Fall 2014

Assessing Positional Accuracy and Correcting Point Data for Digital Soil Mapping at Varying Scales

Minerva J. Dorantes
Purdue University

Graduate School Form 30
(Revised 08/14)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Minerva J Dorantes

Entitled
Assessing Positional Accuracy and Correcting Point Data for Digital Soil Mapping at Varying Scales

For the degree of Master of Science

Is approved by the final examining committee:

Phillip R. Owens

Darell G. Schulze

Zamir Libohova

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Phillip R. Owens

Approved by Major Professor(s): _____

_____

Approved by: Joseph M. Anderson 11/25/2014

Head of the Department Graduate Program Date

ASSESSING POSITIONAL ACCURACY AND CORRECTING POINT DATA FOR DIGITAL SOIL MAPPING AT VARYING SCALES

A Thesis

Submitted to the Faculty

of

Purdue University

by

Minerva J Dorantes

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

December 2014

Purdue University

West Lafayette, Indiana

This is dedicated to all those who dig soils. _\m/

TABLE OF CONTENTS

Page

Page

Page

LIST OF ABBREVIATIONS

FAO – Food and Agriculture Organization
GIS – Geographic information systems
DEM – Digital elevation model
NRCS-USDA – Natural Resources Conservation Service of the United States Department of Agriculture
NASIS – National Soil Information System
NATSGO – National Soil Geographic
STATSGO – State Soil Geographic
SSURGO – Soil Survey Geographic
SCR – Soil characterization record
MUR – Map unit record
TUR – Taxonomic unit record
LiDAR – Light detection and ranging
NDVI – Normalized difference vegetation index
PCA – Principal component analysis
NED – National Elevation Dataset
NAD 83 – North American Datum 1983
UTM – Universal Transverse Mercator
VDCN – Vertical distance to channel network
MRVBF – Multi- resolution index of valley bottom flatness
TWI – SAGA topographic wetness index
OLI – Operational Land Imager
TIRS – Thermal Infrared Sensor
FLAASH® - Fast-Line-of-sight Atmospheric Analysis of Spectral Hypercubes
SAVI – Soil-adjusted vegetation index
FVC – Fractional vegetative cover
ISODATA – Iterative Self-Organizing Data Analysis
FIPS – Federal Information Processing Standard
PLSS – Public Land Survey System
Q, QQ, QQQ – Quarter sections, quarter-quarter sections, quarter-quarter-quarter sections
BLM - Bureau of Land Management
DNR – Department of Natural Resources
Mukey – map unit key
WAAS – Wide Area Augmentation System

ABSTRACT

Dorantes, Minerva J. M.S., Purdue University, December 2014. Assessing Positional Accuracy and Correcting Point Data for Digital Soil Mapping at Varying Scales. Major Professor: Phillip R. Owens.

Accuracy, timeliness, and the effect of scale of soil maps are rarely assessed. The recent increase in the use of GIS technologies and modelling software in natural resources and land management, has increased the demand for soil information at a finer resolution worldwide. Most of the world's developing countries rely on soils information at a scale that is too coarse for practical planning, and have obstacles impeding collection of new data, such as civil war and a lack of collection resources. The United States has an exhaustive collection of soils data at a fine scale. However, its location information is replete with errors and inconsistencies which, if unaccounted for, can affect predictive model estimates. An integrated digital soil mapping methodology is necessary to extract the wealth of knowledge stored in soil survey data for building detailed soil maps and for assessing the positional accuracy of soil pedon data.

Two studies were conducted using public data contained in the U.S. Soil Survey databases. The first study tested the development of an accurate regional-scale digital soil class map by combining new elevation data and satellite imagery. As a result, a model design was created that may be applied in countries with limited soil data. In the second study, several models were developed to assess the locational accuracy of the U.S. Soil Survey pedon points for Indiana. The study resulted in the creation of a more

detailed Public Land Survey System grid, as well as several ArcGIS tools to assign a margin of error to existing soil pedon point locations, which separately or together can be adopted on a national scale.

CHAPTER 1. INTRODUCTION

1.1 Current State of Soil Data

The Food and Agriculture Organization (FAO) of the United Nations (2014) estimates that 842 million people were living with chronic hunger between 2011 and 2013. The problem of food insecurity is just one of many that soils information can help overcome. Global climate change, water scarcity, and land degradation are problems faced by many parts of the world. Even though there is a steady decline in fatalities as a result of these issues, strategies to mitigate the damage too often rely on models lacking concrete soil data. Soils information is an essential input in models that can assess water expenditure, yield, and potential impact on the environment as well as open opportunities for new crop production and improved nutrient management- all of which are vital considerations in combating these global societal issues.

In their 1994 article, "Of Maps and Myths," Estes & Mooneyhan explore the misconception that many environments and services are "well mapped." The myth is perpetuated by reports that show environmental change or the effects of natural disasters displayed through maps. However, accuracy, timeliness, and the effect of scale of these maps are very rarely assessed. They argue that we do not currently have the maps necessary to support detailed land management and public policy decision-making for many parts of the world (Estes & Mooneyhan, 1994). This is true of the currently available soil information in developing countries for several reasons. It is labor intensive and expensive to provide the technical and scientific support needed to develop soil information databases.

Countries often have problems that they consider to be of greater importance than the study and collection of soil data, such as war, sickness, and famine, so the benefits of soil information are not readily apparent to major investors. History shows that agriculture ministries do not often consider soil mapping an imperative until they are confronted with a natural disaster such as severe drought or intense soil erosion (Cook, et al., 2008). Universities and not-for-profit institutions in developing countries may have informative soil legacy data that they cannot publish due to limited funding or limited communication with the rest of the world. Furthermore, several countries have high resolution datasets that could be used in the production of equally detailed soil maps, but they are labeled as "classified" information, or for “internal use only” and are not shared with land management agencies or the international data community.

Developed countries, like the United States, are faced with a different set of issues. The United States holds an extensive collection of soils information compiled over the last century. Of particular importance to current research is the soil pedon point data which includes physical descriptions, lab analysis information, and geographic locations for soils sampled in the field. Its range and versatility make this database a widely used input for predictive and land use models. Unfortunately, due to database migrations, a lack of quality control, and the inherent error associated with soil surveying, the pedon point data is often incomplete, not in a usable format, outdated, or erroneous. Currently, there is no measure of accuracy assigned to these points, leading many modelers to assume that the data location is correct, which can result in inaccurate model estimates and can have serious repercussions.

An integrated digital soil mapping (DSM) methodology is necessary to extract the wealth of knowledge in soil survey data for building detailed soil maps and assessing the validity of currently available soil point data. Advances in computing technology, geographic information systems (GIS) software, modelling, and image analysis techniques, provide the means to examine the current state of available soil data and to develop methods to improve its accuracy and availability. The studies presented in this thesis make use of these tools to address the issues of providing soil information for

developing countries and assessing the positional accuracy of Indiana's soil pedon points. Additionally, they provide methodologies that can be adopted on a national scale.

1.2 Objectives and Hypotheses

Information contained in the U.S. Soil Survey databases was used as the reference for soil information for the two research projects conducted. The first study explores the application of coarser-resolution map units in digital soil mapping (DSM) and the second study explores the high-resolution pedon data. The objectives of the first study were to (i) test the combined use of new elevation data and satellite imagery in developing an accurate regional-scale digital soil class map; and (ii) design a model for regional-scale soil mapping that may be applied in countries with limited soil data inputs. The objectives of the second study were to (i) develop a methodology to improve the positional accuracy of pedon point data for Indiana collected before GPS; (ii) place pedon points in a geographic location within an environment and soil forming factors that were originally described; (iii) assign a measure of positional accuracy to all pedon points; and (iv) use the updated pedon points to interpolate soil organic carbon (SOC) and produce a statewide SOC map of Indiana.

The major hypotheses of the first study were that (i) a combination of topographic data and satellite data in DSM can produce a regional-scale soil class map that is comparable to the Survey's regional-scale maps; (ii) integrating new elevation data with satellite imagery in DSM methods improves the accuracy of resulting regional-scale soil class maps; and (iii) the methodology to produce these maps is time-efficient and requires few inputs so that it may be applied to developing areas with limited soils data. The major hypotheses of the second study were that (i) legacy data and expert knowledge are necessary tools to improve the positional accuracy of pedon points; (ii) pedon points can be moved to a more accurate location by matching their soil

environment to information stored in one of the Survey databases; (iii) ArcGIS tools can be developed to assign a clear measure of positional accuracy; and (iv) the use of updated pedon points in predictive soil models will produce a statewide map that is very different from the current interpolated SOC map.

1.3 Organization

This thesis is organized into four chapters. The first chapter provides the justification for the research and gives a quick overview of the general and specific objectives and hypotheses of the work. The second chapter provides a brief introduction to the structure and contents of the United States Soil Survey databases. The third chapter assesses the usefulness of integrating satellite data with digital elevation data for developing a digital soil map. The third chapter presents the methodology developed to improve the positional accuracy of soil survey pedons in Indiana. A list of references and an appendix with supporting tables and documents follow the fourth chapter.

CHAPTER 2. U.S. SOIL SURVEY

The United States is fortunate to have one of the most extensive soil data collections of any developed nation, greatly due to the early establishment of the U.S. Soil Survey. Soil surveying, which encompasses the study of soil classification, genesis, morphology, performance, and mapping, was first conducted in the 1880's (Lytle, 1999). Financial crises and natural catastrophe, exacerbated by poor land management in the early twentieth-century, sparked great concern for land-use planning. This resulted in a new demand for soil data throughout the country, which was met by efforts to map and describe soils' potential for alternative land management (Gardner, 1998). The National Cooperative Soil Survey (NCSS) was established in 1953 to set standards for field collection and interpretation of soils and to accelerate the mapping process. Today, the Natural Resources Conservation Service of the United States Department of Agriculture (NRCS-USDA) monitors the NCSS and maintains and distributes soil information.

Soil pedon descriptions, lab analysis information, and geographic locations of soils are freely available to the public through the National Soil Information System (NASIS). NASIS is a tool to help manage and maintain information from soil surveys in a hierarchical structure. It is a dynamic database that allows soil surveyors to enter their field data and extract interpreted soil information for use in resource models. NASIS houses three soil geographic databases: National Soil Geographic database (NATSGO), State Soil Geographic database (STATSGO), and Soil Survey Geographic Database (SSURGO). These databases receive information from three different records: Soil Characterization Record (SCR), Map Unit Record (MUR), and Taxonomic Unit Record (TUR) (Grunwald, 2014) and contain maps, tables and metadata.

Each data set contains soil information at a different level of detail. Figure 2.1 shows how the data sets are related.

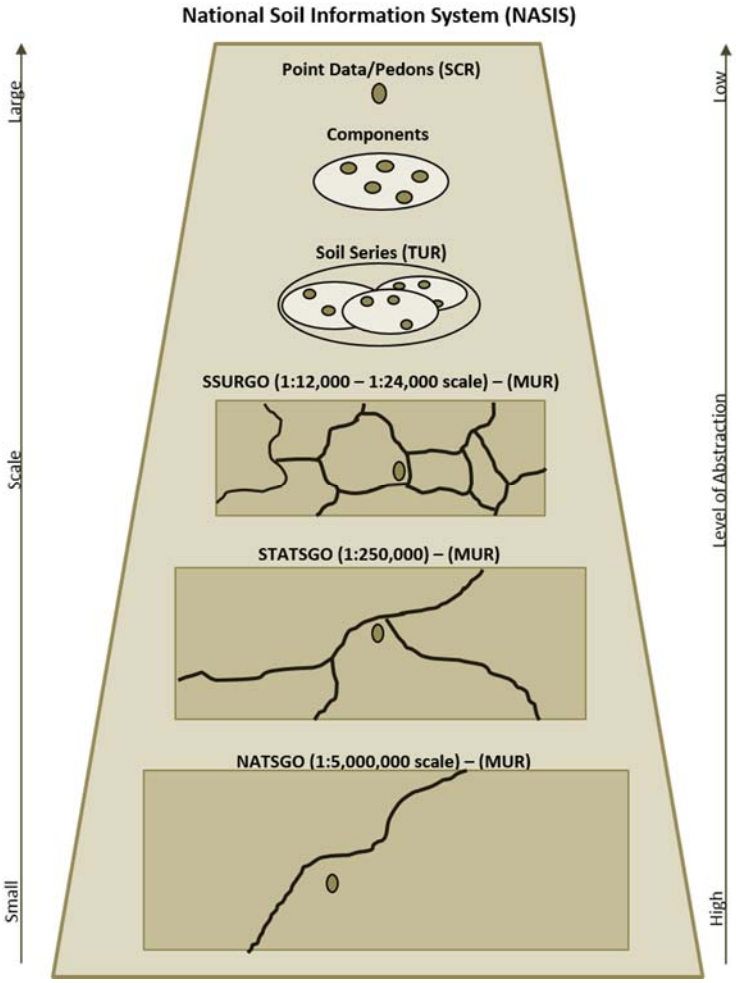


Figure 2.1. NASIS soil data relationships. [Modified from (Lytle, 1999)]

Of the three geographic databases housed in NASIS, NATSGO contains the coarsest information at the smallest (1:750,000 or 1 inch: 11 miles) scale. Its attribute data is derived from MURs from a Major Land Resources Area (MLRA) map and point data from the National Resources Inventory (NRI) that is interpolated to the MLRAs. Due to its coarse resolution and scale, NATSGO data is most appropriate for national or multi-state monitoring and planning.

Of greatest interest to this study are the STATSGO and SSURGO databases which contain much more information at a lower level of abstraction. STATSGO is a set of regional soil maps at a scale of 1:250,000 (1 inch: 4 miles), generalized from more detailed county soil survey maps. They were compiled for use in regional, multi-county, and state and multi-state resource planning to identify potential problem areas, prime farmland, and land cover type in areas with and without soils. Its attribute data is assembled from the MUR and contains information about the components of the map unit and their extent, soil property, and interpretation data. Each map unit can consist of up to twenty-one different components and has a minimum size of approximately 2.3 square miles.

Like STATSGO, SSURGO was digitized from county soil surveys; however, the detail contained in the surveys was retained. This makes SSURGO the most detailed level of soil mapping. Scales of SSURGO maps range from 1:12,000 (1 cm: 0.12 km) to 1:31,680 (1 cm: 0.3168 km) and the smallest areas delineated are 0.005 square kilometers and 0.0388 square kilometers, respectively. SSURGO soil maps were derived from detailed field observations and sampling along transects to delineate boundaries. Aerial photography was most commonly used as the soil map reference base. SSURGO data is useful in field-scale and county-level planning.

The Soil Characterization Record stores point data for over 60,000 sampled pedons in the United States and other countries that was collected from the start of the Soil Survey. It contains descriptive information about the pedon's map location and morphology recorded by NRCS soil scientists and NCSS cooperators from various universities, including Purdue University. The analytical soil information contained in the database was produced by the National Soil Survey Laboratory and collaborating institutions, using standard lab procedures (Soil Survey Staff, 2011). Sampling sites were selected to characterize the central concept of a soil series. These point data are used to populate the MUR and so they are also directly linked to soil series components and the TUR. This soil characterization data is also publically accessible through NASIS and is often used in research environmental models (Lytle, 1999).

Unfortunately, the morphological and geographical data was generally not standardized and much of it was collected before accurate Global Positioning System (GPS) devices were used in the field.

Map units are a multi-level soil group which can contain several components and their soil horizons. The Map Unit Record (MUR) is generated from field observations that were collected as part of transects, grid sampling, measurement of soil properties and for sample collection. The MUR thus contains physical, morphological, and chemical soil property data and interpretations for every part of a soil survey map including the polygon or map unit delineation, line feature, and points. Tables in NASIS are structured to provide information and soil interpretations at the map unit level. The MUR is used in natural resource planning and land management by environmental planners and modelers, land owners and at the local, state and federal level.

The Taxonomic Unit Record (TUR) contains soil taxonomic information at the soil series level from the Soil Series Classification Database (SSCD) and Official Series Descriptions (OSD). The SSCD is managed by the MLRA offices and contains the official taxonomic classification of each established soil series and information about its status, its origin, and geographic range (Lytle, 1999). The SSCD database is continuously updated to reflect changes in Soil Taxonomy and as new soil survey information is provided. The OSD is a text file with a complete physical description of a typical soil series that defines that specific series. A soil series is usually described in the location where it was established and, when possible, exact geographic coordinates are provided. OSDs serve as a guide for identifying and classifying soils in the field as well as in the development of Soil Taxonomy (Lytle, 1999).

CHAPTER 3. AN INTEGRATED APPROACH TO CREATE A DIGITAL SOIL MAP FROM TERRAIN AND SATELLITE DATA

3.1 Introduction

Currently, most developing countries of the world rely on soils information at a scale that is too coarse for practical land use planning. The most-recent global soil map, published by the FAO-UNESCO in 1981 is at a spatial resolution too low for regional-scale land management (Sanchez et al., 2009). More detailed national soil maps (1:1 million or better scale) are available for 109 countries, but these only cover 31% of the earth's land area, leaving the rest of the countries dependent on the FAO-UNESCO map (Nachtergaele, 1999).

3.1.1 U.S. Soil Survey Map Units

Map units are the fundamental units of U.S. Soil Survey maps. They are a soil scientist's interpretation of the soil physical properties and land characteristics bounded by the extent of similar soil forming factors. Soil survey map units are designed to represent observable landscape patterns at a scale that is appropriate for the purpose of a specific soil survey. In this way, they act as substitutes for models of landscape evolution and soil formation (Soil Survey Division Staff, 1993). Map units are described using the taxa of the most extensive single or multiple components, or soil series, contained within the boundaries of that map unit.

There are several kinds of distinguishable map units. “Consociations” are comprised of one dominant component and a set of very similar soils or a non-soil, miscellaneous area. Generally, about half of the pedons of a consociation are the same soil component of the map unit name. “Complexes” and “associations” consist of a group of two or more dissimilar components. Unlike consociations, the components of complexes and associations are very different morphologically and behaviorally. The major components of a complex cannot be separated at a scale of 1:24,000; however, at this scale, the components of an association can be separated (Soil Survey Division Staff, 1993). Undifferentiated groups are the least uniform of the map unit types. They consist of two or more different components that do not always occur in the same map unit. They are included in the map unit name because their application and interpretations for use are very similar.

Different mapping intensities or levels of mapping contain soil map units with varying amount of detail and extent of dissimilarity (Table 3.1). SSURGO maps are produced at a 2^{nd} order of detail. STATSGO maps are produced at a 5^{th} order of detail. Either one can be used in regional-scale planning.

Table 3.1. Mapping detail in U.S. Soil Survey Maps. [Modified from (Soil Survey Division Staff, 1993)]

Level of Mapping	Minimum size delineation (hectares)	Typical components	Kind of map unit
1st order – very intensive (field-scale planning)	1 or < 1	Phases of soil series, miscellaneous areas	Consociations dominate, some complexes, miscellaneous areas
2nd order (SSURGO)	0.6 – 4	Phases of soil series, miscellaneous areas	Consociations, complexes, some associations and undifferentiated groups
3rd order	1.6 - 16	Phases of soil series or taxa above the series or miscellaneous areas	Associations or complexes dominate, some consociations and undifferentiated groups
4th order	16 – 252	Phases of soil series or taxa above the series or miscellaneous areas	Associations dominate, some complexes, consociations and undifferentiated groups
5th order – regional planning (STATSGO)	252 - 4000	Phases of levels above the series and miscellaneous areas	Associations dominate, some consociations and undifferentiated groups

3.1.2 Soil – Landscape Relationships

Due to their scale, soil survey map units do not always capture the complexity of soil properties and soil composition that is observed in nature. However, general as they may appear, soil map units express an understanding of soil-landscape relationships which govern soil development. In this case, 'landscape' refers to the components of topography, including slope, curvature, gradient, permeability, and depth to the water table acting within a three-dimensional area.

Topography, can replace "relief", as one of Hans Jenny's well-recognized factors of soil formation. Soil, he describes, is a function of climate, organisms, and relief acting to change parent material over time.

$$s = f\,(cl, o, r, p, t\ldots)$$

Where:

s = soil, *cl* = climate, *o* = organisms, *r*= relief or topography, *p* = parent material, and *t* = time. The ellipses indicate that other soil forming factors may be included in the relationship (Jenny & Amundson, 1994). Unlike the other major soil forming factors, relief is passive, but it governs potential and kinetic energy in the soil system.

McBratney has modified Jenny's function of soil to express the relationships between soil and the environment in a spatial context, in what is commonly referred as the *scorpan* model for digital soil mapping (McBratney, et al., 2003).

$$S = f(s, c, o, r, p, a, n)$$

Where:

S = (represented as soil classes or attributes) is a quantitative function of: *s*= soil as a class or as a remotely sensed property at a point location), *c* = climate, *o* = organisms, *r*= relief or topography, *p* = parent material, a = age, and n = spatial or geographic position.

If all soil factors except for relief are assumed to be constant, then the function of soil becomes:

$$s = f(r)_{cl,\ o,\ p,\ t}$$

This is commonly recognized as a toposequence and it is often the basis for describing and delineating soil variability in an area.

The interconnectedness between soils along a slope is termed a soil catena. The idea of the soil catena first came about in the 1930's as Geoffrey Milne observed complex soil-landscape relationships in Uganda (Milne, 1936). He coined the term and described it as a sequential change in soil profiles with changing topography within a physiographic region. The distribution of soil types in a catena is dependent on mass fluxes and water drainage governed by slope and elevation. In a catena with a single parent material, soils co-develop along a slope. This leads to the development of an erosional component at the top, a high-energy transport component in the mid-slope, and a depositional component near the base (De Alba, et al., 2004). These different hillslope positions were first defined in *Elements of the Soil Landscape* (Ruhe, 1960) in

terms of their slope profile and slope contour. Ruhe's hillslope positions: summit, shoulder, backslope, footslope, and toeslope, are used by soil scientists to describe pedon locations in the landscape.

Soil surveyors recognize that soils behave as a continuum across short distances. They map soil unit boundaries where they observe discontinuities in the system, because these signal changes in soil properties. These discontinuities include: inflections of slope gradient, topographic divides, and contacts between different rocks or landforms of a different age (Wysocki, et al., 2011). Unfortunately, when used as the basis for soil survey (Hudson, 1992), this soil landscape paradigm is limited by a competing, dominant, taxonomy paradigm.

Soil survey is concentrated on soil taxonomy rather than on soil-landscapes. Soil taxonomy treats soil units or pedons as individuals that can be removed from their setting and placed in a branching classification system for identification. Landform elements such as slope profile and slope contour are often overlooked when delineating soil map units, because the focus is on soil morphology that defines taxonomy. There may be convex, concave, and linear slopes in the same map unit. This produces a unit that groups soils with vastly different functional properties.

McSweeney et al., (1994) suggests that a soil-landscape model can be developed by integrating available topography, vegetation, climate history, and remote sensed data. Primary and secondary attributes can then be use to give values to the landscape raster. This data can be combined to define soil patterns. This approach assumes that map units are primarily derived by tacit knowledge of soil-landscape relationships. Thus, the use of terrain derivatives for comparison of map units from two maps at varying scales of detail, SSURGO and STATSGO, is justifiable.

3.1.3 Remote Sensing Applications in Soil Science

Remote sensing techniques have been used in soil science since the 1970s. Orthophotographs, which are geometrically-corrected aerial photographs that display ground features at X, Y coordinate locations, were used to map soils since they were first developed. As part of reconnaissance mapping, soil scientists studied orthophotographs of their area and drew preliminary boundaries around areas with varying image reflectance. The advent of GIS software and remote sensing technologies, the availability of Digital Elevation Models (DEMs), and a greater demand for digital soil maps have increased the use of satellite imagery in soil science.

Satellite remote sensing technologies can provide information about soil variables. Figure 3.1 highlights portions of the electromagnetic spectrum that are important for soil mapping.

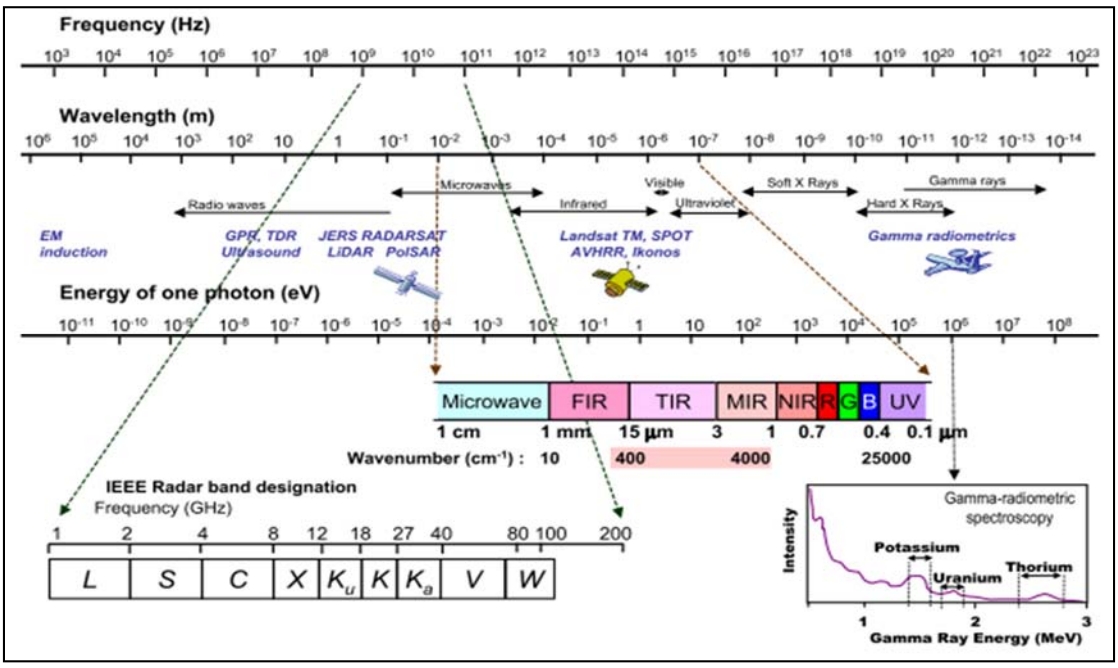


Figure 3.1. Portions of the electromagnetic spectrum which are important for soil mapping. [From (McBratney et al., 2003)]

Although this study focuses on remote sensing techniques, it is important to note that proximal sensing as applied to soil science, is swiftly gaining momentum, especially in radar and gamma radiometric methods to quantify soil surface and subsurface features like surface roughness, soil moisture, and heavy metal content (Rossel, et al., 2010). Light detection and ranging (LiDAR) is another technique that has advanced significantly in the last decade. In addition to elevation models, LiDAR has been used to map soil surface roughness (Bunkin & Bunkin, 2000). Today, LiDAR-derived DEMs, like that used in this study, exist at a 1.5-meter resolution for the entire state of Indiana (Indiana Spatial Data Portal, gis.iu.edu/). Hyperspectral sensors are useful in mapping soil mineralogy and geologic features (King et al., 1995), but they produce an overwhelming amount of information, that it is difficult to sort and use for soil science.

There are currently no methods to accurately map soils using only digital remotely sensed images. This is due in part, to the limited availability of images that are entirely clear of vegetation. Vegetation obscures the soil spectral response and makes it difficult to directly interpret satellite images for soil properties (Campbell, 1987). Soil moisture is also known to interfere with spectral reflectance in the infrared, thermal, and microwave bands. This, in addition to the limitations inherent to all interpretations of remotely sensed imagery, atmospheric disturbances and angle of illumination, are reasons why remote sensing data alone cannot be used to map soils accurately.

Soil distribution can be quantitatively modeled using information derived from a DEM combined with Landsat information. Landsat multispectral data can provide useful environmental covariates that serve as surrogates for the five factors of soil formation for digital soil mapping. Vegetation can be represented by band ratios, such as the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973), that exploit the reflectance properties of photosynthetic plants.

Additionally, models have been developed that adjust for the effects of soil spectral interference and measure soil-related properties (Huette, 1988). Using these methods with satellite and DEM data, Dobos et al. (2001) corrected distortions in their DEM and produced data for soil-landscape modelling. In arid and semi-arid regions, Landsat is also

useful to identify soil parent materials, especially carbonates, sulphates, and gypsic materials. The Landsat 7 band ratios, 3/7, 5/7, and 3/2 were interpreted to represent ferrous iron, hydroxyl radicals, and carbonate radicals, respectively (Boettinger et al., 2008). In their study, Saunders and Boettinger (2007), used the parent material band ratios described above, terrain derivatives (slope, and compound topographic index), and NDVI in digital soil mapping for soil classification. Using principal component analysis (PCA), many spectral bands can be compressed into a few principal components to reduce data size and highlight only the greatest variance. Various supervised and unsupervised classification methods have been used to recognize soil-landscape patterns. Training sites for supervised classification can be selected from available soil surveys or from field-verified points.

3.1.4 Summary of Project Objectives

The development of digital remote sensing techniques that integrate GIS and soil mapping is promising for reducing the cost of producing an accurate regional-scale soil map. Building on previous studies on digital soil mapping using Landsat 7, this study combines Landsat 8 data with DEM data to generate a regional-scale soil map for Tippecanoe County, IN. The objectives of this study were to test the use of new elevation data and satellite imagery in developing an accurate regional-scale digital soil map. This work could prove useful in developing a methodology to create digital soil maps for countries with limited data.

3.2 Materials

3.2.1 Study Area

Tippecanoe County, in west-central Indiana, was chosen as the study area. Tippecanoe County occupies a total land area of approximately 1300 square kilometers.

The five soil forming factors for the study area were considered for generating the appropriate information from the DEM and satellite data. Soil map units from the STATSGO and SSURGO polygon files were used as the “ground truth” soils data for digital soil mapping of Tippecanoe County and a smaller area of interest, respectively (Figure 3.2).

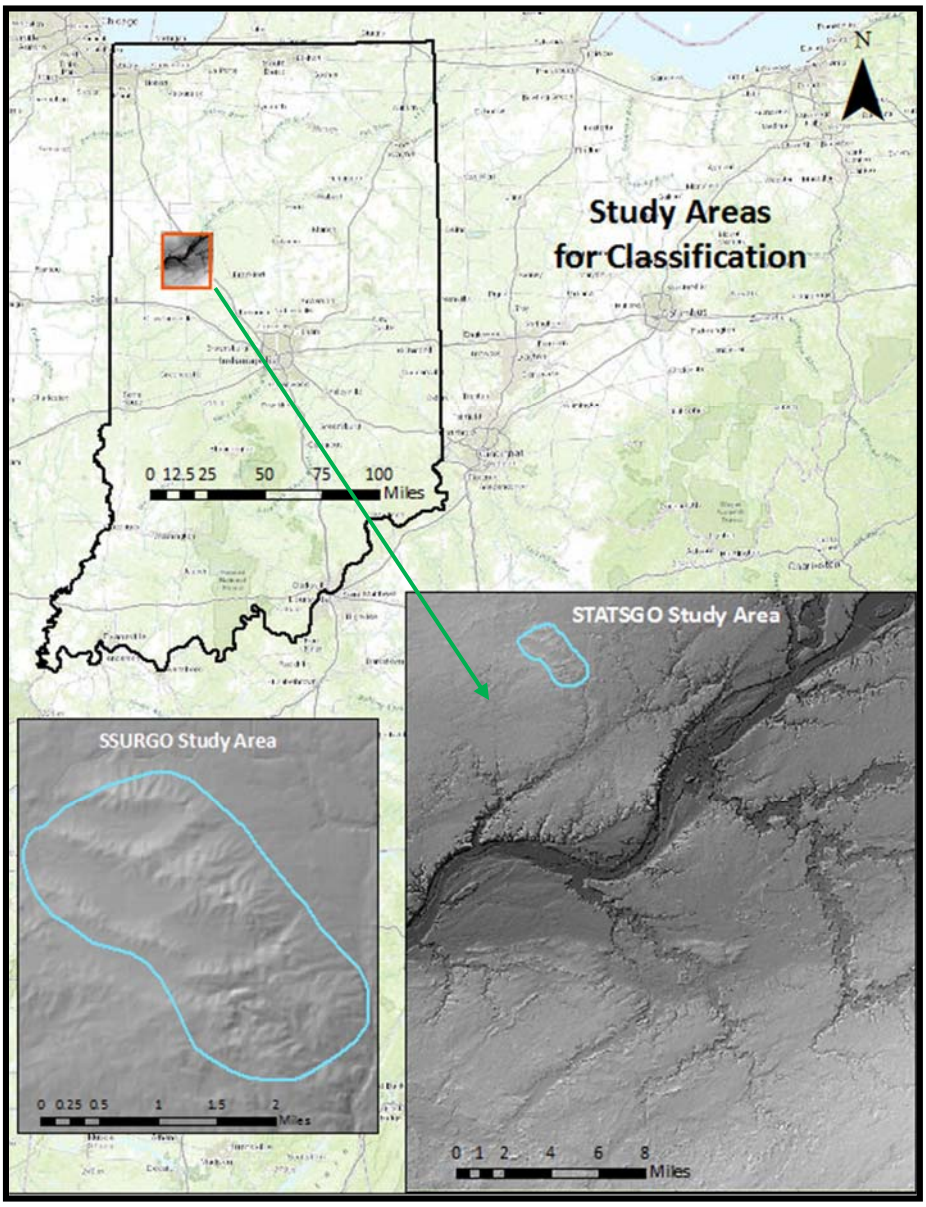


Figure 3.2. Study areas for the classification using the STATSGO and SSURGO reference files. Maps are in North American Datum 83, UTM Zone 16 North projection.

3.2.2 Soils of Tippecanoe County, Indiana

The present-day climate of Tippecanoe County is typical for the Midwest. It is characterized by warm to very warm summers with an average high temperature of 30 degrees Celsius. Winters are cold and often have long periods of subzero temperatures. The annual average amount of snowfall is 54 centimeters and the average low temperature is 4 degrees Celsius (NOAA, 2004). Tippecanoe County, like most of Indiana, is described as having a hot summer continental climate or type *Dfa* in the Köppen-Geiger climate classification system (Peel, et al., 2007). The climate of Tippecanoe County is very different now than it was in the past.

During the period of time known as the Pleistocene Epoch, or more commonly known as the Great Ice Age, Tippecanoe County was entirely covered by glaciers. These glaciers scoured the land and carried materials as they advanced and receded until they disappeared from Tippecanoe about fifteen-thousand years ago (Fleming, 2013). The erosive movement of the glacial ice flattened the land surface, creating the low-relief landscape that we see today. Glacial material, in the form of outwash or water-sorted and stratified coarse sands, gravel, and cobbles, was deposited along the waterways and till or unsorted rocks and sediments of all sizes was deposited everywhere else on the landscape. The glacial geology of Tippecanoe County is also evident in landforms such as kames, eskers, and depressions that are scattered about the till plains. After the glaciers receded north, westward winds deposited a mantle of loess over all of Tippecanoe County.

Following the glaciation, prairie and eastern deciduous forest were the two main plant communities that covered Tippecanoe. The prairie peninsula, or a great expanse of grasses, stretched as far east as the Wabash River in Lafayette. The original prairie vegetation in Indiana was very uniform indicating that the climate throughout this peninsula was also similar during this time. Historically, the prairie peninsula stretched to the west of the Wabash River in Lafayette and the eastern deciduous forest to the

east of the River. The reason for this pattern correlates with climate patterns in that area. Longer, dryer, hotter summers and low precipitation in the winter favors prairie vegetation, whereas more evenly distributed precipitation and less drought frequency favors deciduous plants to the east (Peterson, 1990).

The soils of Tippecanoe County reflect its glacial history, past climate, and the plant organisms that once thrived in the land. The soils can by divided into two categories by their age: (1) soils that are about fifteen-thousand years old and formed from outwash, till or other glacial materials, and (2) soils that are much younger and formed on the floodplains of the Wabash River and smaller waterways and wetlands. Soils that formed from till are dense and poorly drained below the loess layer, but those formed from outwash are more permeable to water. Soils that formed in the floodplains have varying soil properties depending on their location, the force of the water current that deposited their parent materials, and their relative position on the landscape. One thing they have in common is that they often have buried surface horizons (Buol, et al. 2003). Soils formed in organic deposits like those in wetlands are very high in organic matter and ponded, if in their natural state (Franzmeier, 1995). Where the soils are naturally drained and not affected by severe erosion, they exhibit the properties inherent of their past climate and vegetation. Soils formed in prairie vegetation have deep, dark, fertile soil surfaces that are high in organic matter due to the high biomass of prairie grass. Forest soils have a shallow, light-colored surface with lower organic matter. Prairie soils are naturally more fertile than forest soils.

Even though the soils of Tippecanoe County formed from different parent materials and vegetation, the biggest driver of soil functions is the impact of topography and land use. About 8%, or 98 square kilometers of the land area of Tippecanoe County is urban or developed land (USDA-NASS, 2013). Many urban soils are buried under buildings, roads, and sidewalks. Other urban soils have been mixed and moved to level the land for construction. Agricultural soils make up most of the soils outside of the cities. Many of these soils are at least partly eroded from centuries of agricultural use. Their chemical and physical properties have also been somewhat altered by the

installation of tile drains, ditches, fertilizers, and chemical amendments. This study will focus on soil differences that result from soil-landscape relationships. It relies on the catena concept that topography drives water movement and material distribution that creates soil patterns on the landscape, in order to classify different soils.

3.2.3 Soil Map Units from STATSGO and SSURGO

Two sets of reference soil maps were used in this research, a STATSGO map and a SSURGO map. The data was downloaded from the USDA Geospatial Data Gateway (http://datagateway.nrcs.usda.gov/). The STATSGO polygon file was clipped to the extent of Tippecanoe County and the SSURGO file, due to its higher level of detail, was clipped to the extent of a single STATSGO map unit for ease in mapping (Figures 3.3a. and 3.4a.). The resulting soil map units from the STATSGO and SSURGO map sets were reclassified if their extent was too small (Figures 3.3b. and 3.4b.). The STATSGO and SSURGO map units were reclassified in order to combine soil classes that occupied a small area in the study area of interest and to assign these map units or soil groups a more manageable, unique identification number (i.e. 1-13 rather than a 4-digit number). Tables 3.2 and 3.3 provide a list of map units from STATSGO and SSURGO for the areas of interest and their assigned class number for this study. The Reclassified MUKEY column is only used as a reminder of which map units were merged.

Table 3.2. STATSGO map unit legend for small area of interest.

Class #	MUSYM	muname	Reclassified MUKEY
1	s2276	Strawn-Miami-Hennepin	2276
2	s2277	Morley-Markham-Beecher-Ashkum	2277
3	s2315	Treaty-Crosby	2315
4	s2323	Warsaw-Shipshe-Elston	2323
5	s2325	Westland-Ockley-Fox	2325
6	s2327	Sawmill-Lawson-Genesee	2327
7	s2326	Miamian-Fincastle-Brookston	2331
7	s2331	Miamian-Fincastle-Brookston	2331
7	s6056	Miamian-Fincastle-Brookston	2331
8	s2332	Miami-Crosby	2332
9	s2334	Miami-Fincastle	2334
10	s2335	Drummer	2335
11	s2337	Montmorenci-Gilboa-Drummer-Barce	2337
12	s2345	Warsaw-Lorenzo-Dakota	2345
13	s2346	Starks-Rockfield-Fincastle-Camden	2346
13	s2363	Starks-Rockfield-Fincastle-Camden	2363
14	s3263	Spinks-Oshtemo-Houghton-Fox-Boyer	3263

Table 3.3. SSURGO map unit legend for small area of interest.

Class #	MUSYM	muname	mukind	Reclassified MUKEY
1	BgA	Beecher silt loam, 0 to 2 percent slopes	Consociation	163853
2	Du	Drummer soils	Complex	163953
2	Mu	Milford silty clay loam, pothole	Consociation	163953
2	RcA	Raub-Brenton complex, 0 to 1 percent slopes	Complex	163953
2	SyF	Strawn-Rodman complex, 18 to 50 percent slopes	Complex	163953
3	MmB2	Marker silt loam, 2 to 6 percent slopes, eroded	Consociation	164133
3	Pk	Peotone silty clay loam, pothole	Consociation	164133
4	MsC2	Miami silt loam, 6 to 12 percent slopes, eroded	Consociation	164148
4	MsD2	Miami silt loam, 12 to 18 percent slopes, eroded	Consociation	164148
5	MtC3	Miami clay loam, 6 to 12 percent slopes, severely eroded	Consociation	164154
6	OmC2	Octagon silt loam, 6 to 12 percent slopes, eroded	Consociation	164203
7	OpC3	Octagon clay loam, 6 to 12 percent slopes, severely eroded	Consociation	164210
8	Pg	Pella silty clay loam, pothole	Consociation	164238
9	Sn	Sloan clay loam, occasionally flooded	Consociation	164300
10	Cm	Chalmers silty clay loam	Consociation	164327
10	TfB	Throckmorton silt loam, 1 to 3 percent slopes	Consociation	164327
10	TmA	Toronto-Millbrook complex, 0 to 2 percent slopes	Complex	164327
11	TnB2	Toronto-Octagon complex, 2 to 6 percent slopes, eroded	Complex	164331
12	W	Water	Consociation	164359
13	Wb	Wallkill silt loam, coprogenous earth substratum	Consociation	164362

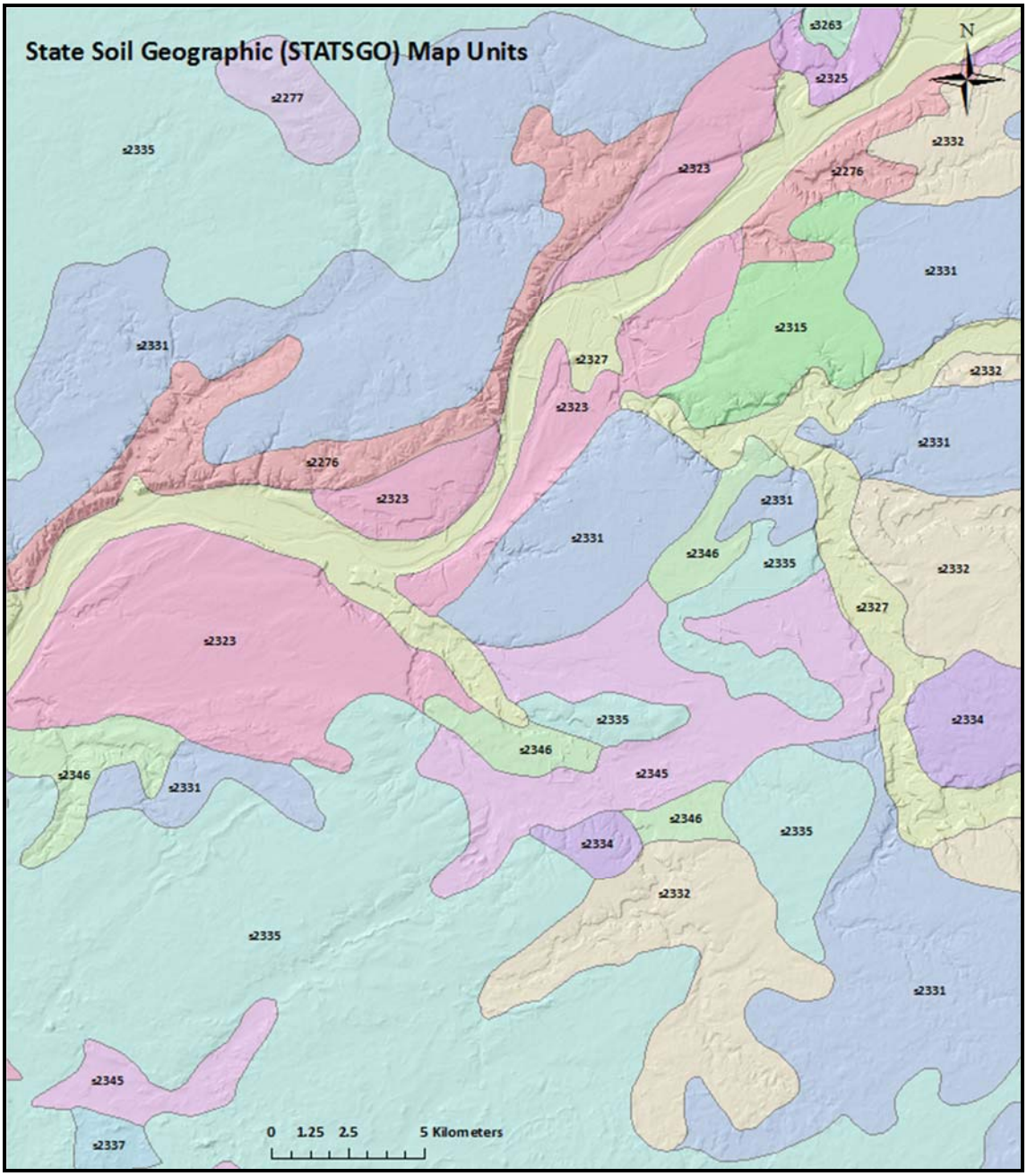


Figure 3.3a. STATSGO map units for Tippecanoe County, Indiana overlain on a shaded-relief map.

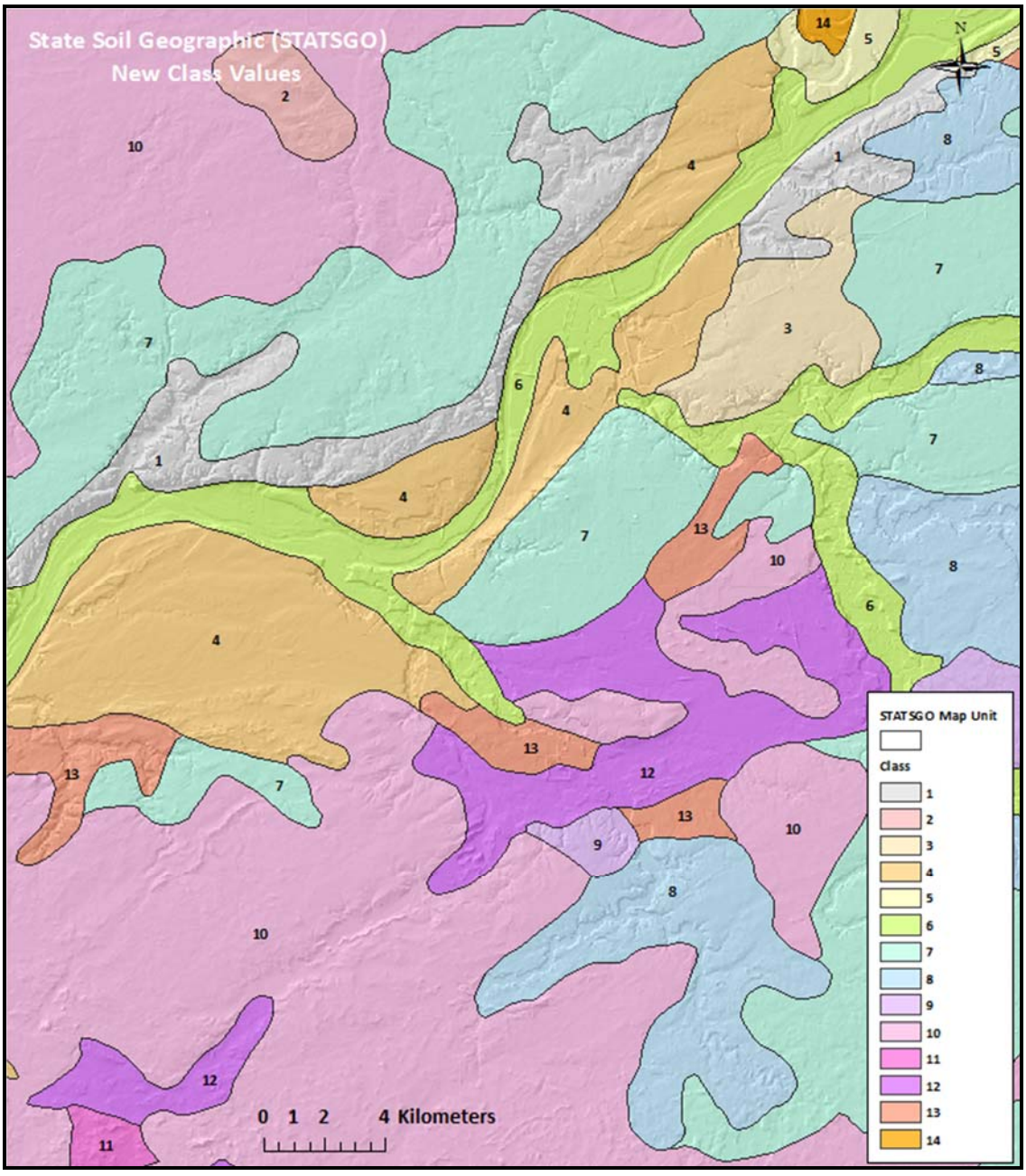


Figure 3.3b. Reclassified STATSGO map units for Tippecanoe County, Indiana overlain on a shaded-relief map.

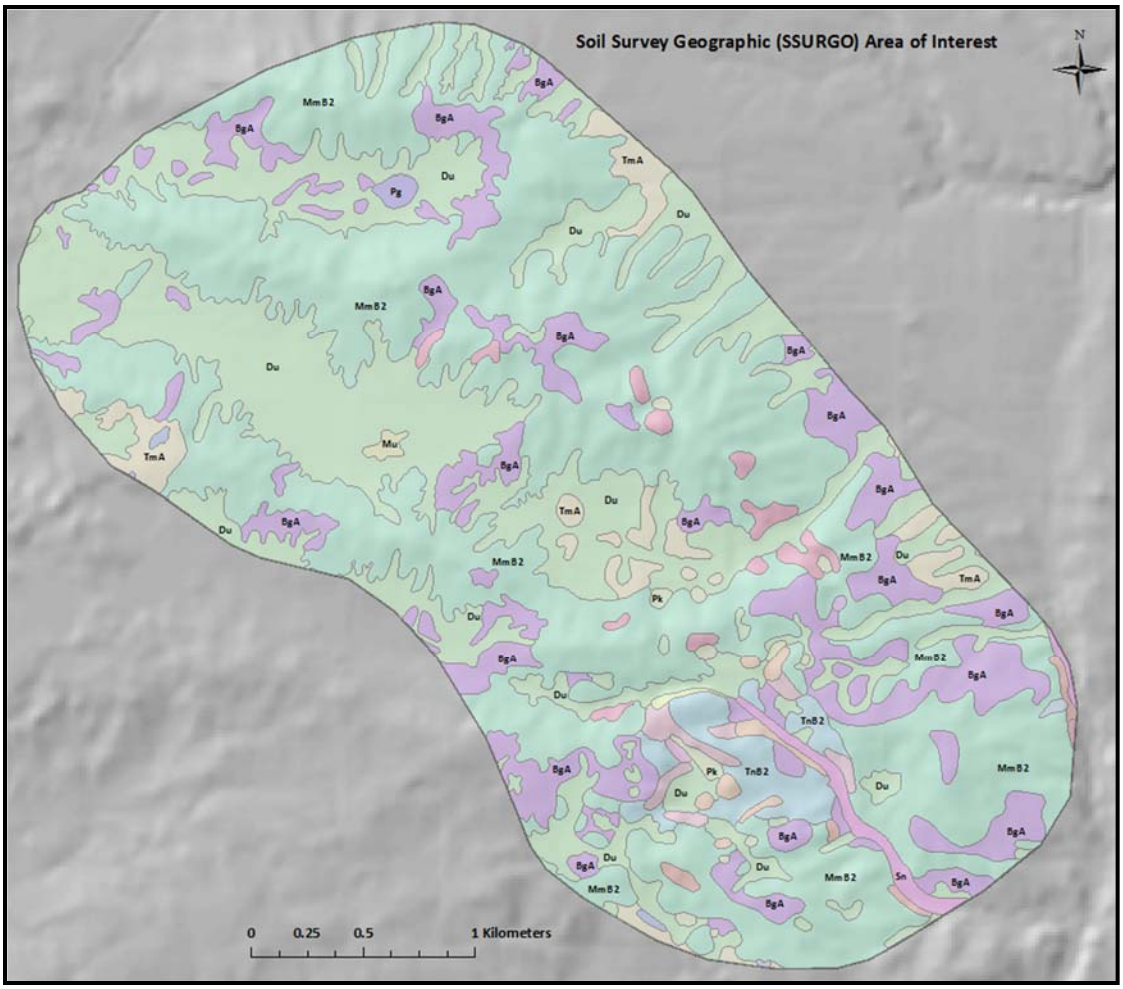


Figure 3.4a. SSURGO map units for a small area in Tippecanoe County, Indiana overlain on a shaded-relief map.

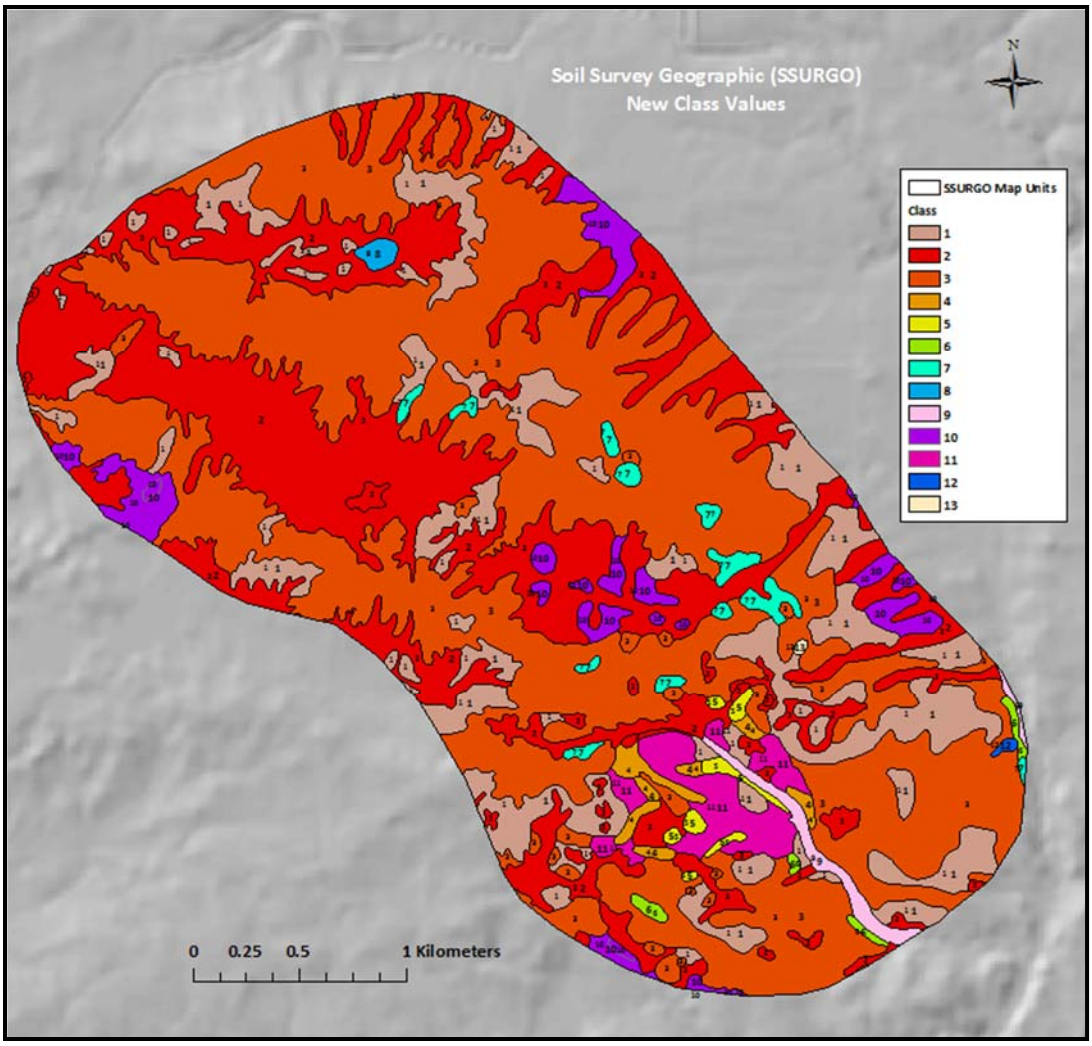


Figure 3.4b. Reclassified SSURGO map units for a study area within Tippecanoe County, Indiana overlain on a shaded-relief map.

3.2.4 Digital Elevation Data

Several DEMs were considered for use in this study: NED, 2005 IndianaMap, and LiDAR. The National Elevation Dataset (NED) was downloaded from the USGS website: http://ned.usgs.gov/. The NED DEM is derived from a combination of photogrammetric terrain data, digitized contours, and interferometric synthetic aperture radar data. The 2005 IndianaMap DEM was downloaded from the Indiana University geospatial data portal (http://www.gis.iu.edu/datasetInfo/elevation). The 2005 IndianaMap DEM is derived from high-resolution color orthophotography. A high-resolution LiDAR DEM, from February 2014, was downloaded from the Indiana University Geospatial Data Portal. All three DEMs were brought into the ArcGIS version 10.2 software and projected to the North American Datum 1983 (NAD 83), Universal Transverse Mercator (UTM) zone 16 north coordinate system. A tool was used to fill sinks in the DEMs, and then they were resampled to 10-, 20-, and 30-meter resolution and clipped to the extent of Tippecanoe County. The files were imported into SAGA GIS (www.saga-gis.org) where the slope was calculated for all DEMs and a visual inspection was conducted to pick the best DEM for this study (Figures 3.5 and 3.6).

The LiDAR DEM was selected because: (i) it has a wider range of elevation values, (ii) it is easier to distinguish between changes in elevation than the other two DEMs, and (iii) details such as a stream's dendritic pattern are much more clearly defined in the LiDAR DEM. Several passes of mean filters were applied to the original, 1.5-meter LiDAR DEM, as recommended by MacMillan, et al. (2003). A similar method of smoothing the DEM was used because it "brings out the longer-range signals and masks the local shorter-range signals" (MacMillan et al., 2003). Slope was calculated for each filtered elevation file, using SAGA GIS (Figure 3.7). A single 3x3 low pass filter was chosen after a visual comparison of several filters, because the resulting elevation map retained its natural appearance and this filter removed many unnecessary areas of micro-relief, such as small ditches and roads, which would have caused noise in the study predictions.

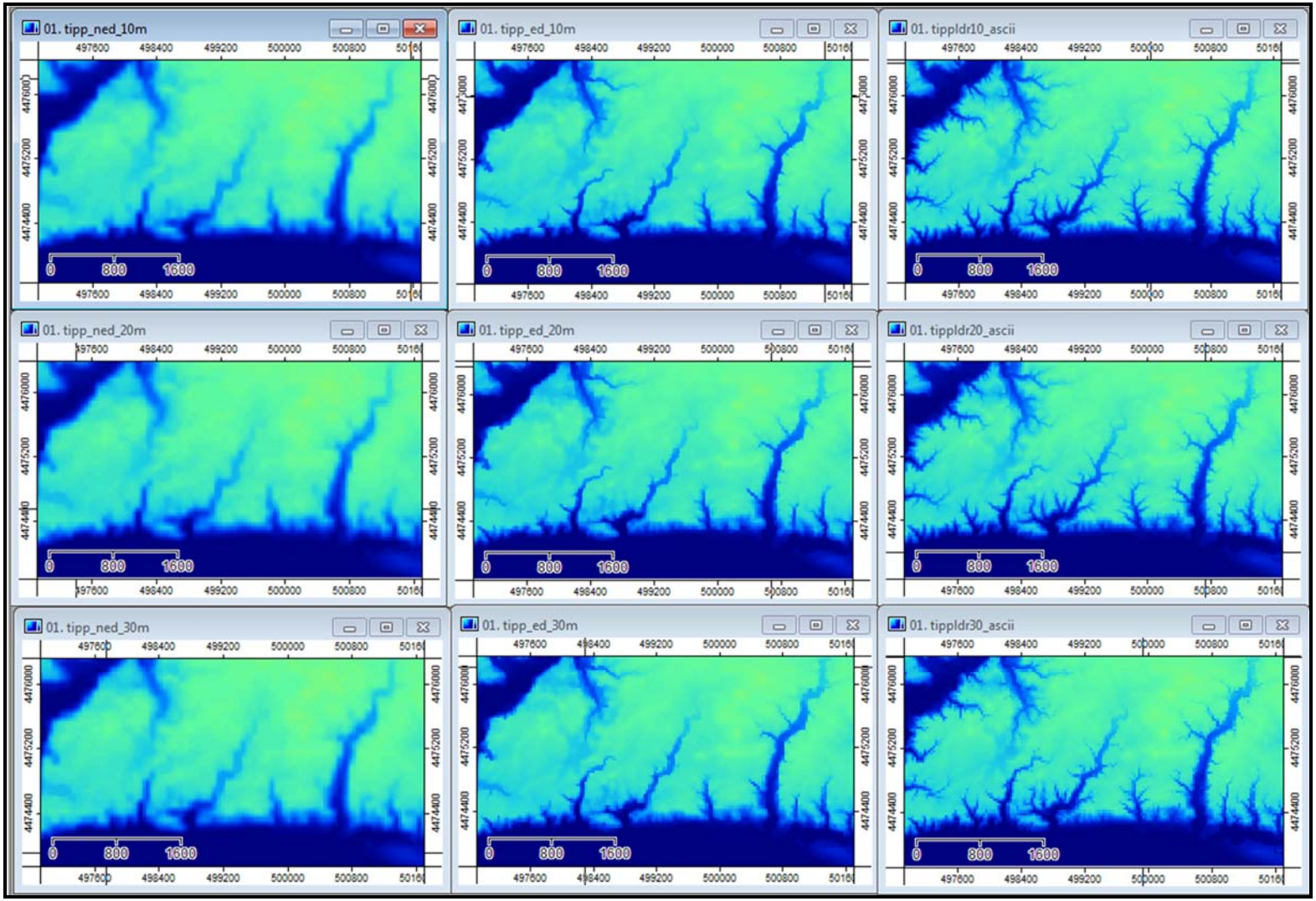


Figure 3.5. Side-by-side comparison of NED (left), 2005 IndianaMap (middle), and LiDAR (right) DEMs

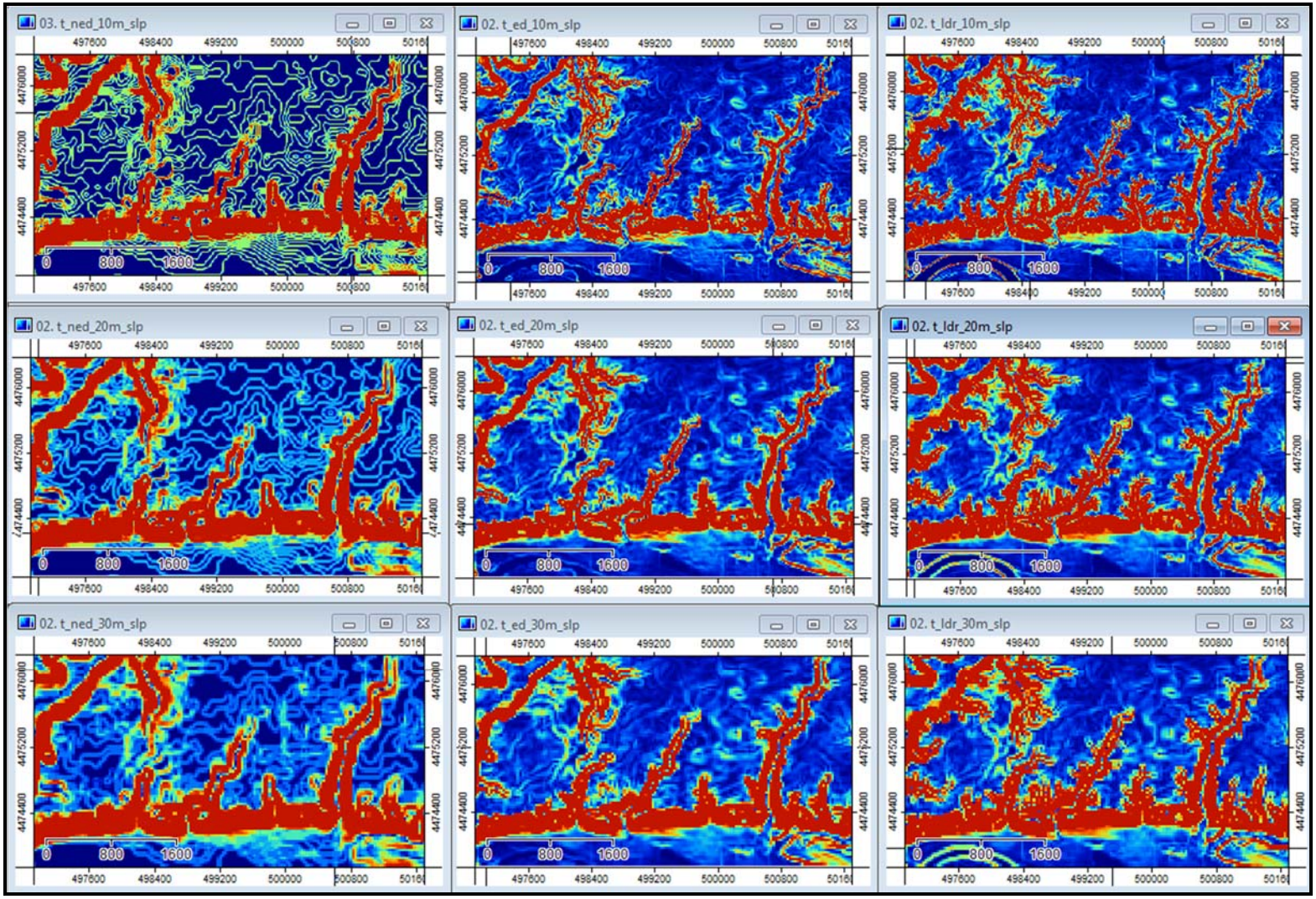


Figure 3.6. Side-by-side comparison of NED (left), 2005 IndianaMap (middle), and LiDAR (right) slope

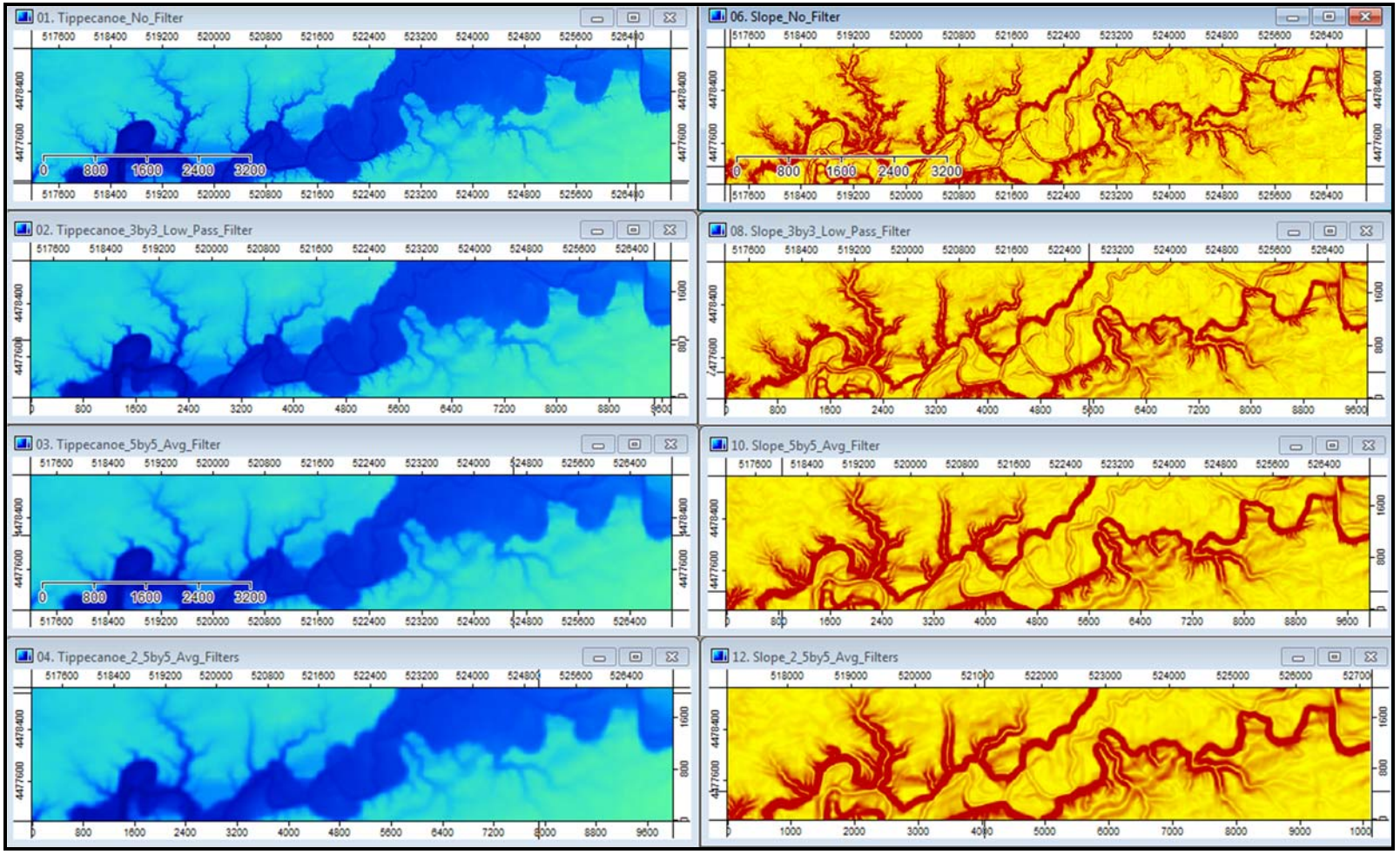


Figure 3.7. LiDAR DEM and slope files with and without low pass filters

3.2.4.1 → Terrain Derivatives to Represent the Soil-Landscape Model

Terrain derivatives provide a means to quantify the soil-landscape relationships by producing a raster layer with values for every pixel (Thompson et al., 2001). Terrain derivatives were generated using SAGA GIS version 2.1.0, ArcGIS version 10.2.2, and GRASS GIS version 7.0 (http://grass.osgeo.org). Several terrain derivatives that are commonly used in digital soil mapping (Evans, 2013; McBratney et al., 2003) were calculated from the 3x3 low pass filtered, 1.5-meter LiDAR DEM, but only those that best represented the natural variability in soil types were used in the study and are described here.

SAGA GIS was used to generate the vertical distance to channel network (VDCN), multi-resolution index of valley bottom flatness (MRVBF) (Gallant & Dowling, 2003), and the SAGA wetness index (TWI) (Sorensen et al., 2006).

The VDCN, pictured in Figure 3.8a, is useful to highlight soils along floodplains. The VDCN is a grid data set of the vertical distance from a grid cell to the closest stream channel line below. It can be expressed as:

$$\mathrm{VDCN} = \mathrm{DEM} - \text{base level or stream channel elevation.}$$

(Bock & Kothe, 2008).

The MRVBF (Figure 3.8b.) is calculated from the inverse of slope and relative depth to surrounding pixels within a circular neighborhood. It is useful in highlighting soils in the highest and lowest positions on the landscape. The MRVBF is computed through a series of steps:

Step1: A nonlinear function is used at several points in the algorithm to map an output value x>0 to a range of 0 to 1:

$$N(x,t,p) = \frac{1}{1+\frac{x^p}{t}}$$

where:

N = new number from 0 to 1

p= a shape parameter

t = a threshold

Step 1:

Slope is calculated and transformed to flatness (0 to 1) using the nonlinear function with shape parameter (*p*) = 4:

$$F_1 = N(S_1, t_{s,1}, 4)$$

where:

F_1 = flatness

S_1 = % slope

$t_{s,1}$ = slope threshold based on the range of slope in DEM used to develop the index

p = 4

Elevation percentile (PCTL) is calculated for this step using a radius of 3 DEM cells. It is transformed to a local lowness value using the nonlinear transformation and then combined with flatness to produce the preliminary valley flatness index (PVF_1):

$$PVF_1 = F_1N\,(PCTL_1, 0.4, 3)$$

where:

threshold (*t*) = 0.4

shape parameter (*p*) = 3

F_1N = flatness

PVF_1 = preliminary valley flatness

The PVF_1 is transformed using:

$VF_1 = 1 - N\,(PVF_1, 0.3, 4)$

where:

VF_1 = index of valley flatness

Values for VF_1 that are less than 0.5 are too steep or high on the landscape to be considered valley bottoms. The higher the value for VF_1, the more dominant the valley bottom character (Gallant & Dowling, 2003).

The TWI is a measure of the potential of a soil to accumulate water. The TWI (Figure 3.8c.) is defined as:

$$ln\ (\alpha/tan\beta)$$

Where:

Ln = natural log, α = upslope contributing area per unit of contour length, and β = slope. The SAGA wetness index is different from the conventional topographic wetness index in that it considers multidirectional flow as opposed to just flow from a single direction, so it is a better model for the natural system.

Miller (2013) calibrated profile curvature, slope, and relative elevation neighborhood sizes to soil scientists' knowledge for identification of hillslope position in the field. Relative elevation was calculated for Tippecanoe County using his recommended neighborhood size of 135 meters. The Hillslope Position Toolbox for ArcGIS (Miller, 2013) was used to generate the relative elevation grid (Figure 3.8e). The tool calculates the relative elevation uses three steps as pictured in Figure 3.8d:

1). A reference ceiling elevation is calculated by summing the minimum and maximum elevations in the neighborhood.

2). Next, the central cell elevation is subtracted by the reference ceiling elevation to produce an inverse elevation grid.

3). Finally, a relative elevation grid is produced by subtracting the inverse grid from the original elevation grid. The mid-point between the maximum and minimum elevation values in the neighborhood is zero. Values above the midpoint are positive and those below the midpoint are negative.

The profile curvature is the curve intersecting with the z-axis and the direction of the maximum change in elevation. Positive values describe a convex profile curvature and negative values describe a concave profile curvature.

The profile curvature (Figure 3.8f.) was derived in GRASS GIS using Miller's recommended neighborhood size of 63 meters.

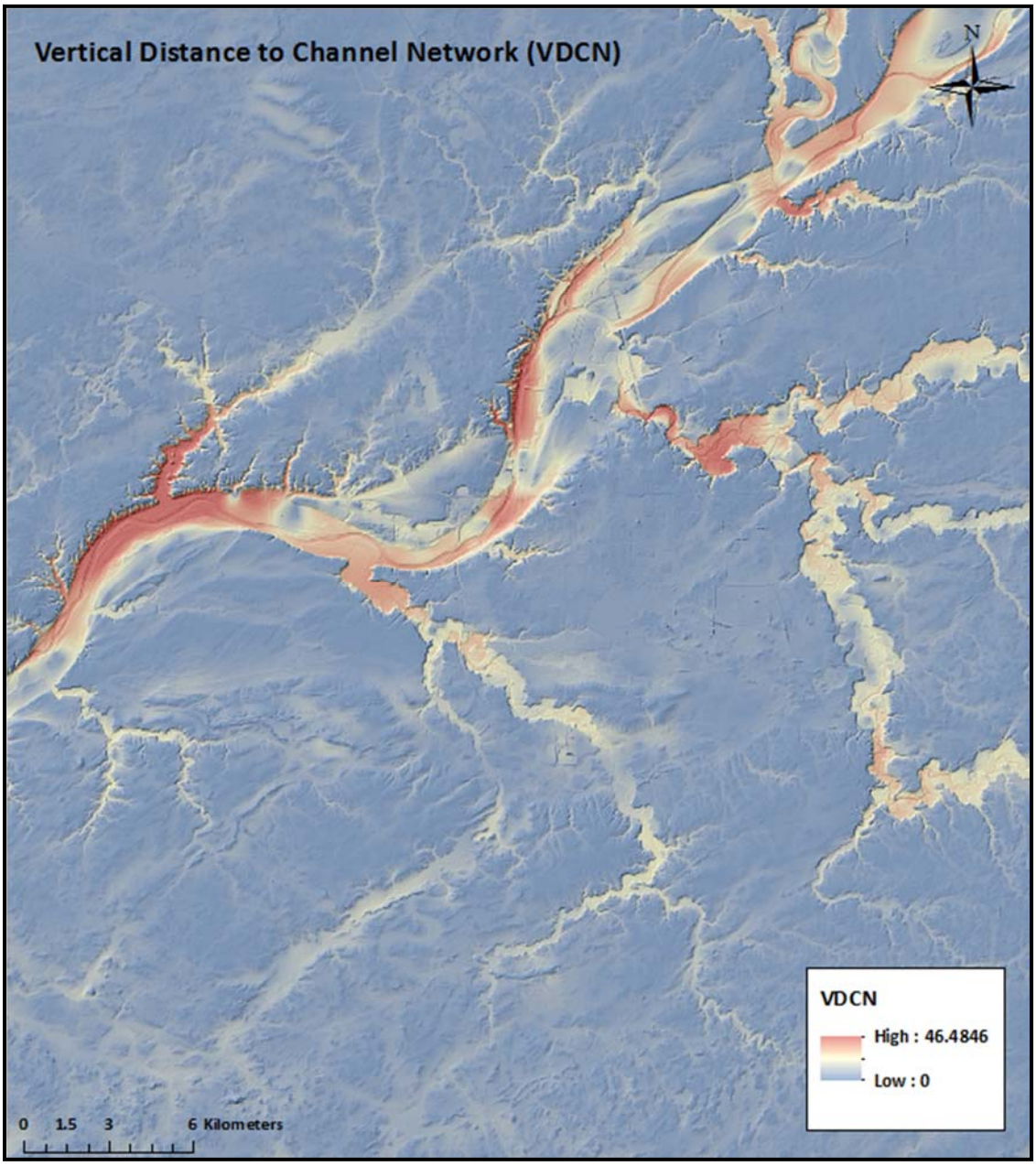


Figure 3.8a. Vertical Distance to Channel Network for Tippecanoe County, Indiana

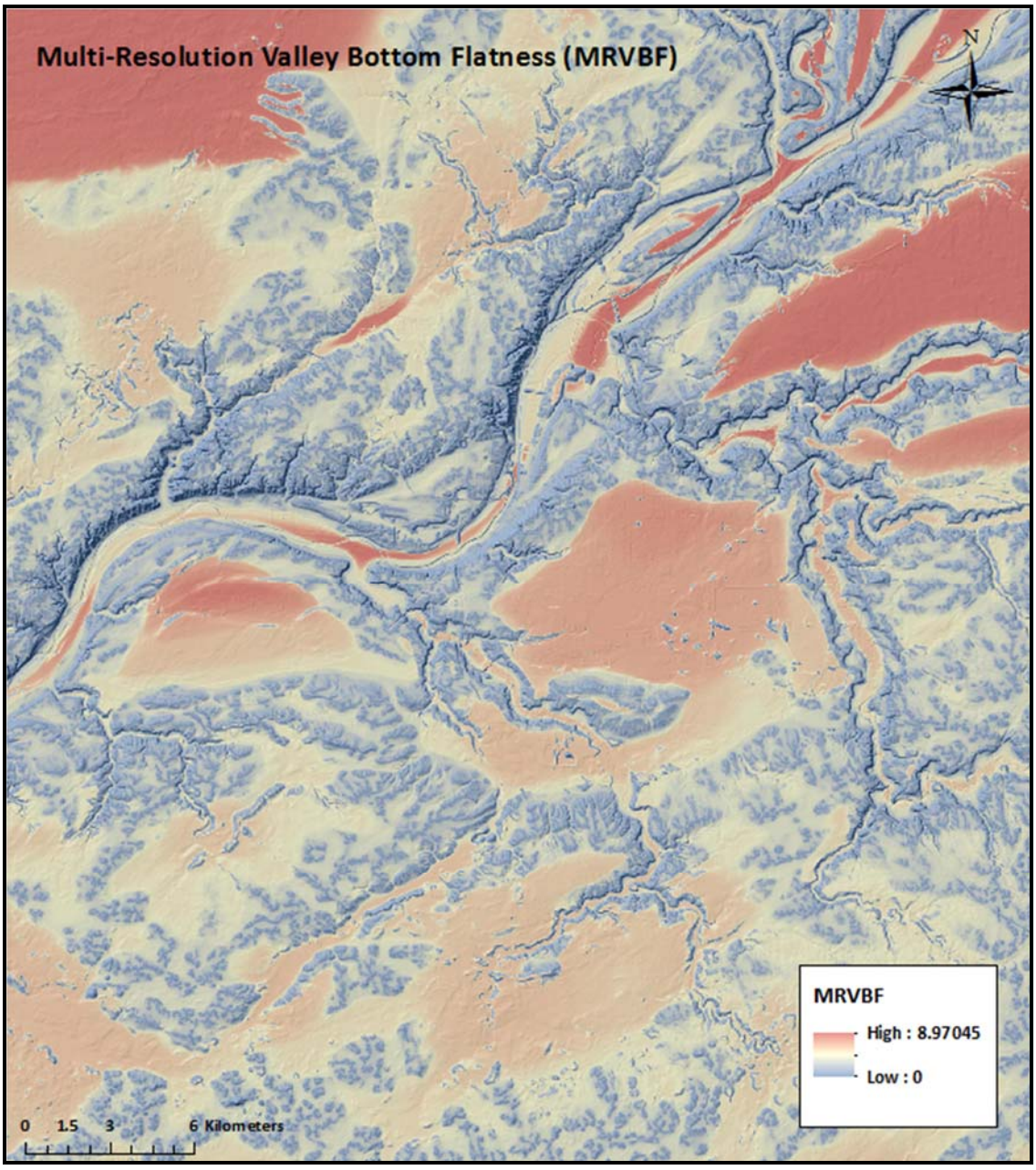


Figure 3.8b. Multi-Resolution Valley Bottom Flatness Index for Tippecanoe County, Indiana

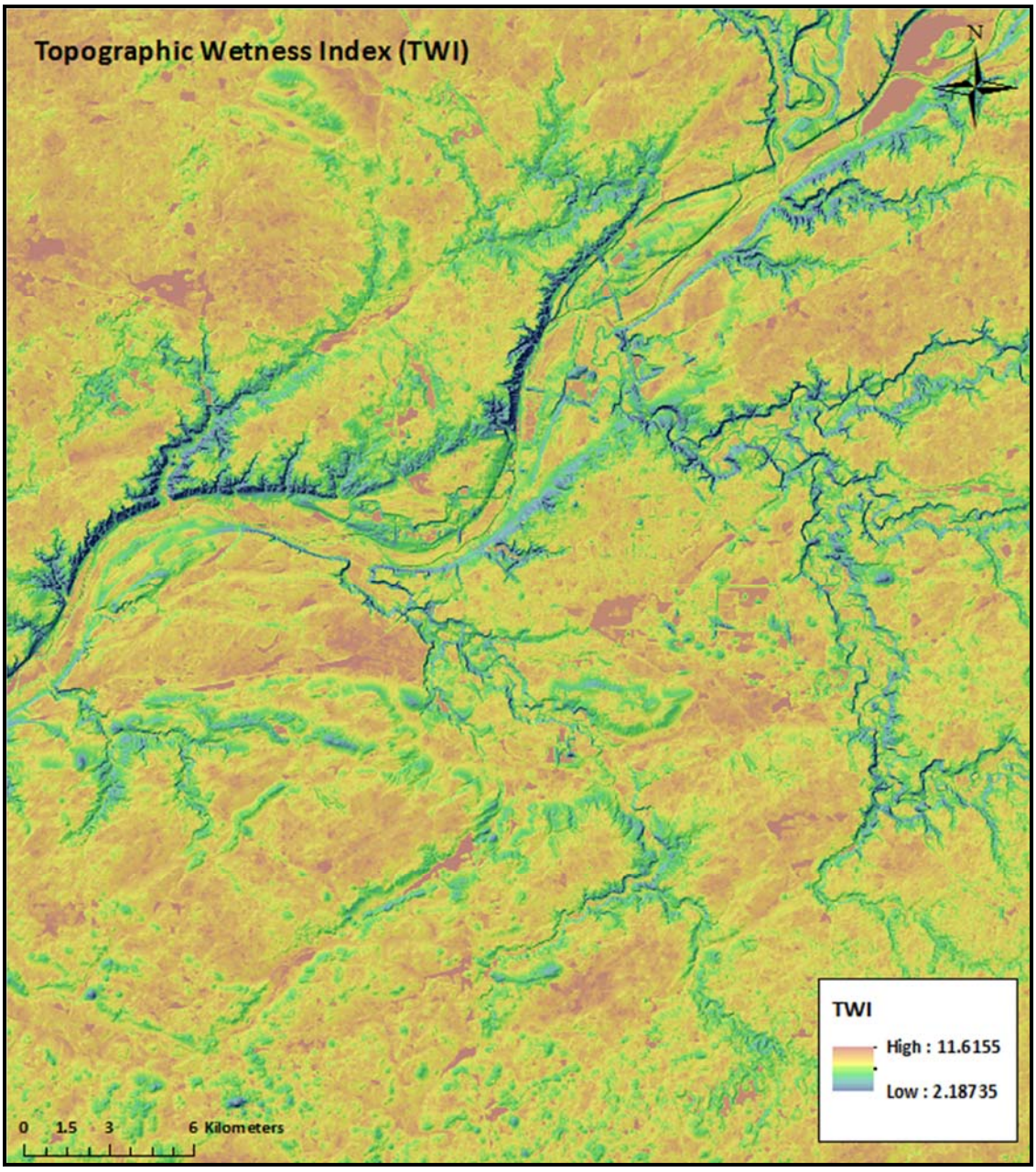


Figure 3.8c. SAGA Topographic Wetness Index for Tippecanoe County, Indiana

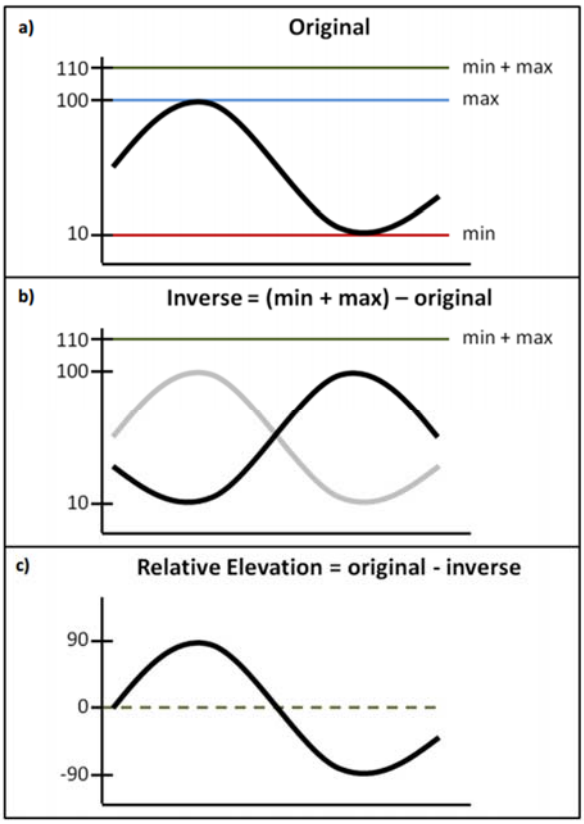


Figure 3.8d. Relative elevation is calculated for each cell on neighborhood basis. Relative elevation is calculated for each cell on neighborhood basis.

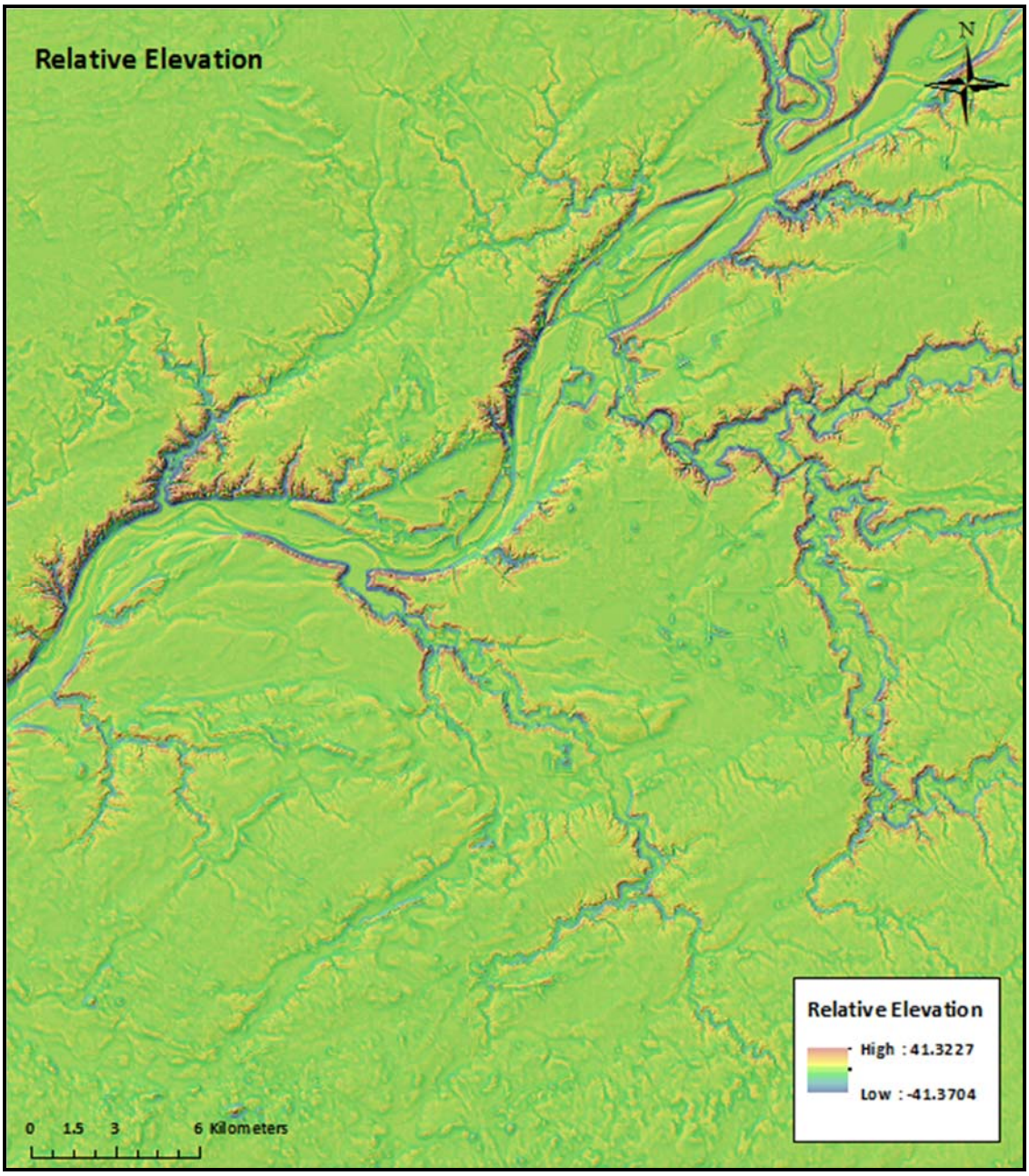


Figure 3.8e.Relative elevation for Tippecanoe County, Indiana

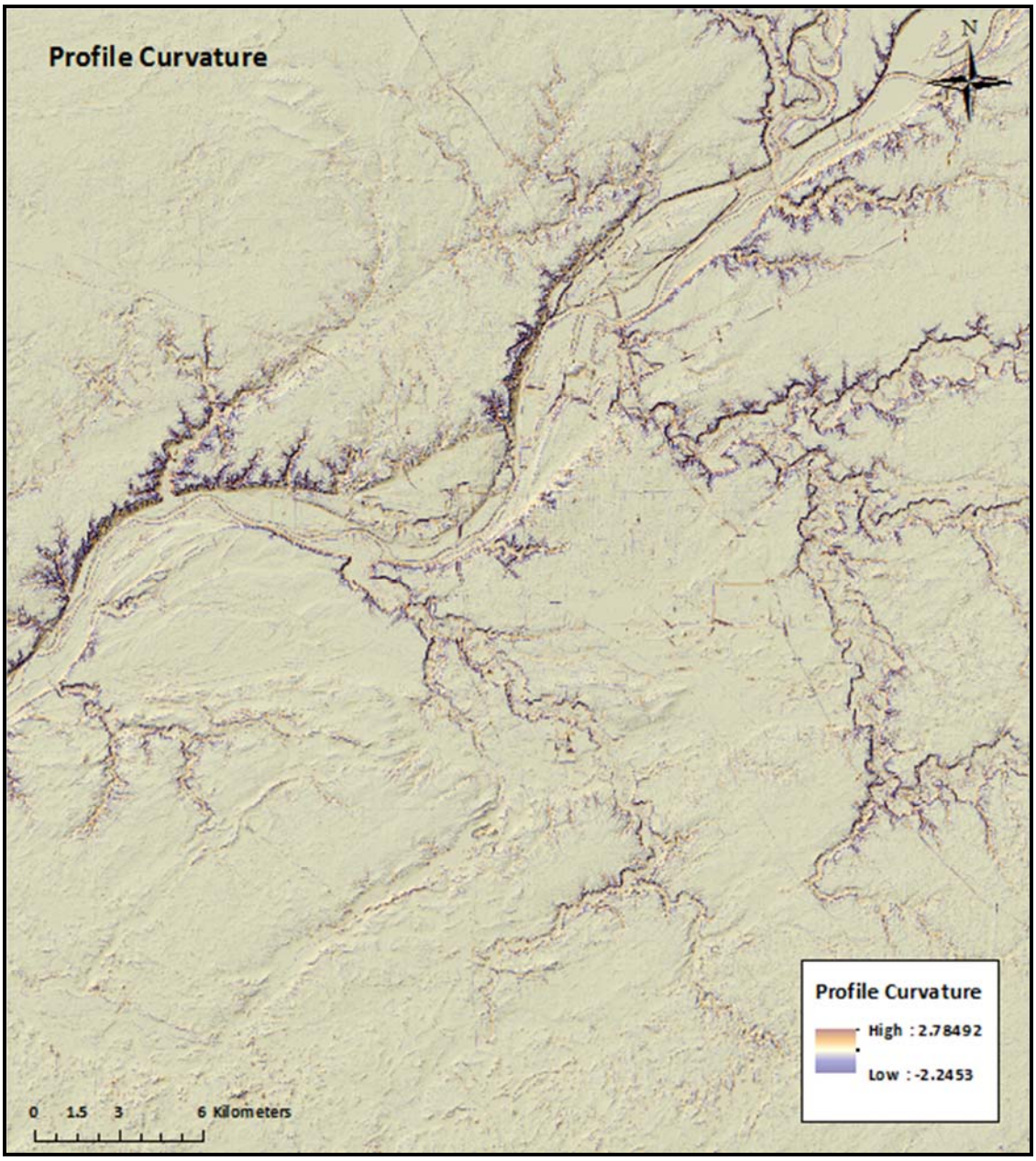


Figure 3.8f. Profile curvature for Tippecanoe County, Indiana

3.2.4.2 → Geomorphons

Geomorphons (Jasiewicz & Stepinski, 2013), a GRASS-GIS add-on, was used to define landforms in Tippecanoe County. Unlike other landform classification systems, geomorphons are not derived from primary or secondary terrain derivatives. Instead, Geomorphons is a pattern recognition, image analysis software approach that makes use of two classic concepts, the viewshed concept and local ternary patterns, to define terrain morphology from a DEM. Geomorphons supports the observation that a soil scientist in the field classifies landforms based on topographic patterns – a form of image analysis- not geomorphometric variables.

The viewshed concept uses an algorithm to determine the difference in elevation between an origin cell and a target cell (Figure 3.9). If cells of higher elevation stand between the origin and target cells within a horizontal distance, then the target cell is excluded from that viewshed. The zenith angle (*DΦL*) is the maximum angle subtended by the origin point and any other point viewed from above the surface along an azimuth (*D*) and radial limit or search radius (*L*). The nadir angle is the maximum angle subtended by the origin point and any other point viewed from below the surface along the *D* and *L*. The viewshed is measured as the mean of the zenith and nadir angles in all eight compass directions within a user-defined radial limit from the origin point (Yokoyama, et al., 2002).

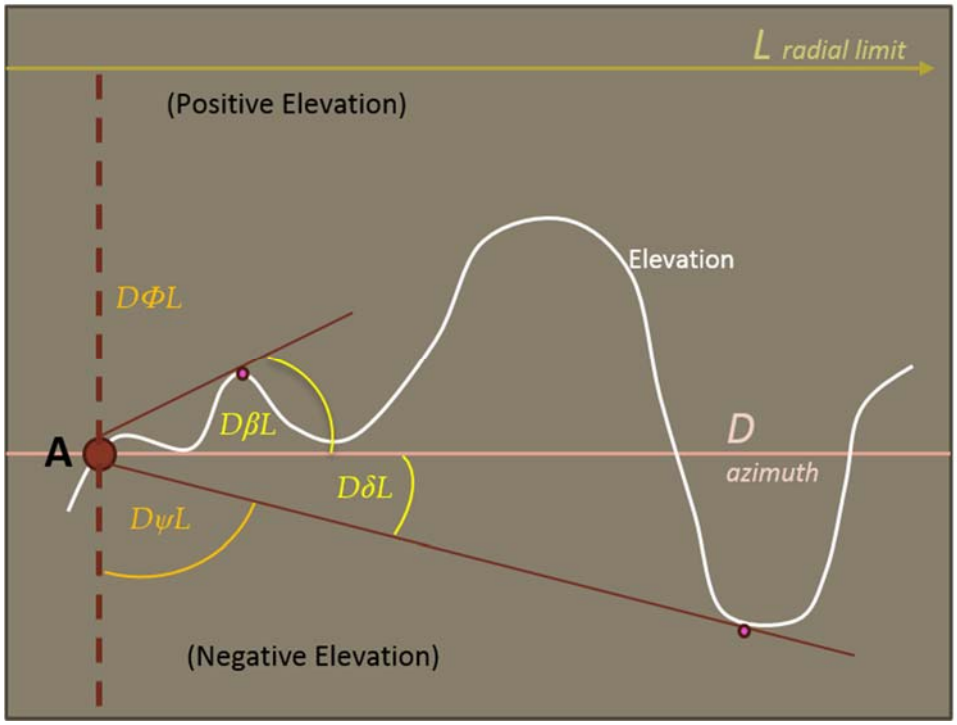


Figure 3.9. The viewshed parameters for origin cell **A** along one azimuth (D) that contains the largest elevation angle within the set. DβL is the maximum elevation angle and DδL is the minimum elevation in an azimuth-radial limit (D-L) set. The zenith angle is equal to 90^{o} - DβL and the nadir angle is equal to 90^{o} + DδL.

A ternary value- 0, 1, or -1- is derived from a comparison between the zenith and nadir angles relative to a user-defined flatness threshold (*d*):

Ternary value = 1 if: DψL - DΦL > *d*

Ternary value = 0 if: |DψL – DΦL| < *d*

Ternary value = -1 if DψL – DΦL < - *d*

The flatness threshold (*d*) is defined as the minimum allowable difference between the zenith and nadir angles in the viewshed that is considered different from the horizon (Jarek Jasiewicz & Stepinkski, 2013). This means that with a higher *d* value, the difference between the zenith and nadir angles would have to be much greater in order for the ternary value at that azimuth to be 1. A higher flatness threshold would thus serve to reduce the effect of small peaks and pits that are anomalies in the DEM.

The combination of eight ternary values, one for each azimuth, forms a ternary pattern that is reclassified into ten common landforms based on the number of translations of the ternary values (Figure 3.10). Each pixel of a DEM is thus classified as one of ten geomorphons: flat, peak, ridge, shoulder, spur, slope, pit, valley, footslope, and hollow.

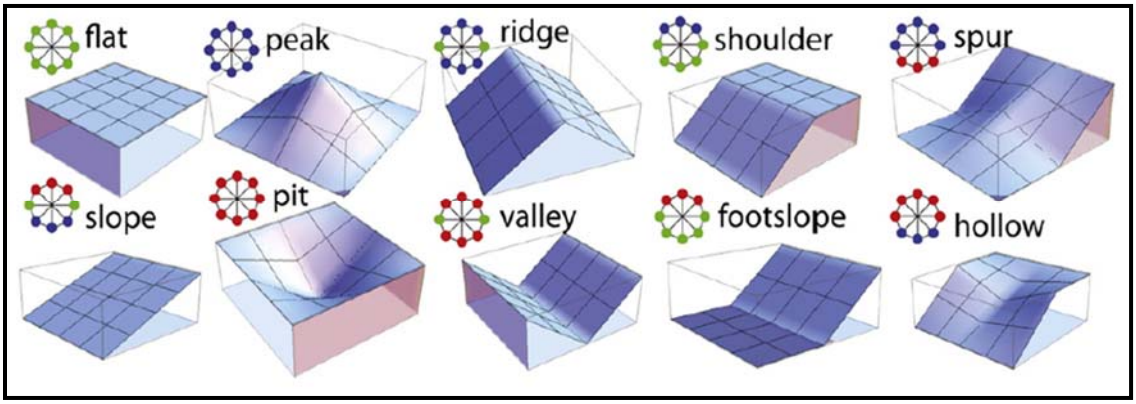


Figure 3.10. The ternary patterns and 3D morphologies of the 10 landform elements defined by Geomorphons. From[(Jasiewicz & Stepinski, 2013)]

Several iterations were run to test for different user-defined parameters. The most appropriate set of parameters was found by comparing each output geomorphons grid to a map of shaded relief and contour lines for Tippecanoe County. The best parameters were determined to be a search radius (*L*) of 600 meters, a flatness threshold (*d*) of 0.75^{o}, and a skip distance of 10 meters to reduce the impact of micro-reliefs in the DEM on the geomorphons classification. The resulting geomorphons grid is displayed in Figure 3.11.

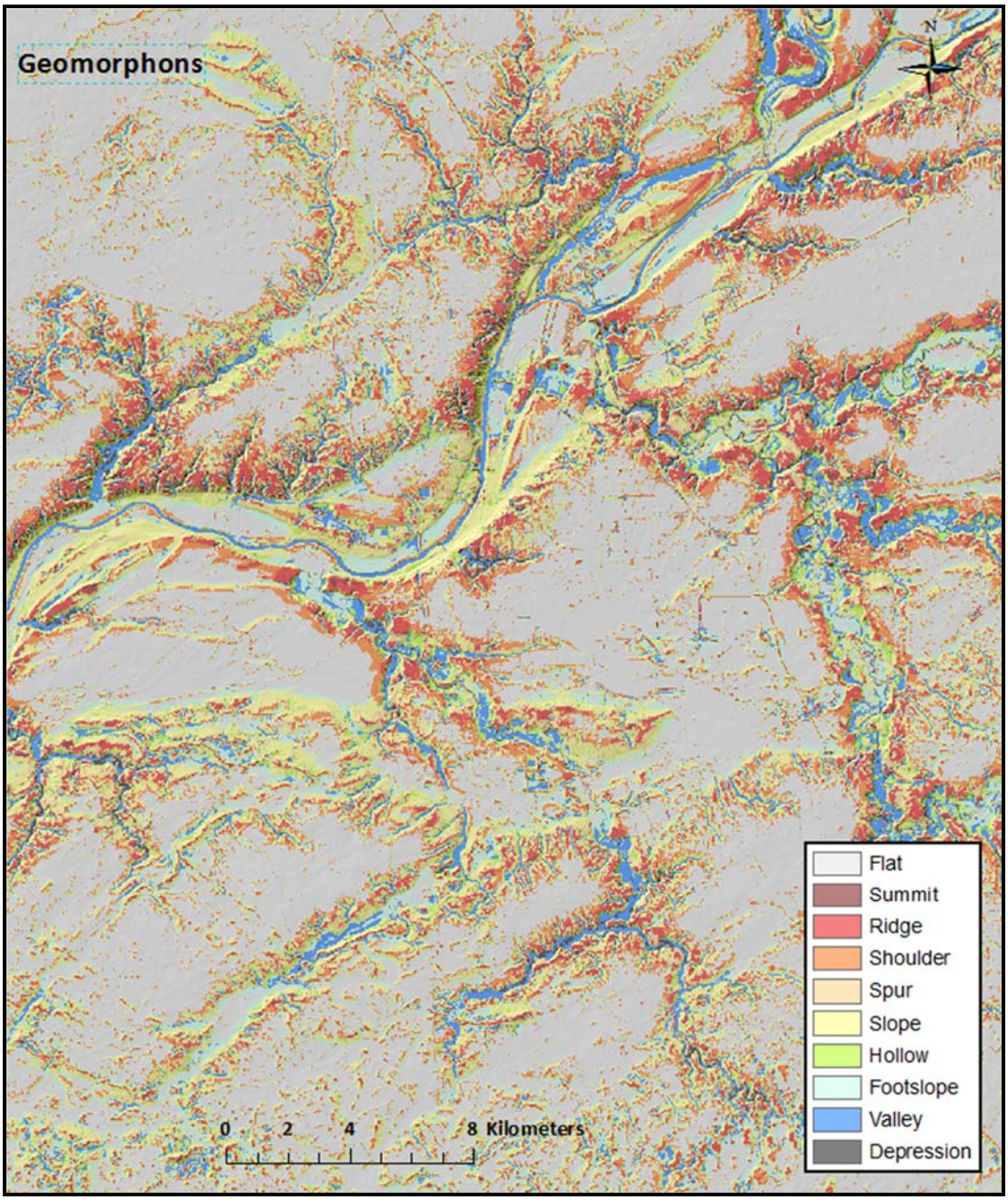


Figure 3.11. Geomorphons for Tippecanoe County, Indiana

3.2.5 Satellite Data

There are many sources of publically available satellite imagery. Several of these have been used in previous soil studies including the Advanced Very High Resolution

Radiometer (Aksoy, et al., 2009; Dobos et al., 2002, 2001; Lozano-Garcia, et al., 1995; Maselli, 2004; Senay & Elliott, 2000; Singh et al., 2006), the Moderate-Resolution Spectoradiometer (Du-blayo et al., 2008; Gu et al., 2008; Tsvetsinskaya et al., 2002; Wang, et al., 2007), and Landsat Thematic Mapper and Enhanced Thematic Mapper (Aksoy et al., 2009; Ali & Kotb, 2010; Boettinger et al., 2008; Demattê et al., 2009; Dewitte, et al., 2012), mostly to quantify soil moisture, and on soils with little vegetation cover. This study uses Landsat 8 data because it is a relatively new satellite and data is collected at a 30-meter and 15-meter resolution, similar to Landsat 7 but without the scan line failure and associated issues.

The Landsat 8 satellite was launched February 13, 2013 and achieved operational orbit on April 11, 2013. It is equipped with two sensors: the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). The OLI provides the multispectral bands and panchromatic bands and the TIRS provides two thermal bands (U.S. Geological Survey, 2014b). Table 3.4 lists the band designations for the Landsat 8 satellite.

Table 3.4. Band designations for Landsat 8.[Modified from (U.S. Geological Survey, 2014a)]

Landsat 8 OLI and TIRS		
Bands	**Wavelength (micrometers)**	**Resolution (meters)**
Band 1 – Coastal aerosol	0.43 – 0.45	30
Band 2 - Blue	0.45 – 0.51	30
Band 3 - Green	0.53 – 0.59	30
Band 4 - Red	0.64 – 0.67	30
Band 5 – Near Infrared (NIR)	0.85 – 0.88	30
Band 6 – Shortwave Infrared 1 (SWIR 1)	1.57 – 1.65	30
Band 7 – SWIR 2	2.11 – 2.29	30
Band 8 – Panchromatic	0.50 – 0.68	15
Band 9 - Cirrus	1.36 – 1.38	30
Band 10 – Thermal Infrared 1 (TIRS 1)	10.60 – 11.19	100 resampled to 30
Band 11 – TIRS 2	11.50 – 12.51	100 resampled to 30

Landsat 8 imagery was downloaded from the USGS EarthExplorer website (http://earthexplorer.usgs.gov/). A single satellite image was selected based on results of an advanced search with criteria suitable for the purpose of this study. The search criteria were for an image covering the full extent of Tippecanoe County in a single swath, with less than 10% cloud cover, with a high image quality rating, and collected in the months June, July, and August because vegetation density is greatest during this time and vegetation greenness density can be used as a marker for subsurface soil variability. The image (Figure 3.12) selected has the following characteristics:

Acquisition date: June 25, 2013

Image number: LC80220322013176LGN00

Swath: Path 022 Row032

Coverage: Full scene; day; 9% cloud cover

Processed: Level 1T

Before using the image for analysis, it was pre-processed using the ENVI version 5.1 software (Exelis Visual Information Solutions, 2014) to enhance the image and reduce atmospheric effects. A radiometric calibration (Exelis Visual Information Solutions Inc., 2014) was performed to calibrate the image data to radiance. The following equation was used to compute radiance:

$$L_\lambda = Gain * Pixel\ value + Offset$$

Where:

L_λ = Radiance in units of W/ (m^2 * sr * µm)

The gain and offset for each band is taken from the image metadata. The image was calibrated for input into the Fast-Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH®) module, so the properties entered were: interleave as BIL, data type as floating-point, and a scale factor of 0.1 to scale the output to units of µW/ (cm^2 * sr * nm).

Figure 3.12. Pan-sharpened Landsat 8 image of Tippecanoe County in natural color bands.

FLAASH® was used to retrieve the spectral reflectance from the Landsat image. The FLAASH atmospheric correction takes into account the scene visibility, water vapor and distribution of aerosols to remove the influence of the atmosphere before calculating reflectance. It differs from other atmospheric correction models in that it incorporates the MODTRAN® model's information for atmosphere and aerosol types and it corrects for pixel mixing from surface scattering of the radiance. FLAASH uses the following equations (Exelis Visual Information Solutions, 2014):

$$L = \left(\frac{A\rho}{1-\rho e S}\right) + \left(\frac{B\rho e}{1-\rho e S}\right) + L_a \qquad \text{(Equation 1)}$$

Where:

ρ = the pixel surface reflectance

ρ_e = the average surface reflectance for the pixel and surrounding area that accounts for spectral mixing

S = the spherical albedo of the atmosphere

L_a = the radiance back scattered by the atmosphere

The values for *A*, *B*, *S*, and L_a are determined from MODTRAN calculations that incorporate the viewing and solar angles and the mean surface elevation. These values are highly dependent on the water vapor column amount which is estimated for every pixel (Yale Center for Earth Observation, 2013).

A pixel surface reflectance is estimated in all of the sensor channels using Equation 1. The final output is a spatially averaged reflectance, ρ_e, estimated from a spatially averaged radiance, L_e (Equation 2):

$$L_e = \left(\frac{(A + B)\rho e}{1 - \rho e S}\right) + L_a \quad \text{(Equation 2)}$$

3.2.5.1 → Band Combinations and Vegetation Indices

Different image band combinations were explored in ENVI to view the image and to help determine which bands had the best quality. Table 3.5 provides the band combinations and the resulting images are displayed in Figures 3.13a – e.

Table 3.5. Band combinations for Landsat 8. [Modified from (Butler, 2013)]

Band Combination	Band Names	RGB
Natural Color	Red Green Blue	4 3 2
False Color (urban)	SWIR 2 SWIR1 NIR	7 6 4
Color Infrared (vegetation)	NIR Red Green	5 4 3
Agriculture	SWIR 1 NIR Blue	6 5 2
Atmospheric Penetration	SWIR 2 SWIR 1 NIR	7 6 5
Healthy Vegetation	NIR SWIR 1 Blue	5 6 2
Land/Water	NIR SWIR 1 Red	5 6 4
Natural with Atmospheric Removal	SWIR 2 NIR Green	7 5 3
Shortwave Infrared	SWIR 2 NIR Red	7 5 4
Vegetation Analysis	SWIR 1 NIR Red	6 5 4

The false color image on the right in Figure 3.13a has the band combination SWIR2-NIR-green. The shortwave infrared 2 band highlights vegetation in moist soils and the NIR highlights plant greenness. The lighter green areas are areas with greater plant vigor. Figure 3.13 b highlights green vegetation and helps identify urban areas. Brighter blues are greener vegetation with higher biomass and very bright yellow areas are urban structures. Figure 3.13c helps detect bodies of water. Water appears black. The image on the left in Figure 3.13d is an infrared image that accentuates areas that are vegetated. It helps to differentiate areas that are not as green, urban areas in teal and water bodies in black. The image on the right in Figure 3.13d is an SWIR 1-NIR-blue image and it helps to distinguish between areas that are more vegetated (bright green)

from those with little to no vegetation (reds, browns, whites). The images in Figure 3.13e both serve to highlight areas of healthy vegetation, water bodies, and different types of urban areas. The SWIR 1-NIR-red image on the left highlights healthy vegetation in bright green, bare soil in pink, different urban areas in purple, and water bodies in black. In the image on the right, healthy vegetation is bright red, water is black, urban areas are a light blue, and land with little vegetation is green.

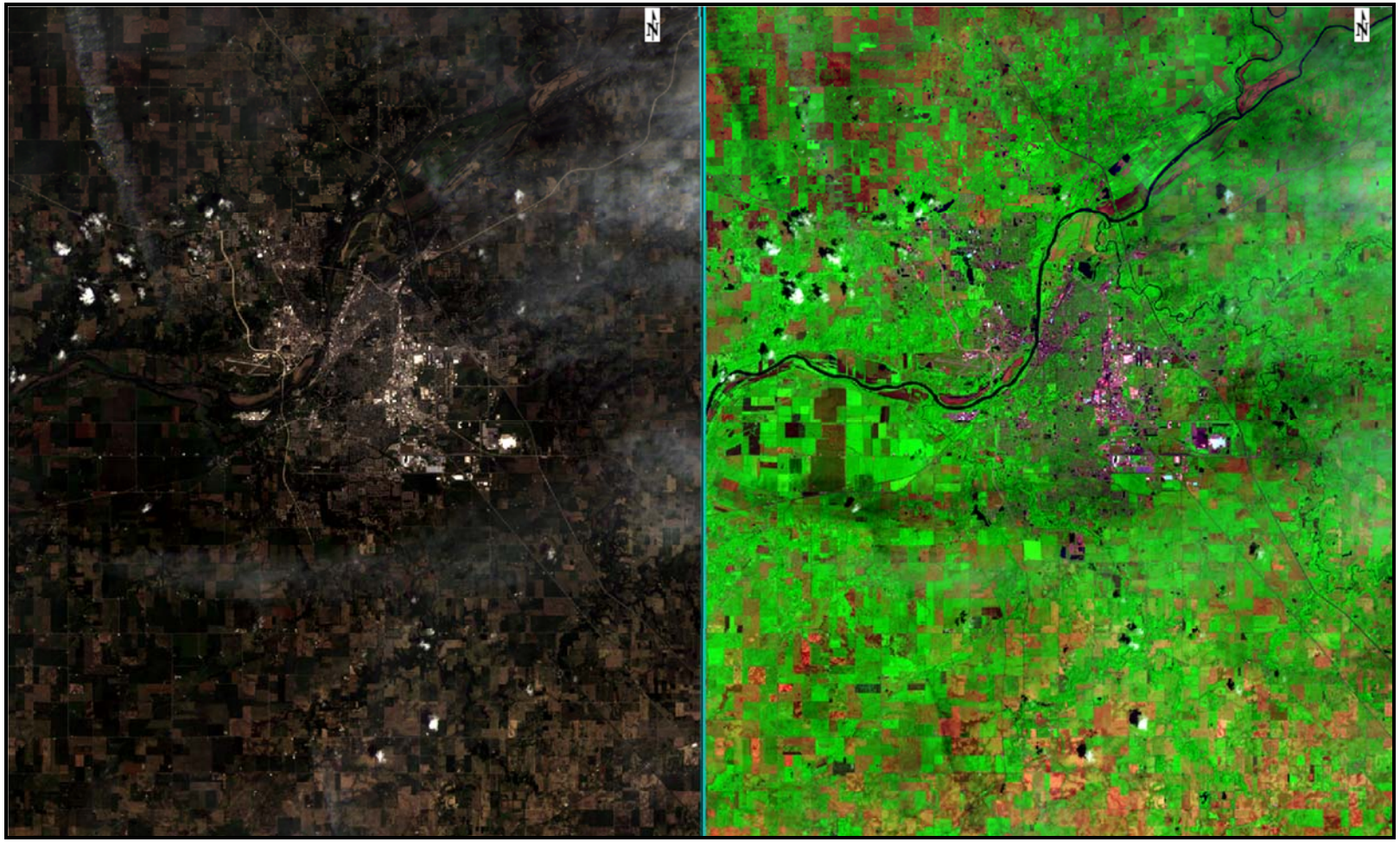


Figure 3.13a. Landsat 8 natural color image (left) and natural color with atmospheric removal image (right)

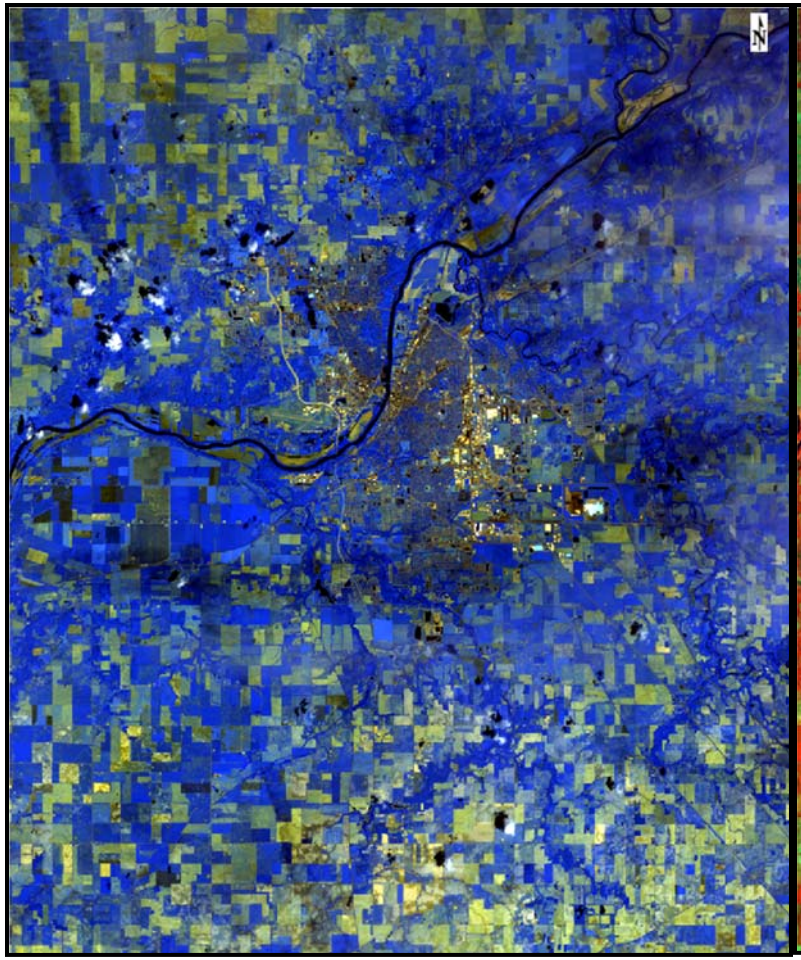


Figure 3.13b. Landsat 8 false color (urban) image

Figure 3.13c. Landsat 8 land and water image

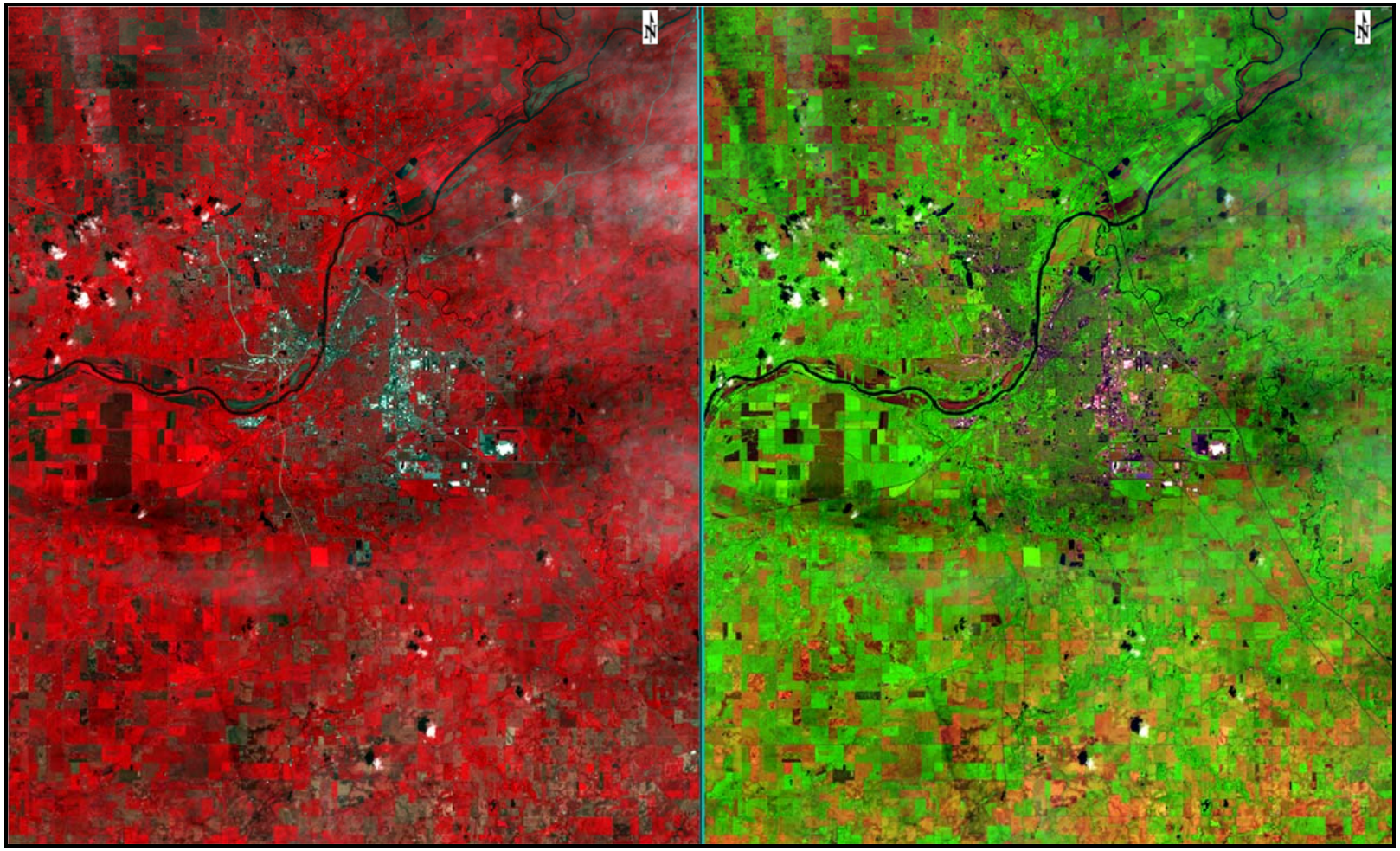


Figure 3.13d. Landsat 8 color infrared image (left) and agriculture image (right)

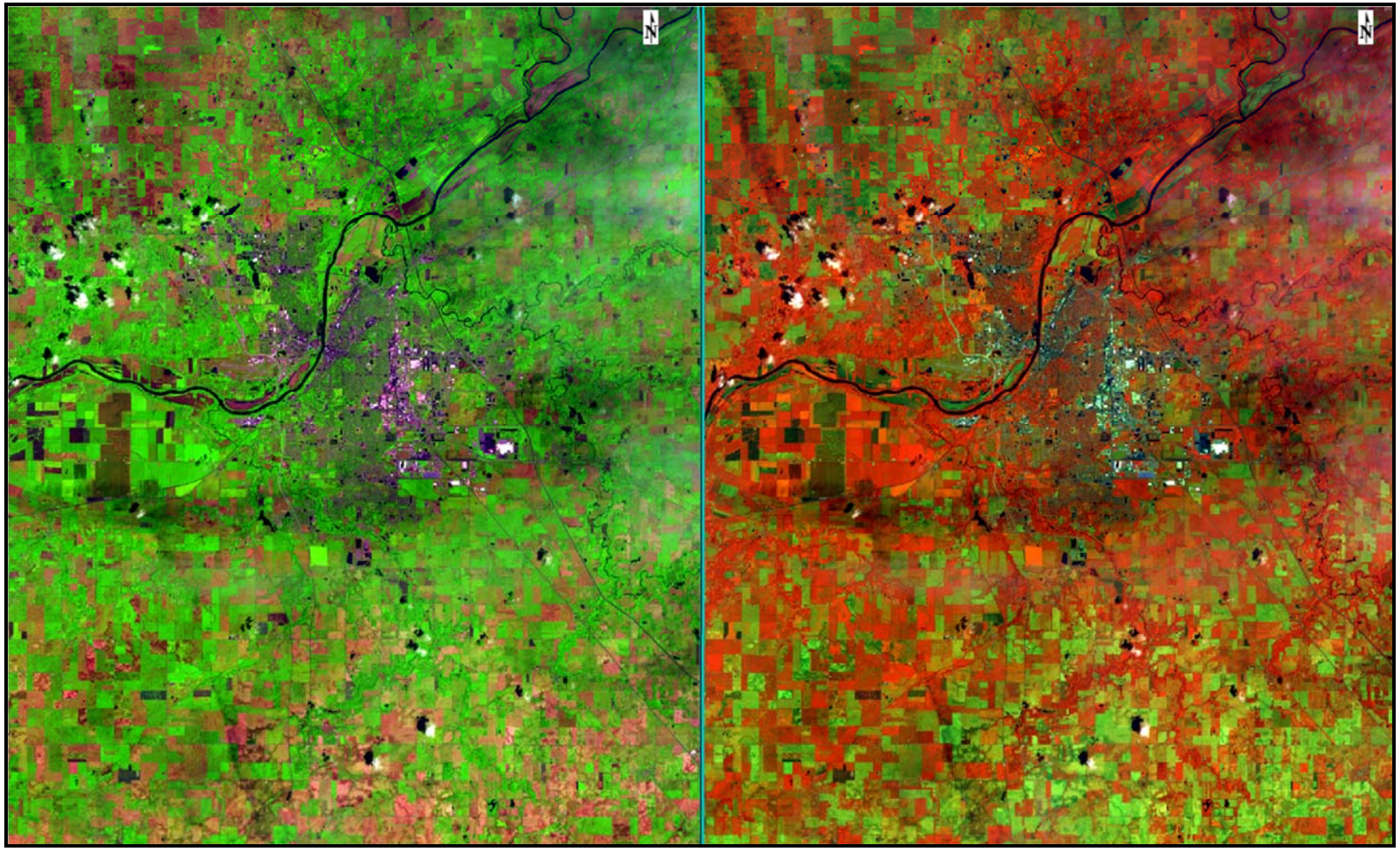


Figure 3.13e. Landsat 8 vegetation analysis image (left) and healthy vegetation image (right)

Various vegetation indices were computed for the Landsat 8 image from the FLAASH reflectance files in the ENVI version 5.1 software. A vegetation index is a ratio of a band where vegetation is bright to a band where it is dark. Vegetation indices exploit the intrinsic characteristics of plant cells in order to highlight and quantify green vegetation density.

Cells in plant leaves are effective at scattering light and appear bright in the NIR range between 1300 nm and 700 nm because of the high contrast between their water-rich interior and their intercellular air space. Vegetation appears dark in the visible spectral range because there is a high absorption of the plant pigments. However, pigments are least absorptive around green in the visible spectrum and this is why they appear green to us. At wavelengths longer than 1300 nm, vegetation appears dark due to absorption by leaf water (Ray, 2014). Even though vegetation is more reflective in the green band than in the red band, a study by Tucker, 1979 showed that NIR-red combinations were preferable and they have become the standard.

The indices that were calculated are NDVI, soil-adjusted vegetation index (SAVI), and a measure of fractional vegetative cover (FVC). The NDVI image, shown in Figure 3.14, was generated using the NDVI tool which uses this standard formula:

$$NDVI = [(NIR - Red)/ (NIR+Red)]$$

Where: *NIR* is the near-infrared band and *Red* is the red band. The output image had an index range of -0.5 to +0.9516 which falls within the acceptable range of -1 to +1.

The SAVI image, shown in Figure 3.15, was produced using the Band Math tool and this expression adopted from (Huete, 1988):

$$SAVI = [(NIR - Red)/ (NIR+Red+ L)] * (1 + L)$$

Where:

NIR is the near-infrared band and *Red* is the red band. The correction factor is *L* and it ranges from 0 for very high vegetation density to 1 for little to no vegetation. The standard value of 0.5 was used. The resulting SAVI index range was -0.7489 to 1.4271.

The SAVI attempts to reduce soil noise in areas where the vegetation cover is low. It takes into consideration the effect of bare soil on surface reflectance and the vegetation. Like the NDVI formula, the SAVI assumes that bare soil forms a line somewhere between the NIR and red bands in the spectrum. Along this line, there is no vegetation and divergent from this line in space are isovegetation lines or lines of equal vegetation. The SAVI also recognizes that soil has varying reflectance due to soil physical and chemical differences and soil moisture so it introduces an adjustment factor to correct for this (Ray, 2014).

The FVC, displayed in Figure 3.16, is an estimated value of the percent of vegetation in a pixel. It is computed by re-normalizing the NDVI (Zeng et al., 2000):

$$FVC = [(NDVI - min\ NDVI)/ (max\ NDVI - min\ NDVI)] * 100$$

Where:

NIR is the near-infrared band and *Red* is the red band. The *min NDVI* and *max NDVI* are the lowest and highest values of the previously-computed NDVI, respectively. The valid range for FVC is 0% for no vegetation and 100% for full vegetative cover.

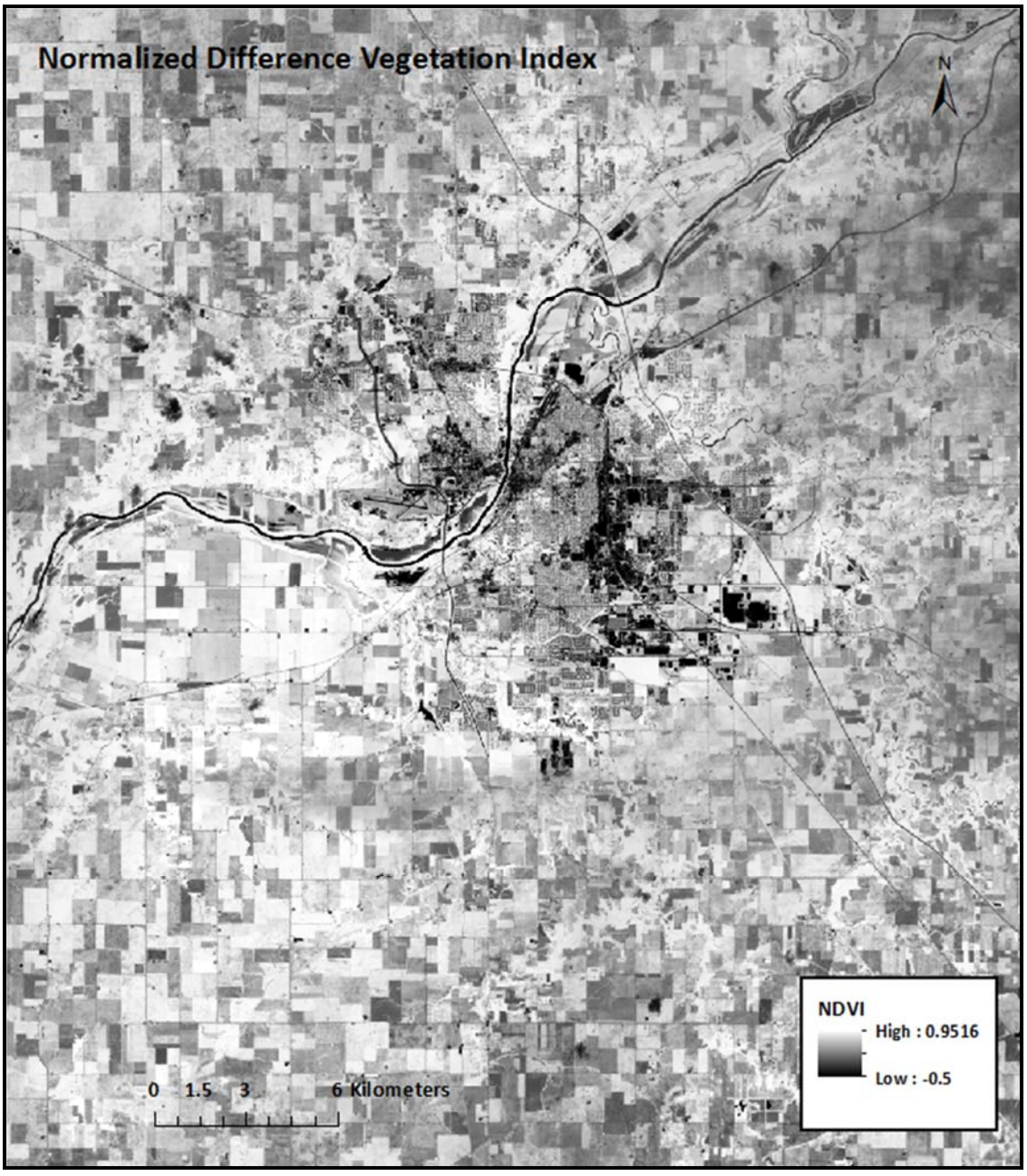


Figure 3.14. Normalized difference vegetation index (NDVI) for Tippecanoe County, Indiana

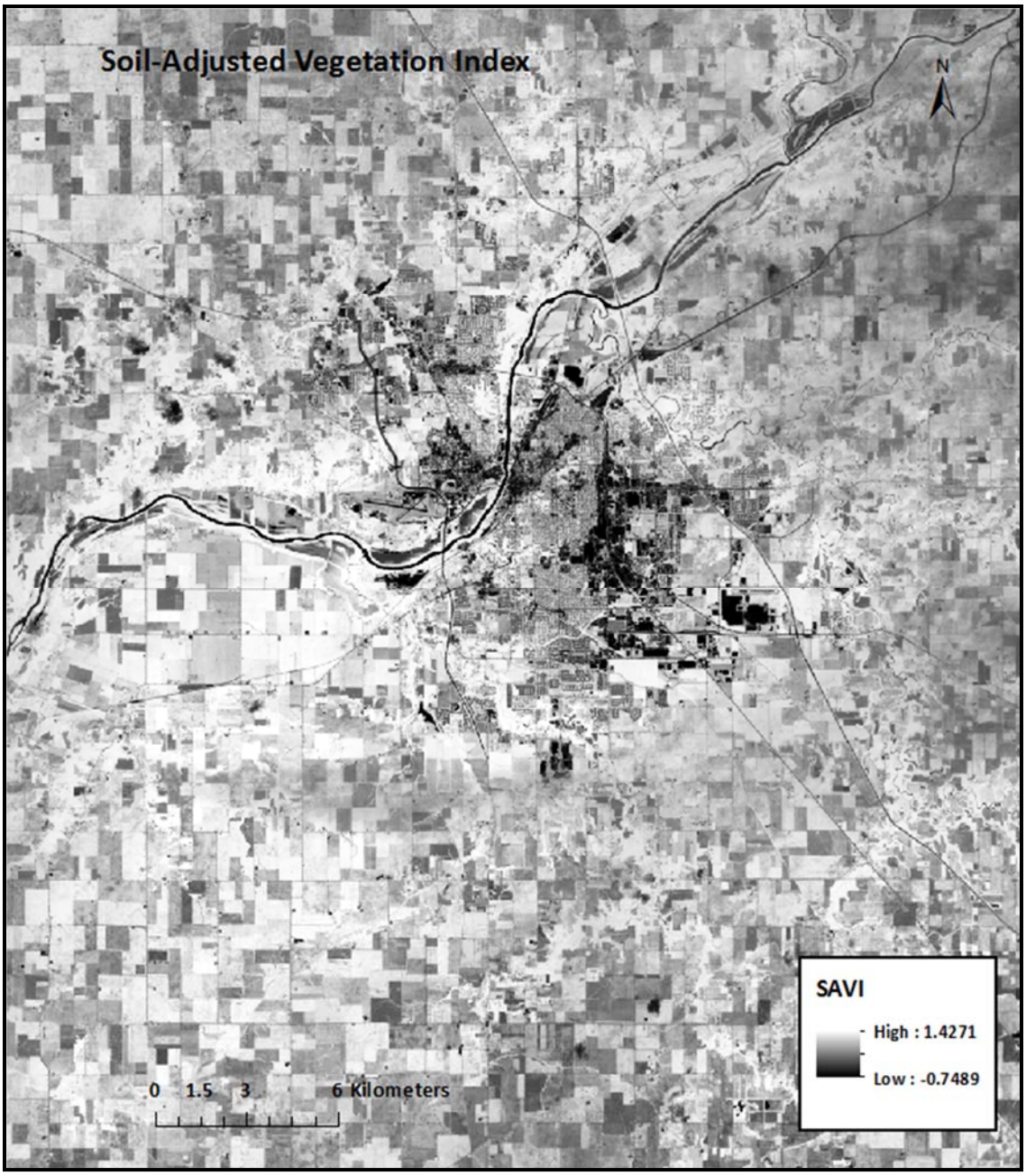


Figure 3.15. Soil –adjusted vegetation index (SAVI) for Tippecanoe County, Indiana

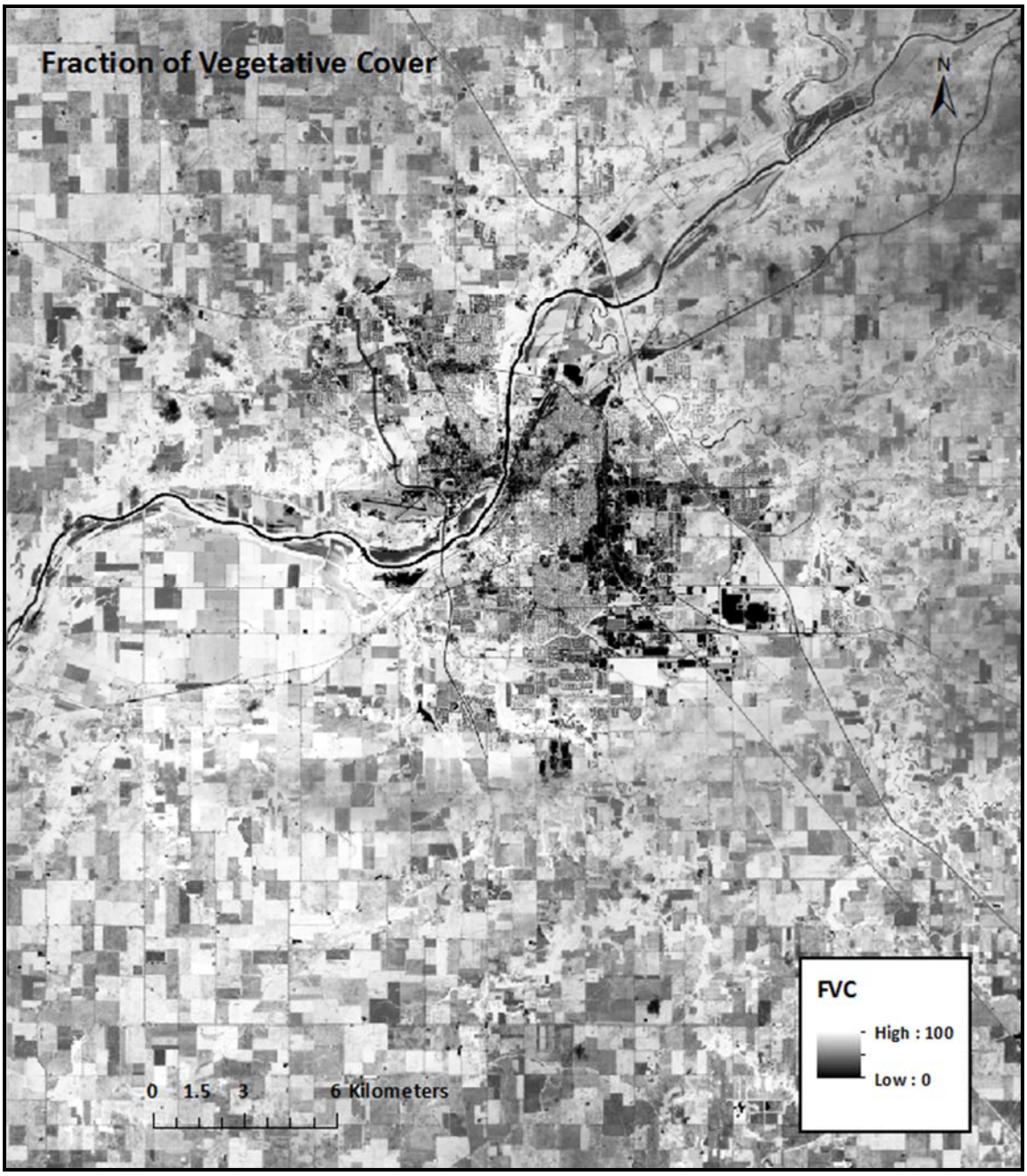


Figure 3.16. Fractional Vegetative Cover for Tippecanoe County, Indiana

3.3 Methods

3.3.1 Principal Component Analysis

After the satellite image was pre-processed and an atmospheric correction was performed, a principal component analysis (PCA) transformation was conducted to reduce the number of bands and unnecessary data that could cloud the classification. The PCA was generated in the ENVI version 5.1 software using a covariance matrix (Exelis Visual Information Solutions, 2009).

Spectral bands of Landsat 8 images and other multispectral images are not completely independent of each other. The bands are correlated and so they must be transformed using coefficients derived from the covariance matrix of the original image. The result is an uncorrelated set of output bands and a decrease in the dimensionality of a data set (Lillesand et al., 2004). Three principal components were produced from this PCA. They are displayed in Figures 3.17 a-d.

3.3.2 Satellite and Digital Elevation Data Integration

After generating the terrain derivatives from the DEM and the measures of vegetation density and principle components from the satellite data, the files were all resampled to 30-meter spatial resolution and NAD 83 UTM zone 16N map projection. They were all clipped to the boundary of Tippecanoe County and exported as ESRI GRID files. Histograms of all input files were analyzed using the Spatial Analyst toolbar in ArcGIS and layers that were not normally-distributed were removed. The data ranges of the remaining layers were normalized by entering this formula into Raster Calculator:

$$Z = \{[(X - \text{old min}) * (\text{new max} - \text{new min})]/ (\text{old max} - \text{old min})\} + \text{new min}$$

Where:

Z = the output raster with new data ranges

X = the input raster

new max = the maximum value assigned for output rasters, or ‘10’ in this case

new min = the minimum value assigned for the output rasters, or ‘1’ in this case

old max = the maximum value of the input raster

old min = the minimum value of the input raster

After normalizing the data, a copy of the layers was clipped to the smaller study area for classification and comparison to SSURGO map units. The band statistics were computed and the resulting covariance and correlation matrices were studied for the two data sets to determine if any layers were correlated. The NDVI and FVC layers were removed from both data sets because these were found to be highly correlated with the SAVI. This is due to the fact that all three are derivations of the same two NIR and red bands. The 3 principal components, SAVI, geomorphons, relative elevation, TWI, VDCN, MRVBF, and profile curvature were the only files used for the classification and comparison to STATSGO. Two additional layers were removed from the SSURGO dataset, the VDCN and MRVBF because they were highly correlated within the smaller extent of the SSURGO map area of interest. Each of these layer sets was integrated into a new data set as a band collection file.

3.3.3 Image Classification

Image classification was used to categorize all pixels in Tippecanoe County into soil class units. Unsupervised and supervised classification methods were applied in this study and performed using version 10.2.2 of ArcGIS. The Iterative Self-Organizing Data Analysis (ISODATA) technique was used for unsupervised classification. The Gaussian Maximum Likelihood classifier was used for supervised classification and STATSGO and SSURGO soil polygon maps were used as the training polygons for this classification.

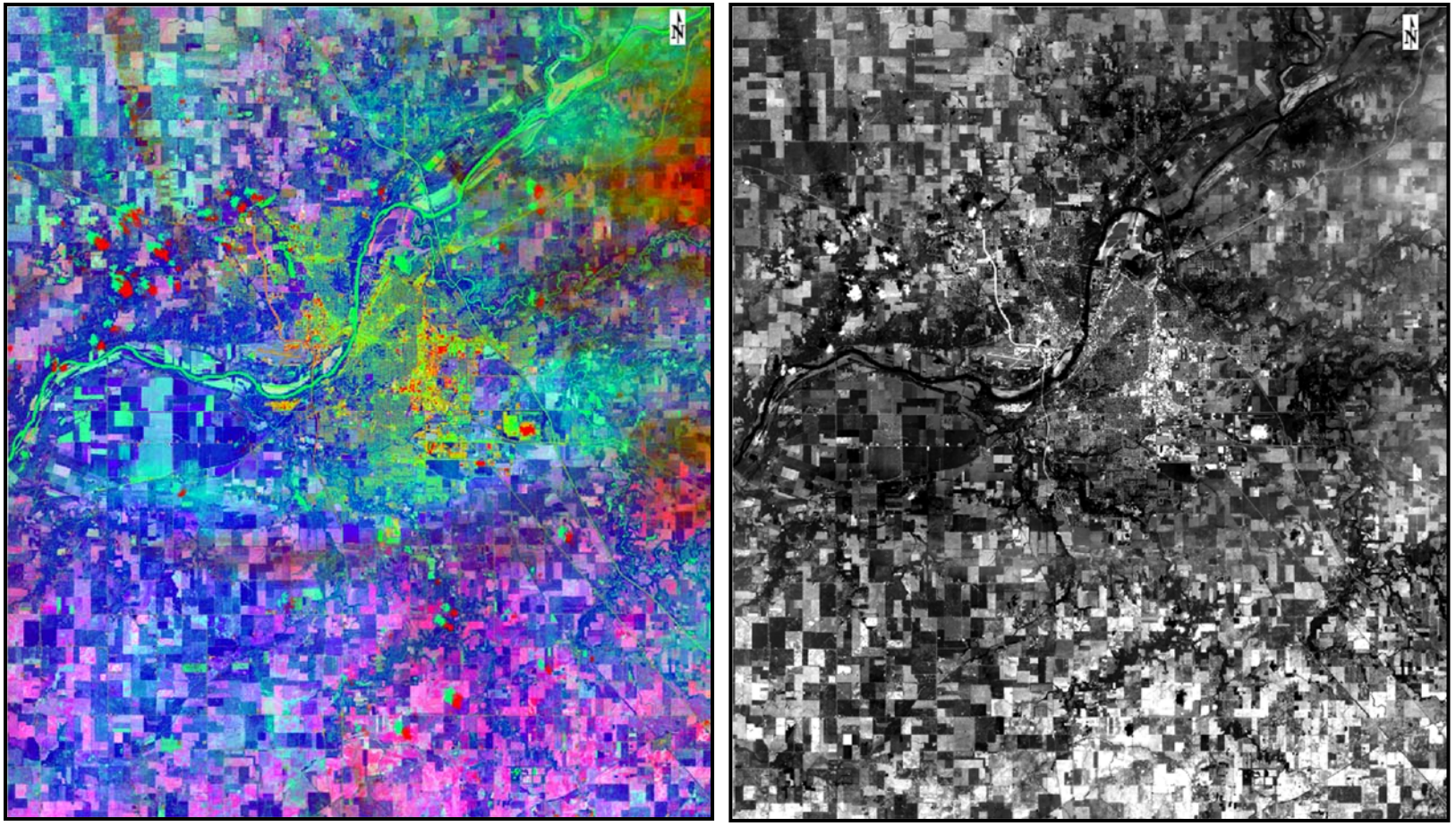


Figure 3.17a. Principal component analysis output aggregate image (left) and 3.17b. First principal component image for Tippecanoe County, Indiana (right).

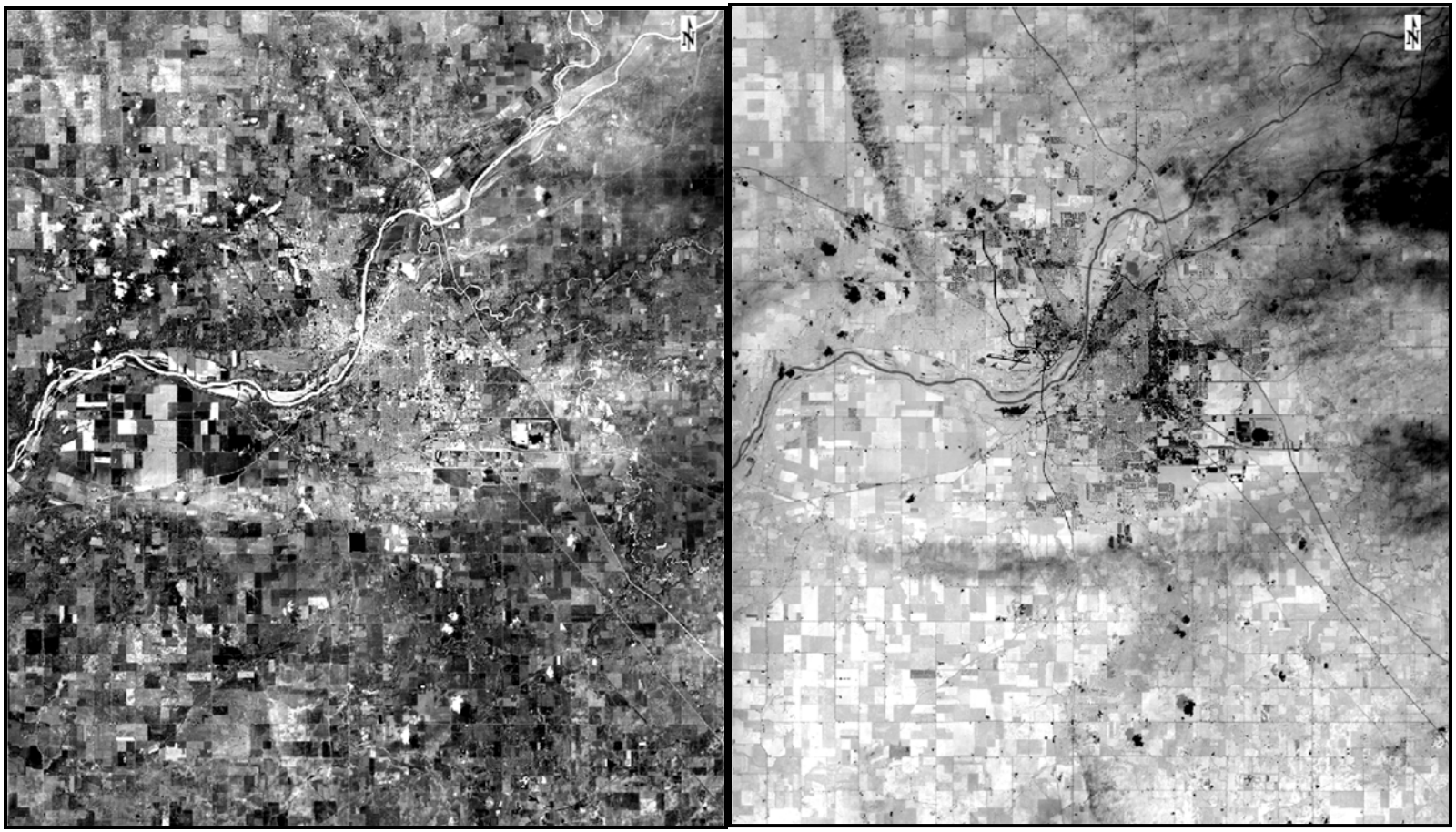


Figure 3.17c. Second principal component image (left) and 3.17d. Third principal component image for Tippecanoe County, Indiana (right).

3.3.3.1 → Unsupervised Classification

Unsupervised classification works through algorithms that aggregate the unknown image pixels into classes based on clusters of the pixel values. The assumption behind this technique is that pixel values within the same class should be relatively closer in the measurement space than those from different classes. Because of the nature of unsupervised classification, the identity of the resulting classes is not immediately known. Unsupervised classification is useful because it can help highlight classes that the analyst may not have considered in the supervised training set or classifier.

In this study, the ISODATA technique was used to estimate the pixel value separability of the band collection files for Tippecanoe County and for the small area of interest. The ISODATA repeatedly classifies the image pixels and allows the number of clusters to change from one iteration to the next. Statistics are calculated after every iteration to determine if the clusters will be merged, split, or deleted. Clusters are merged if the distance between the mean points of two clusters is less than a specified minimum value. A cluster will be split if its standard deviation is greater than a predefined maximum value. Clusters are deleted when they contain fewer than the minimum number of pixels allowed per class (Lillesand et al., 2004). The classification is completed when the cluster statistics no longer change or the maximum number of allowed iterations is reached. The ISODATA unsupervised classification was performed multiple times on the two data sets. Each run used a different set of input parameters, listed in Table 3.6.

Table 3.6. Input parameters for ISODATA classification

# of classes	Iterations	Minimum Class Size	Sample Interval
80	20	4	10
80	100	4	10
80	1000	4	10
80	1000	4	5
80	1000	4	3
80	1000	4	2
80	1000	4	1

3.3.3.2 → Supervised Classification

A supervised classification uses an algorithm to assign a category label to the unclassified pixels. When performing a supervised classification, an analyst provides numerical descriptors known as "training samples" for the different categories. The pixels are labeled with the category that their numerical values most closely resemble. A Gaussian Maximum Likelihood classifier was used for supervised classification. The assumptions for this technique are that the pixel set for each category of the training data is normally distributed and that a category can be characterized by its mean and covariance matrix. Under these assumptions, a statistical class probability is calculated for each pixel to determine its membership to a category. An equal *a priori* probability weighting was used and so a pixel was assigned the class for which it had the highest probability of membership (ESRI, 2014) using the following equation:

$$f_i(X) = \frac{1}{((2\pi)^{p/2})|\Sigma_i|^{1/2}} \; exp\left\{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right\}$$

From (Crawford, 2013)

Where:

f_i = probability density function of class i

p = number of layers or bands

x = p * 1 observation vector

μ_i = mean of class *i*, p * 1 vector

Σ_i = covariance of class *i*, p * p matrix.

3.3.4 Accuracy Assessment

The accuracy of the results obtained from the supervised classification was assessed. Generally, a quantitative measure of the accuracy is achieved by selecting a subset of pixels from the raster thematic map produced from the image classification and checking the pixel class values against those of the reference data. The percentage of pixels that agree and disagree with the class values of the reference or ground truth data is calculated this way. These results are often presented as a table known as an "error matrix" or "confusion matrix" and an overall accuracy can be estimated.

To examine whether or not the classified pixel class values agreed with the same pixels in the reference map, the classified thematic raster for the study area was overlaid with the reference map for that study area. In order to check for matches in the class values for the pixels, the reference feature map layers were converted to raster type map layers. Then, the classified raster and the reference raster were combined to display the pixel values for each pixel pair. The accuracy assessment was performed by using an error matrix to calculate the user's and producer's accuracy, overall model accuracy and Cohen's kappa coefficient.

3.4 Results

Results of the unsupervised and supervised classification will be presented for the extent of Tippecanoe County and the small area of interest separately. Only the best classified images from the unsupervised and supervised method will be presented. Visual comparison with the reference file and the number of resulting classes were used as guidelines for selecting the best results from the unsupervised classification. The

values for overall accuracy and kappa coefficient were used to select the best results from the supervised classification. The results of the soil association map adapted from STATSGO will be presented first. This study integrated five terrain derivatives and geomorphons from a LiDAR DEM and Landsat 8 imagery data for Tippecanoe County and used the STATSGO association map unit polygon file as the 'ground truth' data.

The results of the soil consociations and complexes map adapted from SSURGO will be presented next. This study integrated three terrain derivatives and geomorphons from a LiDAR DEM and Landsat 8 imagery data for a small area of interest in the northern part of Tippecanoe County. It used the SSURGO complex and consociation map units polygon file as the 'ground truth' data. Different layer combinations were used for supervised classification for each of these studies. The layers used will be listed either in the map title or in the figure description of the results maps and in the results tables.

3.4.1 Soil Associations Derived from STATSGO

This study integrated five terrain derivatives, geomorphons, and Landsat 8 data in order to classify STATSGO soil associations in Tippecanoe County. Five different combinations of satellite and DEM data were used for the classification. The combinations are listed in Table 3.7.

Table 3.7. Description of data sets used for supervised classification of Tippecanoe County.

Combination of Satellite and DEM data	Code Used for Dataset
3 Principal Component Bands from Landsat 8, SAVI, Geomorphons, Relative Elevation, TWI, VDCN, MRVBF, Profile Curvature	3PC+SAVI+G+RE+TWI+VDCN+MRVBF+PrCv
3 Principal Component Bands from Landsat 8, SAVI, Relative Elevation, TWI, VDCN, MRVBF, Profile Curvature	3PC+SAVI+RE+TWI+VDCN+MRVBF+PrCv
Geomorphons, Relative Elevation, TWI, VDCN, MRVBF, Profile Curvature	G+RE+TWI+VDCN+MRVBF+PrCv
Relative Elevation, TWI, VDCN, MRVBF, Profile Curvature	RE+TWI+VDCN+MRVBF+PrCv
3 Principal Component Bands from Landsat 8, SAVI	3PC+SAVI

The results of the unsupervised classification are presented in Figures 3.18a-c. The results for agreement with the full STATSGO map units file were observed for all five data set combinations and are presented in Table 3.8. In addition, the error matrices, commission and omission, user's and producer's accuracy, overall accuracy, and Cohen's kappa coefficient were calculated. These are presented in Tables 3.9a-c for the classified image with the highest accuracy (Figure 3.19).

A visual comparison with the STATSGO class raster shows that the "3PC+SAVI+G+RE+TWI+VDCN+MRVBF+PrCv" composite data set produced the best result from the unsupervised classification. The map in Figure 3.18a appears to have the highest agreement with the STATSGO reference raster for classes 1 and 7. However, this is only a qualitative, visual estimate. Results of the supervised Maximum Likelihood classification show that the same ten-layer composite data set gave higher agreement with the reference raster than the other composite data sets. Slightly lower agreement with the reference pixels was observed with the composite data set that included all but the geomorphons layer. The composite data set that only included the satellite data (3PC+SAVI) resulted in much lower accuracy than when topographic data was integrated

with the satellite data. When only topographic data was used for the classification, that model that used terrain derivatives produced higher-accuracy results than the model that used the terrain derivatives and the geomorphons layers.

The class accuracies for the classes derived from the ten-layer composite data set are presented in Table 3.9b. Overall agreement was very high for classes 2 (87%) and 11 (88%) and very low for classes 4 (10%), 8 (6%), and 13 (8%). Only five classes had at least 50% agreement; classes 2 and 11 had the highest overall agreement across all composite data sets. Conversely, classes 3, 8, and 13 had the lowest overall agreement across all composite data sets.

Table 3.8. Results of supervised Maximum Likelihood classification of the Tippecanoe County study area using different data sets

Dataset	3PC+SAVI+G+RE +TWI+VDCN+ MRVBF+PrCv	3PC+SAVI+RE +TWI+VDCN+ MRVBF+PrCv	G+RE+TWI+ VDCN+ MRVBF+PrCv	RE+TWI+VDCN+ MRVBF+PrCv	3PC+SAVI
Overall Accuracy	35.090%	34.002%	28.557%	29.138%	12.966%
Cohen's Kappa Coefficient	0.231	0.221	0.154	0.159	0.077

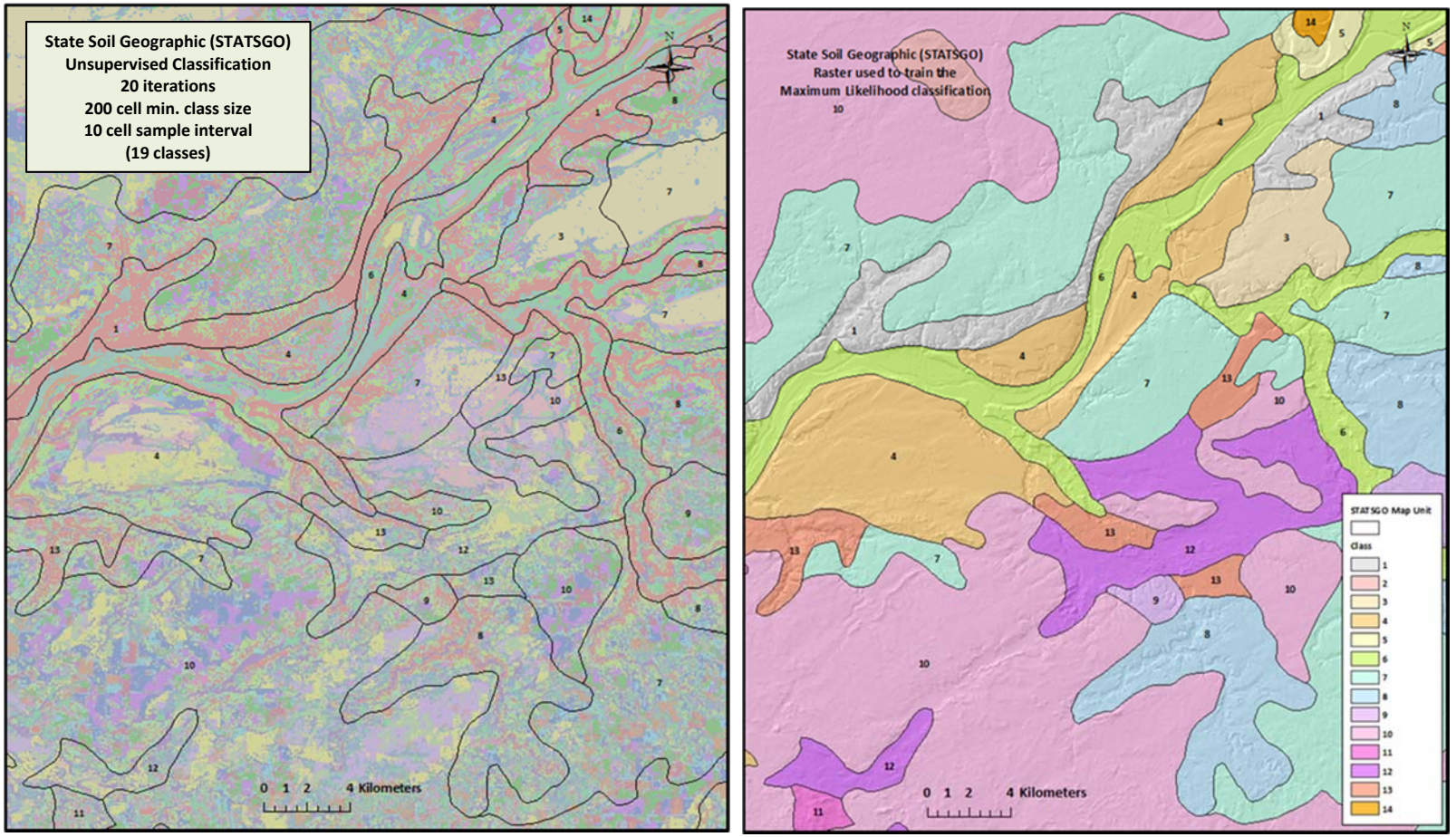


Figure 3.18a. Result of unsupervised classification using the "3PC+SAVI+G+RE+TWI+VDCN+MRVBF+PrCv" data set (left) and STATSGO classes for comparison (right).

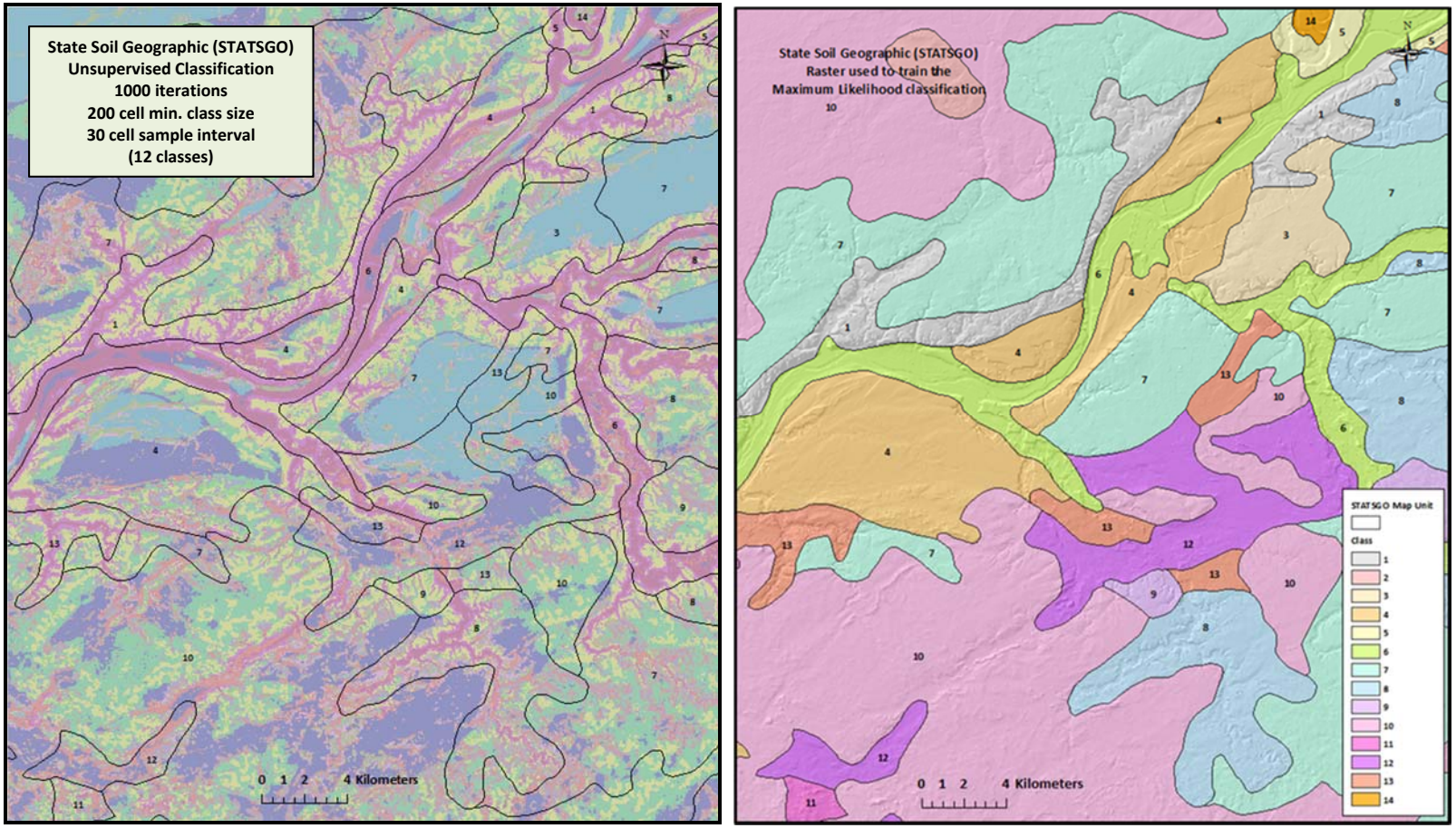


Figure 3.18b. Result of unsupervised classification using the "G+RE+TWI+VDCN+MRVBF+PrCv" data set (left) and STATSGO classes for comparison (right).

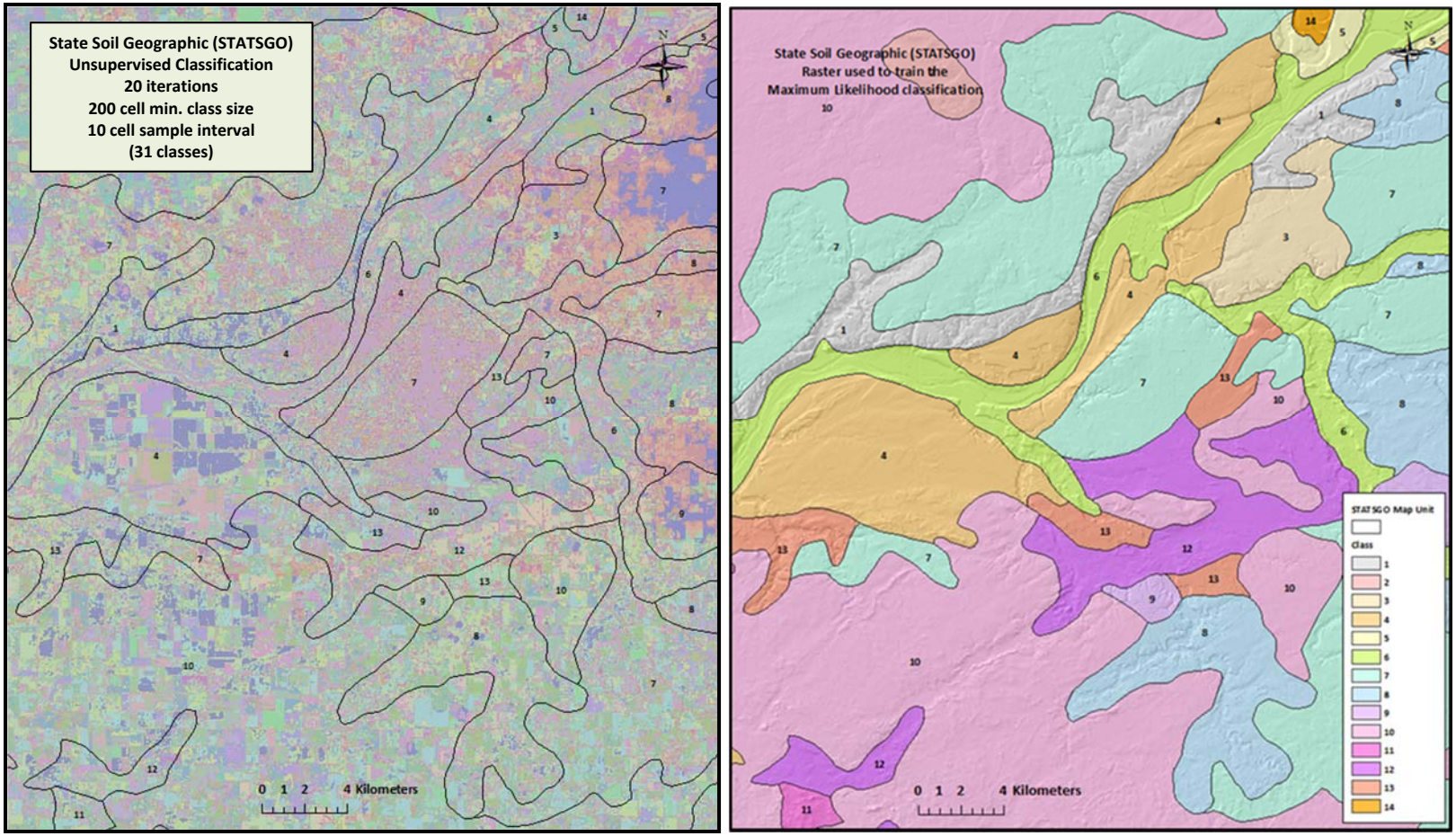


Figure 3.18c. Result of unsupervised classification using the 3PC+SAVI data set (left) and STATSGO classes for comparison (right).

Table 3.9a. Error matrix for 14 STATSGO classes derived from the integrated data set, "3PC+SAVI+G+RE+TWI+ VDCN+MRVBF+PrCv"

All 10 Layer	Refer C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	Ground Truth
Test C1	**30188**	36	2749	14146	1145	27685	14640	10342	1266	2808	10	2483	4307	213	112018
C2	3516	**10760**	2355	10379	349	2005	47886	8613	1053	50193	84	10062	3817	18	151090
C3	2592	105	**6774**	2899	863	2102	15183	7367	341	4044	0	921	608	68	43867
C4	1675	27	225	**14680**	159	5727	4559	2140	290	3069	3	2571	1236	52	36413
C5	5973	1	1633	4971	**3962**	6749	2310	5135	103	170	0	466	485	355	32313
C6	3523	0	373	7365	197	**52934**	2639	1408	206	1113	3	891	1086	145	71883
C7	1497	147	3210	4025	473	486	**39105**	7338	388	8514	32	1831	1599	3	68648
C8	1759	28	482	2419	180	1207	7206	**6983**	329	5377	44	1771	770	0	28555
C9	5085	384	3654	6557	420	2694	26862	14195	**14177**	15800	67	3366	4567	0	97828
C10	1464	608	7853	51278	797	1394	128341	34132	1913	**302381**	294	27455	10084	24	568018
C11	1724	0	135	10125	0	237	11847	5363	775	50076	**4572**	9018	2737	0	96609
C12	593	245	966	10919	7	2575	23341	3254	184	17377	74	**16284**	2872	26	78717
C13	3290	67	323	5490	32	1815	10809	4733	319	6569	35	2496	**3024**	0	39002
C14	2706	0	64	2242	1239	11814	483	1188	13	46	0	235	80	**1976**	22086
Total	65585	12408	30796	147495	9823	119424	335211	112191	21357	467537	5218	79850	37272	2880	**1447047**

Table.3.9b. Class accuracies for 14 STATSGO classes derived from the integrated data set, "3PC+SAVI+G+RE+TWI+ VDCN+MRVBF+PrCv"

All Ten Layers: 3PC+SAVI+G+RE+TWI+VDCN +MRVBF+PrCv	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
C1	**46**	0	9	10	12	23	4	9	6	1	0	3	12	7
C2	5	**87**	8	7	4	2	14	8	5	11	2	13	10	1
C3	4	1	**22**	2	9	2	5	7	2	1	0	1	2	2
C4	3	0	1	**10**	2	5	1	2	1	1	0	3	3	2
C5	9	0	5	3	**40**	6	1	5	0	0	0	1	1	12
C6	5	0	1	5	2	**44**	1	1	1	0	0	1	3	5
C7	2	1	10	3	5	0	**12**	7	2	2	1	2	4	0
C8	3	0	2	2	2	1	2	**6**	2	1	1	2	2	0
C9	8	3	12	4	4	2	8	13	**66**	3	1	4	12	0
C10	2	5	26	35	8	1	38	30	9	**65**	6	34	27	1
C11	3	0	0	7	0	0	4	5	4	11	**88**	11	7	0
C12	1	2	3	7	0	2	7	3	1	4	1	**20**	8	1
C13	5	1	1	4	0	2	3	4	1	1	1	3	**8**	0
C14	4	0	0	2	13	10	0	1	0	0	0	0	0	**69**

Table.3.9c. Model accuracies for 14 classes derived from the integrated data set, “3PC+SAVI+G+RE+TWI+ VDCN+MRVBF+PrCv”

All Ten Layers: 3PC+SAVI+G+RE+TWI+ VDCN +MRVBF+PrCv	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
Commission	73.1	92.9	84.3	59.0	67.3	15.3	30.9	35.2	29.5	6.9	14.0	4.4	8.4	91.1
Omission	54.0	13.3	78.0	90.0	59.7	55.7	88.3	93.8	33.6	35.3	12.4	79.6	91.9	31.4
Producer's Accuracy	46.0	86.7	22.0	10.0	40.3	44.3	11.7	6.2	66.4	64.7	87.6	20.4	8.1	68.6
User's Accuracy	26.9	7.1	15.4	40.3	12.3	73.6	57.0	24.5	14.5	53.2	4.7	20.7	7.8	8.9
Overall Accuracy	**35.1%**													
Cohen's Kappa Coefficient	**0.231**													

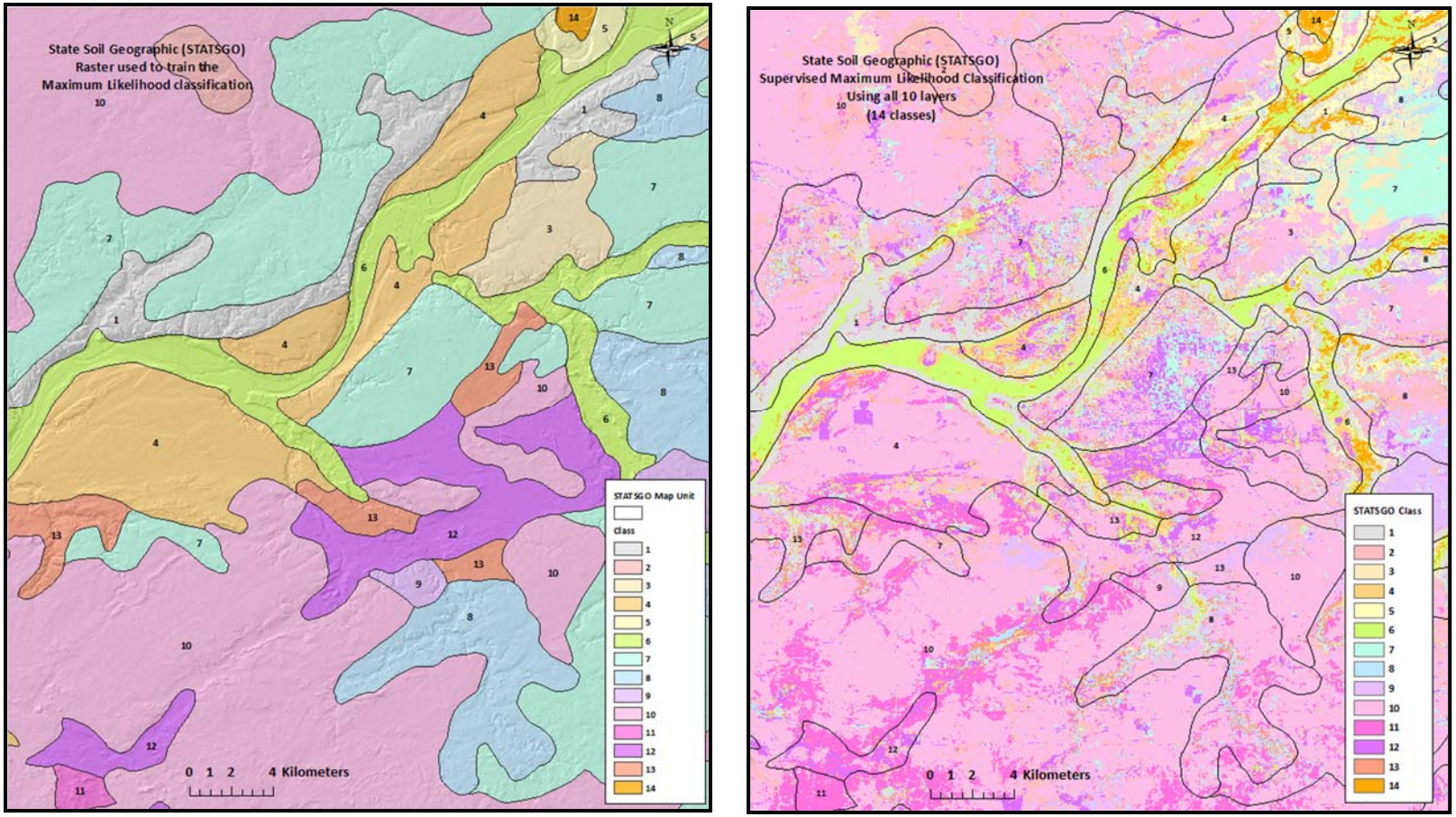


Figure 3.19. Best result of supervised Maximum Likelihood classification using the "3PC+SAVI+G+RE+TWI+VDCN+MRVBF+PrCv" data set (right) and STATSGO classes for comparison (left).

3.4.2 Soil Consociations and Complexes Derived from SSURGO

This study integrated three terrain derivatives, geomorphons, and Landsat 8 data in order to classify SSURGO soil complexes and consociations in a small area in the northern portion of Tippecanoe County. Five different combinations of satellite and DEM data were used for the classification. The combinations are listed in Table 3.10.

Table 3.10. Description of data sets used for supervised classification of Tippecanoe County.

Combination of Satellite and DEM Data	Code Used for Dataset
3 Principal Component Bands from Landsat 8, SAVI, Geomorphons, Relative Elevation, TWI, Profile Curvature	3PC+SAVI+G+RE+TWI+PrCv
3 Principal Component Bands from Landsat 8, SAVI, Relative Elevation, TWI, Profile Curvature	3PC+SAVI+RE+TWI+PrCv
Geomorphons, Relative Elevation, TWI, Profile Curvature	G+RE+TWI+PrCv
Relative Elevation, TWI, Profile Curvature	RE+TWI+PrCv
3 Principal Component Bands from Landsat 8, SAVI	3PC+SAVI

The results of the unsupervised classification are presented in Figures 3.20a-c. The results for agreement with the full SSURGO map units file were observed for all 5 data set combinations and are presented in Table 3.11. In addition, the error matrices, commission and omission, user's and producer's accuracy, overall accuracy, and Cohen's kappa coefficient were calculated. These are presented in Table 3.12a-c for the classified image with the highest accuracy (Figure 3.21).

A visual comparison with the SSURGO class raster shows that the "3PC+SAVI+G+RE+TWI+PrCv" composite data set produced the best result from the unsupervised classification. It is difficult to determine which classes were best classified through unsupervised classification because there are multiple classes within one map

unit boundary in the results map (Figure 3.20a). Results of the supervised Maximum Likelihood classification show that the "G+RE+TWI+PrCv" gave significantly higher agreement with the reference raster than the other composite data sets. Using the "3PC+SAVI+RE+TWI+PrCv," eight-layer composite data set produced the lowest overall accuracy (5.1%). Removing the geomorphons layer from this data set produced slightly higher agreement with the reference pixels. When the terrain derivative layers and satellite data layers were used separately for classification ("Re+TWI+PrCv" and "3PC+SAVI," respectively), the satellite data produced better results.

The class accuracies for the classes derived from the "G+RE+TWI+PrCv" composite data set are presented in Table 3.12b. Overall agreement was very high for classes 8 (100%), 9 (85%), 12 (100%), and 13 (100%). Accuracy was very low for classes 7 (8%) and 11 (10%). Only five classes had at least 50% agreement. Classes 10 and 13 had the highest overall agreement across all composite data sets. Misclassified pixels were distributed randomly across all classes in the other resulting rasters.

Table 3.11. Results of supervised Maximum Likelihood classification of the SSURGO, small study area using different data sets.

Dataset	3PC+SAVI+G+RE+TWI+PrCv	3PC+SAVI+RE+TWI+PrCv	G+RE+TWI+PrCv	RE+TWI+PrCv	3PC+SAVI
Overall Accuracy	5.122%	7.613 %	38.105%	13.845%	17.574%
Cohen's Kappa Coefficient	0.027	0.049	0.206	-0.002	0.093

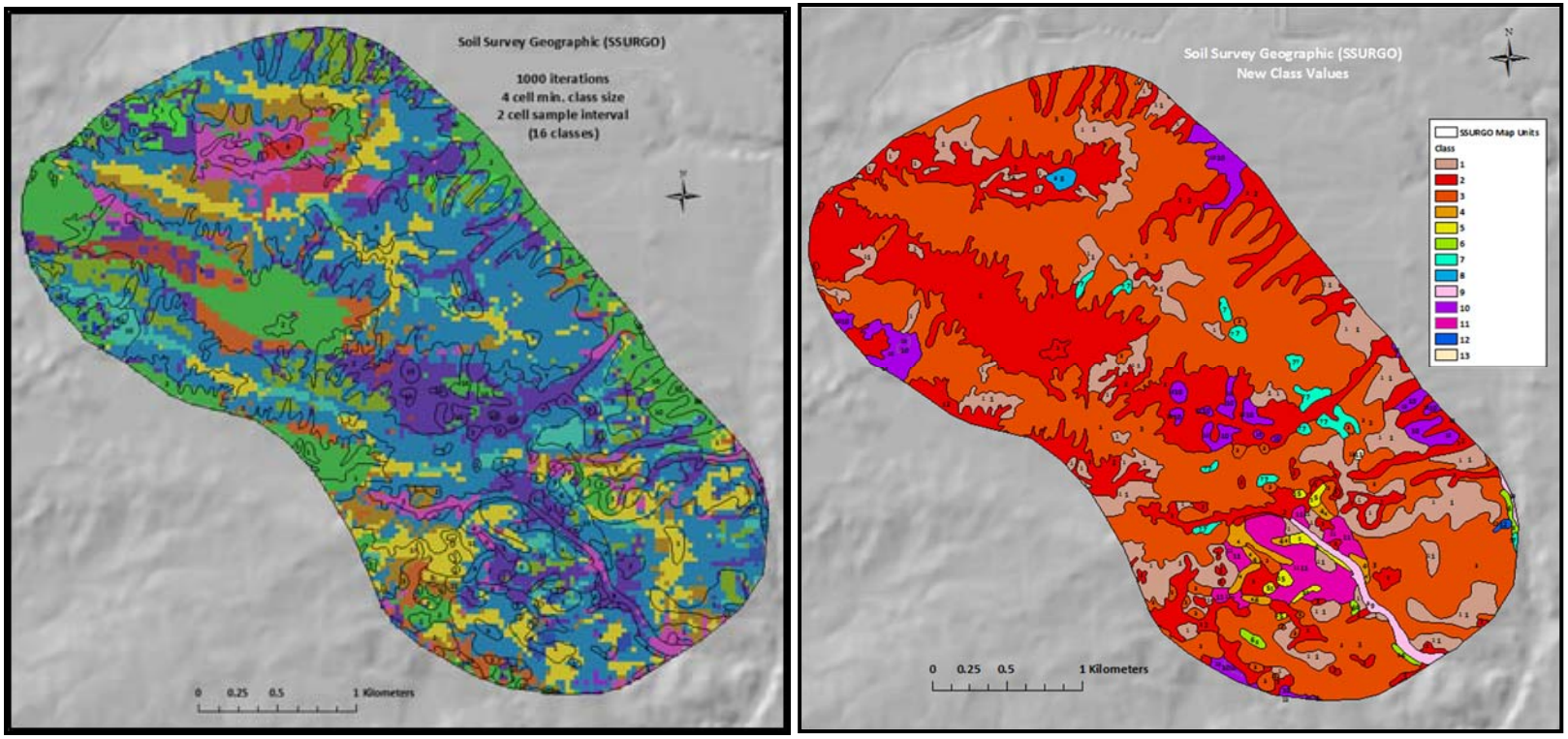


Figure 3.20a. Result of unsupervised classification using the "3PC+SAVI+G+RE+TWI+PrCv" data set (left) and SSURGO classes for comparison (right).

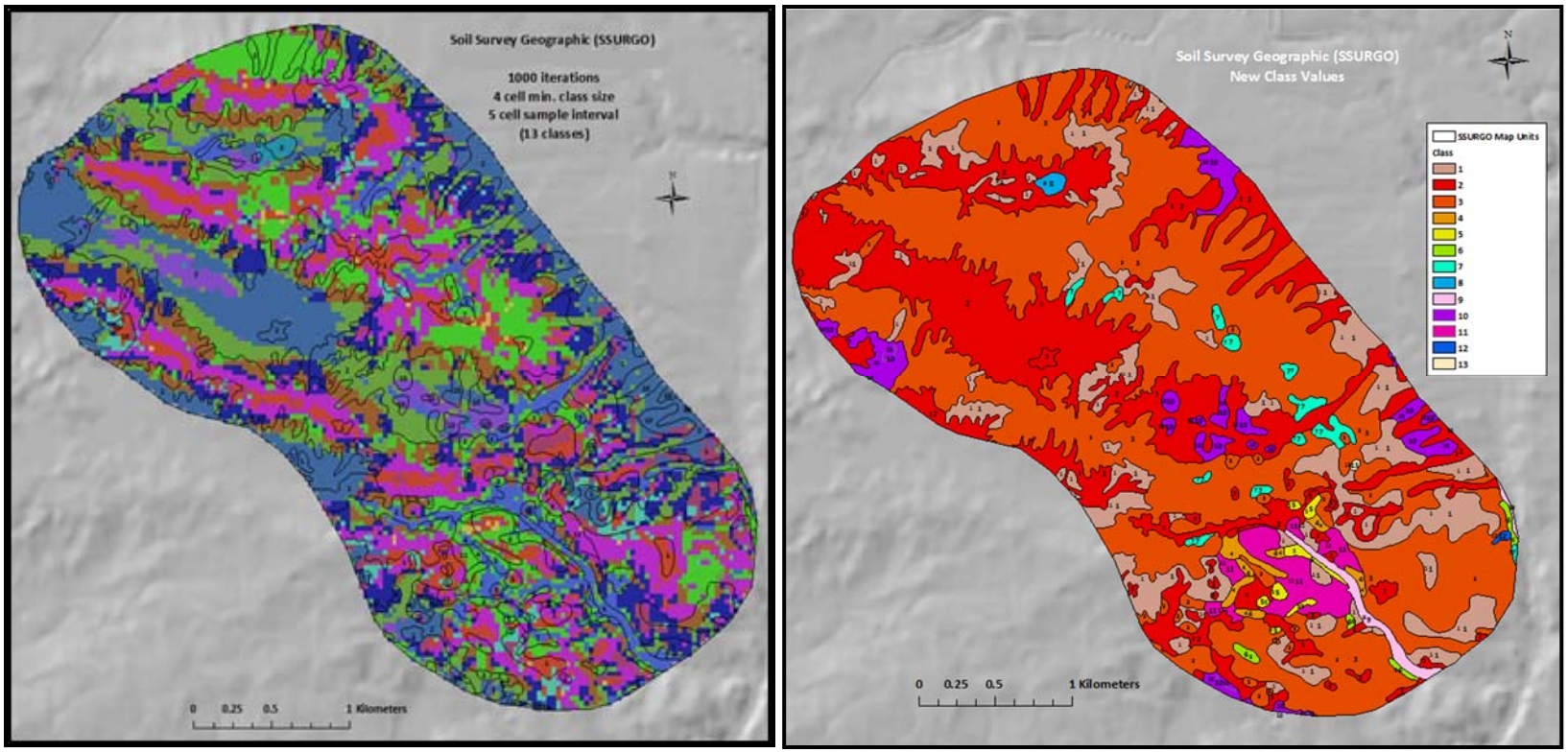


Figure 3.20b. Result of unsupervised classification using the "G+RE+TWI+PrCv" data set (left) and SSURGO classes for comparison (right).

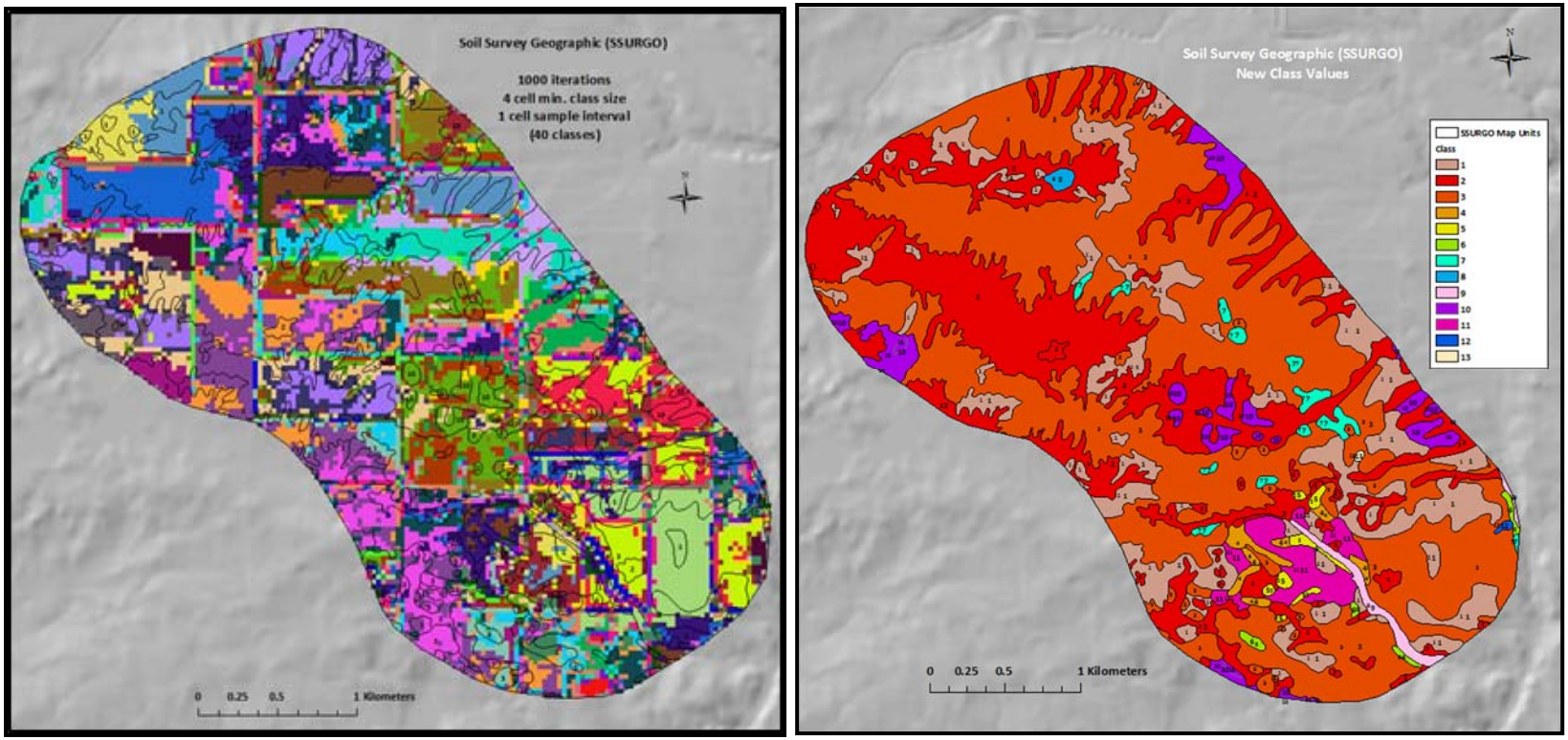


Figure 3.20c. Result of unsupervised classification using the "3PC+SAVI" data set (left) and SSURGO classes for comparison (right).

Table 3.12a. Error matrix for 13 SSURGO classes derived from the integrated data set, “G+RE+TWI+PrCv”

Layer Set: G+RE+TWI+PrCv	Reference C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	Ground Truth
Test C1	**314**	41	416	1	2	0	1	0	0	40	8	0	0	823
C2	182	**910**	439	3	0	3	6	0	3	34	22	0	0	1602
C3	454	319	**2716**	15	10	7	23	0	3	32	75	0	0	3654
C4	12	78	464	**45**	7	4	36	0	2	0	45	0	0	693
C5	42	98	579	7	**28**	4	25	0	1	0	31	0	0	815
C6	21	48	220	8	6	**17**	19	0	3	1	32	0	0	375
C7	10	27	103	11	1	1	**12**	0	0	0	15	0	0	180
C8	59	517	106	0	0	0	0	**23**	1	28	1	0	0	0
C9	37	383	156	0	3	3	10	0	**82**	6	17	0	0	697
C10	534	1094	388	0	1	0	1	0	2	**266**	7	0	0	2293
C11	42	17	253	8	3	3	8	0	0	0	**28**	0	0	362
C12	52	7	102	0	0	0	2	0	0	3	6	**6**	0	178
C13	1	4	0	0	0	0	0	0	0	0	0	0	**4**	9
Total	1760	3543	5942	98	61	42	143	23	97	410	287	6	4	**11681**

Table.3.12b. Class accuracies for 13 SSURGO classes derived from the integrated data set, “G+RE+TWI+PrCv”

Layer Set: G+RE+TWI+PrCv	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
C1	**18**	1	7	1	3	0	1	0	0	10	3	0	0
C2	10	**26**	7	3	0	7	4	0	3	8	8	0	0
C3	26	9	**46**	15	16	17	16	0	3	8	26	0	0
C4	1	2	8	**46**	11	10	25	0	2	0	16	0	0
C5	2	3	10	7	**46**	10	17	0	1	0	11	0	0
C6	1	1	4	8	10	**40**	13	0	3	0	11	0	0
C7	1	1	2	11	2	2	**8**	0	0	0	5	0	0
C8	3	15	2	0	0	0	0	**100**	1	7	0	0	0
C9	2	11	3	0	5	7	7	0	**85**	1	6	0	0
C10	30	31	7	0	2	0	1	0	2	**65**	2	0	0
C11	2	0	4	8	5	7	6	0	0	0	**10**	0	0
C12	3	0	2	0	0	0	1	0	0	1	2	**100**	0
C13	0	0	0	0	0	0	0	0	0	0	0	0	**100**

Table 3.12 c. Model accuracies for 14 classes derived from the integrated data set, "3PC+SAVI+G+RE+TWI+ VDCN+MRVBF+PrCv"

G+RE+TWI+PrCv	C1	C2	C3	C4		C5	C6	C7	C8	C9	C10	C11	C12	C13
Commission	61.8	43.2	25.7	93.5		96.6	95.5	93.3	0.0	88.2	88.4	92.3	96.6	55.6
Omission	82.2	74.3	54.3	54.1		54.1	59.5	91.6	0.0	15.5	35.1	90.2	0.0	0.0
Producer's Accuracy	17.8	25.7	45.7	45.9		45.9	40.5	8.4	100.0	84.5	64.9	9.8	100.0	100.0
User's Accuracy	38.2	56.8	74.3	6.5		3.4	4.5	6.7	0.0	11.8	11.6	7.7	3.4	44.4
Overall Accuracy	**38.10%**													
Cohen's Kappa Coefficient	**0.206**													

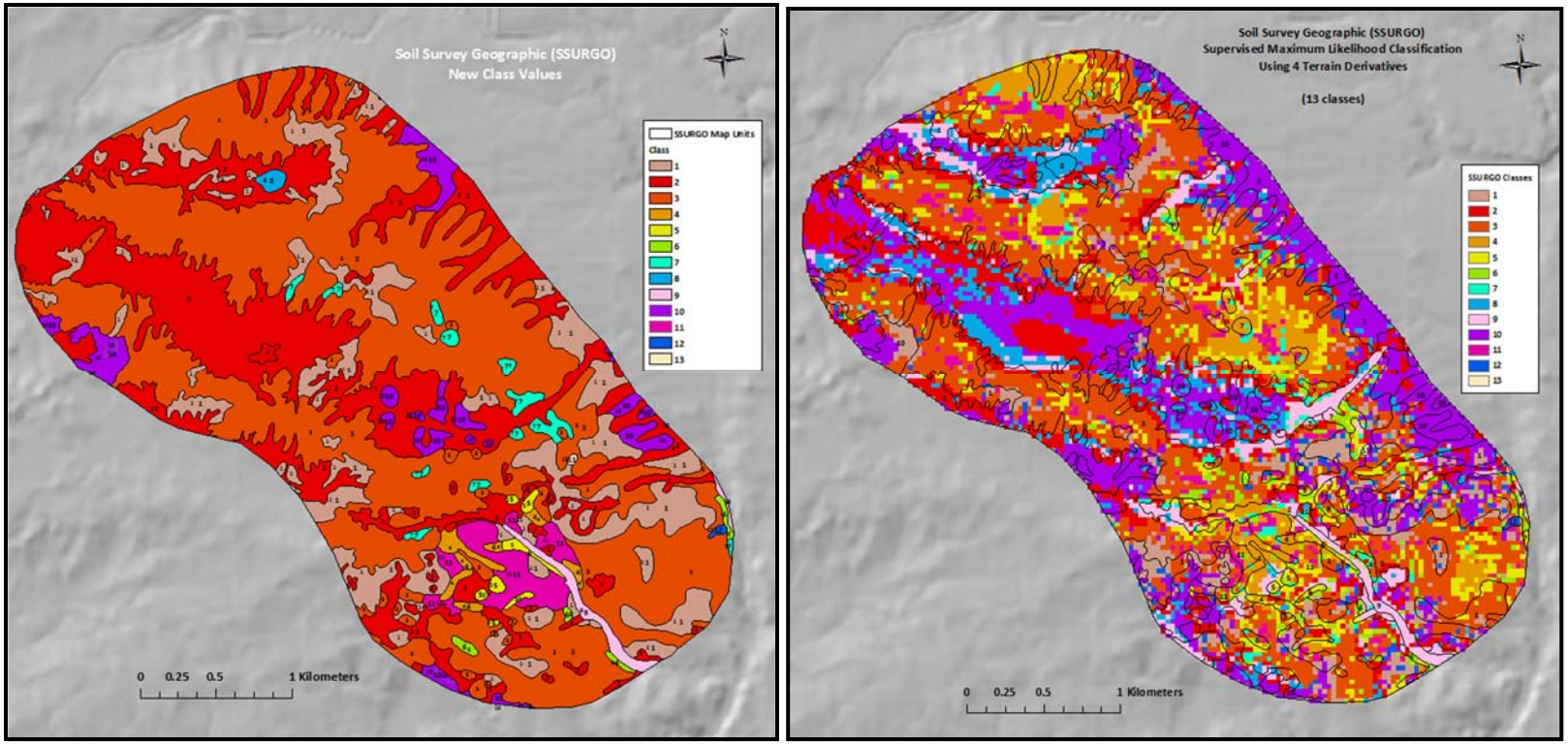


Figure 3.21. Best result of supervised Maximum Likelihood classification using the "G+RE+TWI+PrCv" data set (right) and SSURGO classes for comparison (left).

3.5 Summary and Discussion

The main objective of this study was to examine the effectiveness of combining new multispectral satellite data with topographic data to generate a regional-scale soil class map. The intent of the study was to develop a quick methodology for working with new, freely-available data that others may be able to adopt for regional-scale mapping with limited soils data. This study used a set of existing class maps which represent soil-landscape patterns in Tippecanoe County, Indiana. The two soil class maps used represent mapping at different levels of detail, but both could be used for regional-scale planning. The generalized representations of the soils of Indiana were used as a reference or "ground truth" data for quantitative image analysis of maps derived from the integration of new topographic data and satellite data. The particular approach employed is useful in that it may be customized for a different study area; however, the quality of the input data must be carefully analyzed, and without a clear understanding of the level of generalization in the representative soil map units, it is difficult to truly assess the efficiency of the model.

Much knowledge was gained from this study about preparation, analysis, and interpretation of topographic and satellite data sets. The approach used in this study made several assumptions that may or may not prove accurate in a natural setting. Some of the error in this approach may be attributed to these assumptions, error in data preparation, and error inherent in the data. However, it is impossible to quantify the amount of error that is attributed to a particular source.

Error is introduced any time that spatial data is resampled, such as the DEM that was resampled from a higher pixel resolution to a lower pixel resolution. Using a projected coordinate system like UTM also introduces some error in geo-location, but this is the case with any coordinate system that attempts to model the imperfect sphere of the earth. Soil map units from STATSGO and SSURGO both contain varying degrees of generalized information that is difficult to extract to a larger scale. This study made the

assumption that the map units were delineated using breaks in the soil-landscape relationships observed. It is impossible for anyone except for the soil scientist who delineated the map units to truly understand the soil relationships that they were modelling. The study also assumed that a thirty-meter resolution was sufficient to properly observe these changes in topographic features, even though it was known that the map units encompass a much larger area than just thirty meters. Also, even though SSURGO map units are smaller than STATSGO map units, the same thirty-meter composite data sets were used to classify both. Map units that covered a very small area were combined with the map units adjacent to them on the map rather than simply excluding them from the classification. This would be one likely source of error.

Certain error should also be attributed to the quality and assumptions made about the satellite data. The date of collection of satellite imagery can have dramatic effects on the classification results. The study used multispectral data from a single satellite image and the assumption was made that this data was a good approximation of average spectral values for vegetation for the study area. It is known that this constancy in spectral values does not exist over the period of a year and, depending on climatic conditions, it may not exist over multiple years either.

Even with these temporal, spectral, and spatial variables involved, it is possible to draw some conclusions about the relationships between the data inputs and the results of the classification. In both the STATSGO study and the SSURGO study, when the topographic data was used separately from the satellite data, the classification results from the topographic composite data set had higher agreement with the reference file than did those from the satellite composite data set. Integrating satellite data with topographic data for classification of the STATSGO map units produced a map with higher overall accuracy than using an integrated composite data set for classification of the SSURGO map units.

An evaluation of the accuracy of the soil map units used as the standards for comparison and as training sets for classification was not included in the scope of this study. The existing STATSGO and SSURGO soil maps were assumed to be true data. This

is an assumption that is somewhat flawed by the fact that the soil map units are mere representations of a natural system and furthermore, they may be several scientists' interpretations, which would introduce errors and biases into the model. For this reason, the level of accuracy assigned to a classified image that uses soil map units as training sets, suggests that the existing map units are 100% correct. It is imperative to realize that even results with better than average overall accuracy may be better representations of the natural soil diversity in an area than the model results suggest. The best measure of model accuracy would be to statistically assess the reality of the reference soil map units and use that to verify the accuracy of the classification.

It is important to reiterate that the STATSGO and SSURGO soil maps represent differences in Soil Taxonomy. Soil map units are sets of soils that exhibit structural heterogeneity and they are delineated based on soil physical and chemical properties rather than functional properties. Spectral responses in a heavily vegetated area such as Indiana will highlight variability in vegetation biomass which is directly linked to water availability. However, this is just one characteristic of soil variability in a landscape. Satellite data is also limited by only capturing surface readings, whereas soil morphology varies with depth. More research is needed to relate spectral signatures to soil morphological properties to be able to better classify soils with satellite data.

Supervised classification may be improved by classifying images collected at different times of the year, such as when the soil is bare and during drought years. A method to improve the supervised classification results may be to incorporate weightings into the Maximum Likelihood function. Weighting could be performed by multiplying the total pixels in each class by the average class value. This model would account for the difference in size of the classes. With the right software tools, the maximum likelihood supervised classification could be implemented using a discrete or continuous algorithm that better fits the data input distribution rather than assuming Gaussian distribution.

Soil mapping at the regional-scale is an expensive, labor-intensive undertaking. It requires the direction of knowledgeable soil scientists that can recognize soil properties

in the field and who can accurately interpret soil-landscapes. The advent of GIS tools, higher computing power, and the availability of free topographic and satellite data, offers soil scientists the opportunity to explore soil relationships and patterns in space. However, further research is needed to combine these tools and datasets and retrieve useful soil information for developing countries in great need of it.

CHAPTER 4. DEVELOPMENT OF A SEMI-AUTOMATED MODEL TO IMPROVE POSITIONAL ACCURACY OF SOIL SURVEY PEDONS FOR INDIANA

4.1 Introduction

A rapid increase in recent years in the use of GIS technologies and modelling software in natural resources and land management has increased the demand for quantitative soil information. Soil point data is a vital input in predictive environmental models such as those used to generate continuous soil property maps to make field-scale management decisions and recommendations for precision agriculture. For these reasons, it is necessary to examine the integrity and spatial accuracy of this data set. This study focuses on soil pedon data for Indiana. However, the methods and tools developed can be adopted for other states.

4.1.1 Indiana Pedon Data

In soil surveying, the pedon is defined as the smallest unit that captures the characteristics of a soil individual and allows a thorough evaluation of its horizons and vertical extent (Buol et al., 2003). For purposes of this study, a soil pedon is defined as a U.S. Soil Survey soil sample with a single georeferenced location and detailed site description with attributes of one or more of the following: soil morphological properties measured in the field, and chemical, physical, and mineralogical properties measured in the lab. Traditionally, pedon samples have provided reference data useful for calibrating soil scientists' field estimates, for establishing soil relationships between its forming factors, and for validating and establishing the soil's classification.

A recent paradigm shift from traditional soil surveys that rely on tacit, expert knowledge to the modern use of digital soil mapping used to produce continuous property maps has increased the demand for concrete, accurate pedon data (Ashtekar & Owens, 2013). Presently, this data is used in predictive models to interpolate soil property values to areas that were not sampled.

Pedon data for Indiana was first collected in 1967 as part of an accelerated soil survey program called the Soil Characterization Program (Crum, et al., 1977). This was a cooperative effort between the U.S. Soil Survey, known at the time as the Soil Conservation Service, and Purdue University. Soil sampling was conducted for two separate purposes: for research and to define characteristics for soil map units for county soil surveys. If for research, soils were commonly sampled on a transect or as part of a set with a single varying soil forming factor (i.e. along a toposequence if the variable was relief or a genosequence if the variable was parent material). For delineating soil map units, reconnaissance sampling was performed where several soils were observed and a representative soil was chosen, described, sampled and analyzed.

Of prime importance to the pedon data records and this study are the site descriptions collected at the sampling location. The surveyors followed the guidelines outlined in the Soil Survey Manual (Soil Survey Division Staff, 1951) for conducting soil descriptions in the field. The manual's guidelines for site selection recommended choosing a site with good accessibility and avoiding roads, fences, farmsteads, or disturbances which may change the soil morphology. The site description consisted of the series name as described by the soil scientist in the field, slope, erosion class, drainage, landform, vegetation, parent material, location, county, a record soil number, file number, date described, and describer's name.

The *Soil Characterization in Indiana: Field and Laboratory Procedures* report (Crum, et al., 1977) describes in detail the methods used in the field and in the lab for soil sample descriptions and analysis. Soil pedons were assigned the series name of the map unit consociation or one of the series names of a complex map unit. Series names were subject to change pending the lab results. Depending on the level of inconsistency

between the field descriptions and lab data, the pedon was declared a taxadjunct, variant, or reassigned a series and/or map unit name. If a pedon had diagnostic characteristics distinct from its series concept, but the same interpretations as the series, it was called a taxadjunct. If the pedon had characteristics distinct enough from all other series, but its extent in the county was not great enough to warrant a new series, it was called a variant. If lab and field studies revealed that a pedon did not belong in the map unit it was originally assigned, its map unit and series name was changed.

These data were stored in a database created and managed by Purdue University in a detailed form. An example is presented in Appendix A. Standard abbreviations from the Soil Survey Manual, presented in Appendix B, were adopted for the soil descriptions conducted as part of this joint effort. Over the course of twenty-five years, thousands of pedons from Indiana were analyzed by the Purdue Soil Characterization Lab and the National Soil Survey Lab. Before merging the pedon data collected over the years into a single database, certain issues were addressed and corrected, including updating the series names to fit the current definitions and standardizing the horizon designations used. In the mid-1990s, the data were transferred into NASIS, where each pedon point was assigned a unique “Pedon ID” and “User Site ID” comprised of the year the pedon was sampled, the FIPS (Federal Information Processing Standard) code for the county where it was sampled, and a record number. Pedon points continued to be sampled by NRCS scientists after 1990 and data collected was subsequently entered into NASIS.

The greatest distinction between the pedon data collected before and after the 1990s is in the sources used to georeference the pedon sample site. Although no documentation exists for this, it is consensus among Indiana’s soil scientists that GPS units were not widely used in the field until after 1994. Therefore, this study assumes that soil pedon points that were sampled before 1995 were not georeferenced using GPS and that all points sampled in 1995 or later, were georeferenced using a GPS unit in the field. Prior to 1995, the best estimate of the site location was made by marking 7.5-minute topographic maps and aerial photos with the sample site location and then using

those as a reference to obtain a precise point location by citing the United States Public Land Survey System (PLSS). The PLSS for Indiana and the five-point system of location are described in detail in the following sections.

4.1.2 Indiana Public Land Survey System Grid

The push to survey was driven by settlers' desire to move west of the Ohio River in search of more plentiful and inexpensive land. After gaining control of the Northwest Territory, the United States established a new survey system. The original thirteen colonies had been surveyed in a system of metes and bounds that described the perimeter of a property through unstandardized boundary marks and thus, was not consistent across the state (Steinhardt, et al., 2013). In 1804, newly appointed Surveyor General, Jared Mansfield, introduced a more successful, standardized system which became the PLSS (Linklater, 2003).

The United States Public Land Survey System (PLSS) is a square-grid system of land parcels that extends through 30 states including Indiana. Its organized structure is a result of a fast-growing American population that demanded an accurate description of their property boundaries. Its well-defined subdivisions and extent deem it a useful tool for delineating property in rural areas and for locating soil sampling sites prior to the development of GPS.

The PLSS is a tier system of grids of decreasing area. The different components of the PLSS are presented in Figure 4.1 below. The first level in this system is a set of principal meridians and baselines, known collectively as Standard Lines. The First Principal Meridian is a north-south line that lies one mile west of the border between Ohio and Indiana. The Base Line intersects the Principal Meridian at a right angle and continues west. The Standard Lines serve as the zero point from which the second level, composed of townships and ranges, is labeled. Townships run parallel to the Base Line and are referenced north or south to this line. Similarly, ranges run parallel to the

Principal Meridian and are referenced east or west to this line. In 1796, Congress required that lands be split by lines running parallel and perpendicular to the Earth's true meridians at six-mile intervals, thus, townships were defined as 6-mi^2 tracts by the boundaries of the meridians and base lines (White, 1926). Townships are further divided into thirty-six sections - the third level in the PLSS system. Sections are roughly 1-mi^2 or 640 acres each. Sections were numbered in a boustrophedonic fashion, that is, in a right to left then left to right movement with Section 1 in the northeast corner, Section 6 in the northwest corner, Section 31 in the southwest corner, and Section 36 in the southeast corner. The fourth level of grids are the quarter (Q) sections of roughly 160 acres. The fifth level is comprised of quarter-quarter (QQ) sections, approximately 40 acres each and the sixth level is the quarter-quarter-quarter (QQQ) sections of 10 acres each.

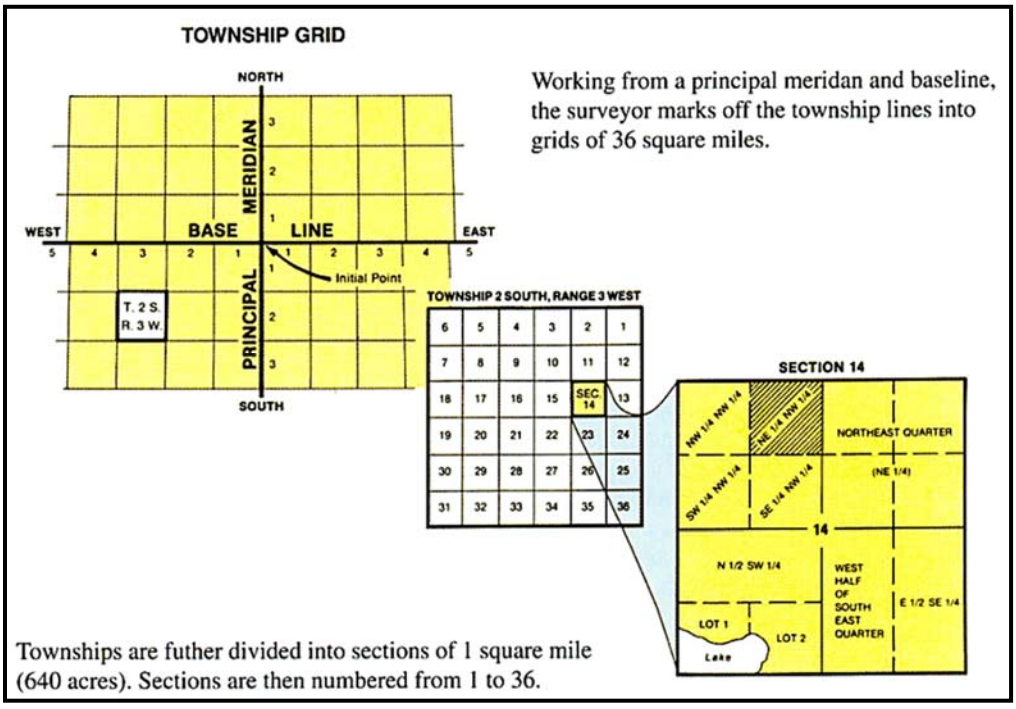


Figure 4.1. Component of the Public Land Survey System grid. [From Bureau of Land Management, 2002]

Land in Indiana was sold in sections, half-sections, quarter sections, and even quarter-quarter sections. For this reason, property boundaries and county roads in rural areas usually follow the PLSS grid. In some areas, the components of the PLSS can be identified fairly accurately using an aerial photograph. From the air, the graphing-paper structure of the Indiana PLSS may appear immaculate; however, it is not perfect.

More often than not, the northwest sections of a township are much smaller than the 640 acres that were intended because these were the last sections surveyed and any excess or deficiency in the measure of the sections was carried to this corner. Furthermore, as with any model that attempts to fit the Earth to a planar system, the standard lines of the PLSS encountered some issues. Due to the Earth's curvature, there was a convergence of the meridians north of the state and a divergence south of the state. To compensate for the model error, "correction lines" to move part of the meridians fifty or sixty yards further east or west of their original location, were installed prior to completion of the survey. The western side of Section 31 was extended to more than a half mile in length and the western part of Section 6 was reduced to less than a half mile in length (White, 1926). Some of the inconsistencies in the Indiana PLSS grid have nothing to do with human error in surveying or with the shape of the Earth, but rather they are a result of land grants that were not were not surveyed in the PLSS, but rather French surveys conducted before the PLSS (Steinhardt et al., 2013). Even with these imperfections, the PLSS grid of Indiana allows for an accurate location description using the five point-system.

The general success of the PLSS grid was due to the fact that it fractally divided the land into the smallest tracts and thus provided people the opportunity to purchase land easily. In soil survey, the PLSS grid made it possible to accurately describe the location of a sampling site. According to Mike Wigginton, Resource Soil Scientist in Indiana, explained that soil scientists would plot their location on an orthophoto, then they would measure the distance from the plotted point to a reference corner or center. Some soil scientists paced off their sampling location to the nearest recognizable reference point. They were instructed to use the five-point system which described their

specific location as a certain distance and cardinal direction from the corners or center of a township, section, Q section, QQ section, or QQQ section. Location descriptions for soil pedon points were written as general as to the center of a township and as specific as to an x and y distance from the corner or center of a quarter-quarter-quarter section. To identify a land parcel using the PLSS grid, a description would be written from the smallest component to the largest component. For example in Figure 4.2, "Parcel B would be described as the Southeast ¼ of the Southeast ¼ of the Southwest ¼ of Section 10 T2N R3W, Second Principal Meridian [and it would be] written: SE1/4 of SE1/4 of SW1/4 of Sec.10 T2N R3W Second Principal Meridian" (Steinhardt et al., 2013) in a soil scientist's field notes.

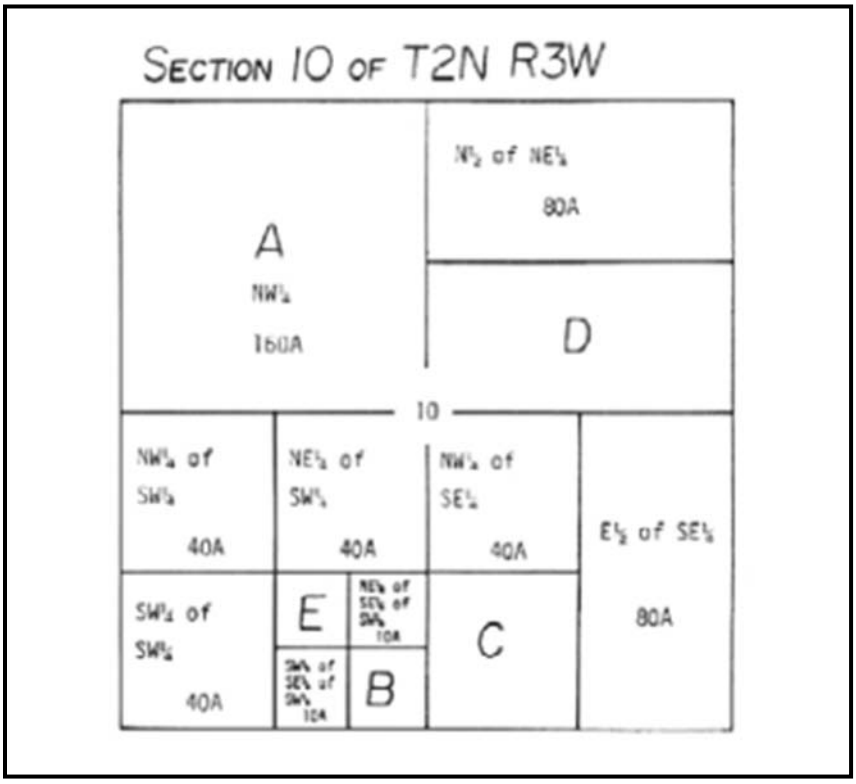


Figure 4.2. Detail of Section 10 of Township 2 North, Range 3 West, Second Principal Meridian of Indiana.[From (Steinhardt et al., 2013)].

Detailed written notes and records of the PLSS grid were kept by the General Surveyor and his deputies and maintained by the U.S. General Land Office. With the help of these records and aerial photographs, the USGS National Mapping Program created Quadrangle Maps for the United States that accurately depicted PLSS grid components. The *National Mapping Program Technical Guide* (USGS & USDA-FS, 2003), provides a description of the PLSS features shown on USGS single edition quadrangle maps. In 1998, the Indiana Geological Survey, published a digital PLSS grid of Indiana that was derived from the 7.5-minute USGS topographic maps. Only the townships, ranges, and sections were digitized and are currently available to the public. This digitized PLSS grid was used by the NRCS to georeference the U.S. Soil Survey pedon sample locations through a computer-based algorithm.

4.1.3 NRCS Method to Locate Pedon Points

A few years before 1994, Thomas Reinsch of the National Soil Survey Center, developed a computer algorithm to convert soil scientists' textual pedon location descriptions to a coordinate point on a NAD 83 geodetic reference system. His PLSS Conversion model required a PLSS grid's section center and the distance and direction from the center or corners in order to assign a geographic location to these points. Such a model was necessary in order to georeference pedon points that had been surveyed before the use of GPS units in the field and for which no geographic location existed.

Reinsch's PLSS Conversion model was written in Visual Basic for Microsoft Access and standardized to process land descriptions from all thirty states that had adopted the PLSS grid. It was pre-loaded with a list of pedon points, their textual location descriptions and a table with x and y latitude and longitude coordinates for all section centroids for every state. In preparation for running the model, the location descriptions were parsed and the text was split into columns for each of the major PLSS components. As a first step in locating the point, the model used the components to locate the most

specific land parcel described (either township or section) and the centroid of that parcel was used to assign a temporary geolocation for that pedon at the corner of a section or the center of a section. Even if the pedon sample point was described from a more detailed location than the section, the temporary point was placed at the section centroid. For points that were described as a cardinal distance from a section corner, the assumption was made that the section center was a 804 meters (0.5 mile) east or west and 804 meters (0.5 mile) north or south of the section corner (Reinsch, 1994). The next step in the model transformed the linear distance and compass direction to the corner or center of a section, if provided, to a projected cardinal distance through polyconic inverse ellipsoidal equations from Snyder (1987). The projected distance was then used to move the pedon point from its temporary location to a final georeferenced coordinate location. The x and y coordinates generated by the PLSS conversion model were exported to NASIS and assigned as the official locations of the pedon points. These coordinates are currently used as the pedon point locations.

After examining the PLSS Conversion model structure and a phone conversation with Thomas Reinsch, it was concluded that there are several issues with the model. The accuracy of this method depends on equal-area polygon sections which are not the only types of sections present in Indiana's PLSS grid. Furthermore, the PLSS Conversion model ignores points described beyond the section. This is due in part to the level of detail of the best-available digitized PLSS grid from the Indiana Bureau of Land Management (BLM) (Indiana Geological Survey, 1:24,000, PLSS Section Polygon Shapefile) which does not explicitly demarcate any of the quarter sections. This means that points with location descriptions that reference Q, QQ, or QQQ sections are not located to the most specific land parcel possible, but rather to a section and, thus, error is inadvertently introduced into the pedon location.

4.1.4 Summary of Issues with the Pedon Data

Despite the efforts of the NRCS and its collaborators to maintain the integrity of the soil pedon data and provide an accurate location for the pedon point, there are several inaccuracies present in the pedon site descriptions. One of the issues is missing or incomplete information in the location description and other portions of the pedon site description. Another issue is that the PLSS Conversion model used to derive the official pedon point coordinates makes assumptions that are not entirely accurate and in doing so, introduces error into the positional accuracy of the pedon points. These and other issues will be discussed in the following sections and a methodology to assess the positional accuracy of the current pedon points will be presented.

4.2 Materials and Methods

This study focuses on U.S. Soil Survey pedon data available for the state of Indiana. The geographic boundaries for Indiana in NAD 83, UTM Zone 16N are: North 4625481.094636 meters, South 4180918.750000 meters, West 403539.062500 meters, and East 692187.625431 meters. There were data within the database without any geographic location. However, because the aim of the study is to assess the positional accuracy of the pedon data, only data with location descriptions and site information was used. Within this study, GIS layers were projected to the NAD 83, UTM Zone 16N coordinate system. Some layers, like the PLSS grid layers for Indiana were already projected in UTM, while most of the NASIS pedon point data was in WGS 84. All methods and tools presented hereafter are intended for use in any U.S. state that has adopted the PLSS grid, even though the methods developed in this study are currently calibrated for Indiana.

4.2.1 Acquiring a Statewide Set of Pedon Points

In order to improve the positional accuracy of the pedon points in the USDA Soil Survey Characterization Database, it was necessary to become familiar with the NASIS data structure and format. Pedon point data is maintained as part of the Soil Characterization Record in NASIS. A search for the pedons for Indiana was conducted with the help of expert soil scientists from the NRCS (Mike Wigginton, Henry Furgeson, Gary Struben, and Rick Neilson). A total of 6459 Indiana pedon points were extracted from the complete collection of pedon data for the United States stored in NASIS which contains over 60,000 site records. Approximately 64% of the points for Indiana had location data associated with the point data. Any missing information in the location descriptions was populated whenever possible, using computer-based algorithms that were developed.

A number of tools exist to query for data in the numerous tables within the NASIS system. An up-to-date collection of all pedon points for the United States was received from the National Soil Survey Office in tabular form as an Analysis PC version 2.1 database. Analysis PC is a stand-alone Microsoft Office database program that is equipped with a pedon database and a link to the SSURGO database (United States Department of Agriculture - NRCS, 2008).

Unique identifiers for Indiana were found by reading the NASIS version 7.0.4 Table Column Descriptions guide (Soil Survey Staff, 2014). The identifiers were used to select pedons that were sampled in Indiana. Each state has a unique State Area ID (stateareaiid column) and Indiana's unique number is "6677." The User Site ID (usiteid column) is the unique value assigned to individual sample pedon points with associated site descriptions. Commonly, the two-letter state abbreviation is contained in the User Site ID. Taking this information into account, the pedons sampled in Indiana were extracted from the Analysis PC database in three steps:

1. The “stateareaiid” column was filtered to select only records with a value of “6677.” The resulting pedons were copied into a blank table in Access.
2. The previous filter was cleared to show all records. Following that process, two new filters were applied concurrently. The “usiteid” column was filtered to select all records which contained “IN.” The “stateareaiid” column was filtered to show only those records with blanks or no data for that column. The resulting pedons were copied to the table with results from Step 1. The “stateareaiid” filter was applied together with the “usiteid” filter in order to prevent duplicate pedons in the results by excluding those which had already been selected in Step 1.
3. The Indiana User Site IDs were exported into a blank Microsoft Excel spreadsheet and duplicates were removed.

From the years, 1961 to 2014, a total of 6459 pedons were collected for Indiana; however, only 4141 pedons had site descriptions with location details, so only these were used for this study.

The NASIS pedon details for all points within the Indiana data were set to include the minimum site description details provided by the Purdue University pedon data, so as not to exclude any location information necessary for the study. Table 4.1 lists the NASIS columns from which data was collected for each pedon. The full set of complete pedon points for Indiana was copied to an Excel spreadsheet where it was examined for errors. The “error free” set of pedon data, which includes site description information from NASIS, will be referred to as the “NASIS pedon data.”

Within the Indiana data, many pedons were missing portions of their location descriptions. With the help of the PLSS grid and supplemental location information associated with the pedon, missing data was entered manually. For example, many pedons were missing an entry for the principal meridian, so the FIPS code was used to identify the county in which the pedon was sampled. The PLSS details for that pedon were used to navigate to the location described on the PLSS grid and the appropriate principal meridian was assigned depending on where on the map the point fell (Figure

4.3). This procedure to populate the principal meridian was automated as part of the PLSS model that will be discussed in the following section. Some pedons were missing the directionality of their township or range, so this data was strategically added using knowledge about the PLSS components and spatial structure. More errors and inconsistencies emerged as the PLSS details and the location descriptions were parsed and the pedon points mapped. Figure 4.4 shows that a few NASIS pedon points for Indiana have coordinates that place them outside of the state and even the country. Some errors could not be corrected with confidence therefore, these points were eliminated from the study.

For the samples collected prior to 1994, the pedon PLSS details provide the highest level of detail for the location of the pedon points. The PLSS details were parsed in order to separate out the PLSS components. For pedons which were missing PLSS details in the "plss details" column, the location description column often contained complete or partial PLSS details, and so it was parsed into the corresponding PLSS component columns for that point. Common errors identified in the PLSS details included mixed compass directions between the township and range (i.e. T2W, R3N rather than the correct: T2N, R3W), missing or incorrect section parcel numbers and section directions, and incorrect and missing principal meridians (i.e. pedon details listing the first principal meridian when all other PLSS details placed the point in the second principal meridian). For example, some of the locations described through the PLSS details were just not possible considering the structure of the PLSS grid. If no direction and distance details were provided in the location details, it was assumed that the point was sampled in the center of the smallest parcel that the location description references.

Table 4.1. NASIS version 7.0.4 column data copied for every Indiana pedon used in study and the tables where it is located.

Column Physical Name	Column Label	Purpose	Table Name	Table Label
usiteid	User Site ID	Unique identifier for pedon	site	Site
pedlabsampnum	Pedon Lab Sample #	ID number assigned to pedon by lab; links morphological data with analytical data	ncsspedonlabdata	NCSS Pedon Lab Data
upedonid	User Pedon ID	A short label to identify a particular pedon	pedon	Pedon
countyareaiidref	County Name	County of pedon location	site	Site
locdesc	Location Description	Geographic location in terms of landmarks and roads. Commonly contains details to support PLSS details	site	Site
plssmeridian	PLSS Meridian	Identifies a line along an astronomical meridian that is the reference for township boundaries. Component of the PLSS	site	Site
plssrange	PLSS Range	Identifies a township quadrangle, when used in conjunction with township. Component of the PLSS	site	Site
plsstownship	PLSS Township	Identifies a quadrangle of the PLSS grid when combined with the PLSS Range.	site	Site
plsssection	PLSS Section	Numeric identifier of a subdivision of the PLSS Township	site	Site
plssdetails	PLSS Section Details	Textual description of the location of a site with reference to one of the corners of the section and distance and direction to location the point within the section	site	Site
latdegrees	Lat. Degrees	Latitude in degrees generated using PLSS Conversion program	site	Site
latminutes	Lat. Minutes	Latitude in minutes generated using PLSS Conversion program	site	Site
latseconds	Lat. Seconds	Latitude in seconds and decimal seconds generated using PLSS Conversion program	site	Site
latdir	Lat. Direction	Latitude position north or south of the equator generated using PLSS Conversion program	site	Site
longdegrees	Long. Degrees	Longitude in degrees generated using PLSS Conversion program	site	Site
longminutes	Long. Minutes	Longitude in minutes generated using PLSS Conversion program	site	Site
longseconds	Long. Seconds	Longitude in seconds and decimal seconds generated using PLSS Conversion program	site	Site
longdir	Long. Direction	Longitude position north or south of the equator generated using PLSS Conversion program	site	Site
horizdatnm	Datum Name	Reference system used for defining coordinates of the points	site	Site
utmzone	UTM Zone	Zone of the UTM projection system	site	Site

Table 4.1. continued. NASIS version 7.0.4 column data copied for every Indiana pedon used in study and the tables where it is located.

utmnorthing	UTM Northing	The distance, in meters, from the equator	site	Site
utmeasting	UTM Easting	The distance, in meters, proceeding east for the UTM zone.	site	Site
Column Physical Name	Column Label	Purpose	Table Name	Table Label
latstddecimaldegrees	Std. Latitude	Standardized latitude value in decimal degrees, in the WGS 84 geographic coordinate system. Values are auto-populated from a GPS, or computed from the original latitude coordinates using conversion algorithms.	site	Site
longstddecimaldegrees	Std. Longitude	Standardized longitude value in decimal degrees, in the WGS 84 geographic coordinate system. Values are auto-populated from a GPS, or computed from the original latitude coordinates using conversion algorithms.	site	Site
drainagecl	Drainage Class	Identifies the natural drainage conditions of the soil and refers to the frequency and duration of wet periods.	site	Site
pmkind	Kind	A term that describes the general physical, chemical, and mineralogical compositions of the material, mineral or organic, from which soil develops.	sitepm	Site Parent Material
taxorder	Order	Highest level in Soil Taxonomy	petaxhistory	Pedon Taxonomic History
taxsuborder	Suborder	Second level in Soil Taxonomy	petaxhistory	Pedon Taxonomic History
taxgrtgroup	Great Group	Third level in Soil Taxonomy	petaxhistory	Pedon Taxonomic History
taxsubgrp	Subgroup	Fourth level in Soil Taxonomy	petaxhistory	Pedon Taxonomic History
taxpartsize	Particle Size	Particle size classes are used to differentiate between soil families. They refer to the particle-size distribution of the entire soil.		

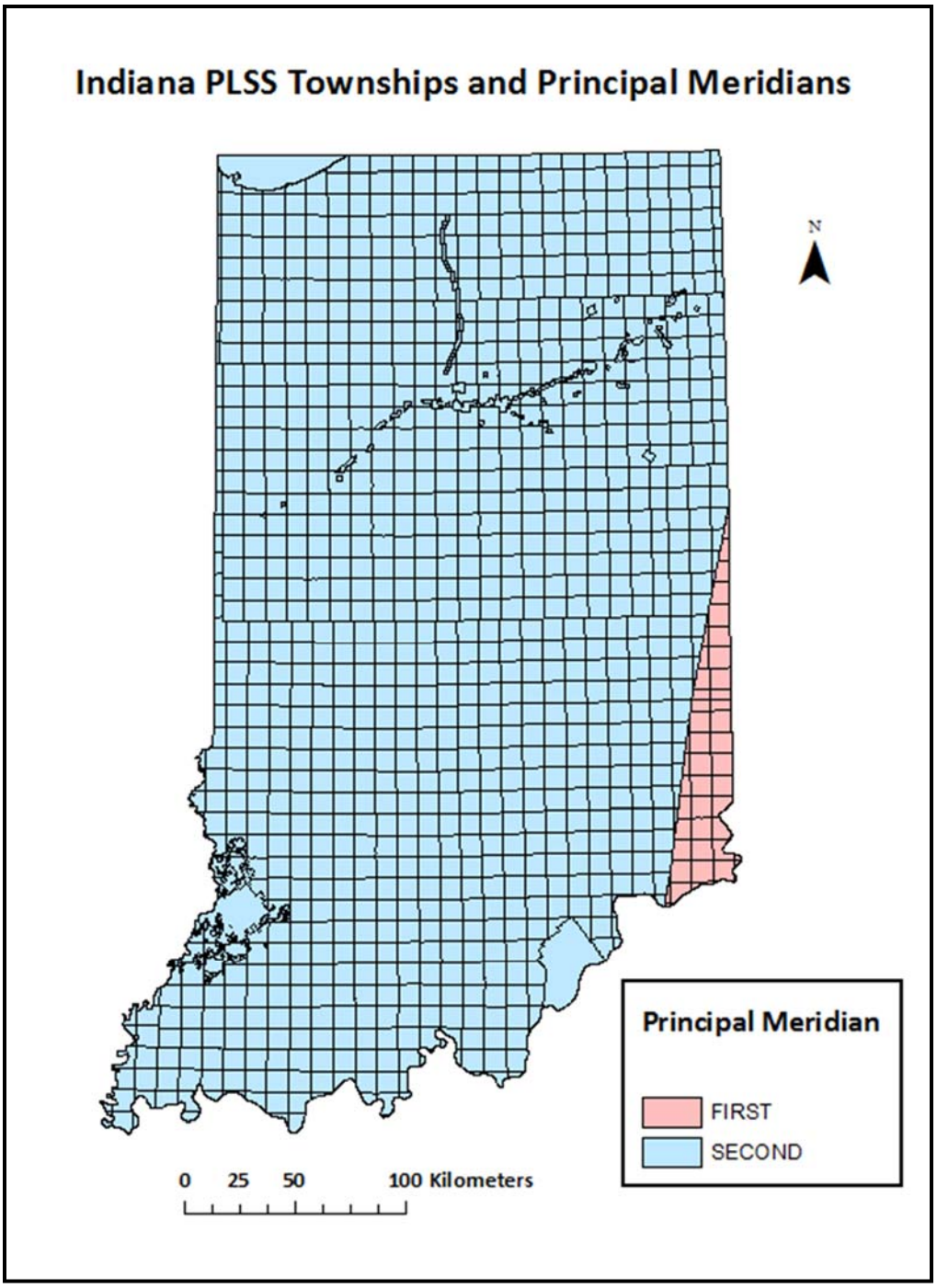


Figure 4.3. PLSS Township grid used to populate missing principal meridians in the pedon site descriptions. The majority of Indiana is on the United States Public Land Survey System grid, but some areas were surveyed through a different survey before the PLSS was implemented. These land units appear as irregular shapes on the map.

Figure 4.4. Figure showing the spatial extent of the NASIS pedon data. The arrows point to pedons whose coordinates place them outside of the state and even outside of the country. Errors like these could not be corrected assuredly, so these data were removed from the study.

4.2.2 Creating a More Accurate PLSS Grid

The positional accuracy of the pedon points relies heavily on a detailed, georeferenced PLSS grid for Indiana. Even for points that are collected with a GPS, it is helpful to have an alternative georeference base for areas with little or no satellite signal which can allow for validation of the GPS reading. The PLSS grid currently available for Indiana does not support land parcels smaller than sections, which limits the degree of positional accuracy that can be achieved for pedon locations described beyond a PLSS section. An accurate, more detailed PLSS grid for Indiana was developed for the sole purpose of improving the positional accuracy of the soil pedon points. Moreover, the grid can also serve as a multi-purpose reference map for Indiana.

The multi-step process of understanding the origin of the information required an investigative process. After a teleconference with Thomas Reinsch and a better understanding of the strengths and limitations of the current PLSS Conversion program for NRCS, it became apparent that improving the positional accuracy of pedon data would have to begin with an improvement to the digitized PLSS grid for Indiana. Research investigations were conducted to learn about the development of the PLSS grid for Indiana and understand the details of how each of the PLSS components were surveyed with the goal of tracing surveyors' steps and generating a more detailed PLSS grid.

Albert White's, *A History of the Rectangular Survey System* (1926) provided the records and documentation that were key in making the new PLSS grid for Indiana. In his report, White provides the meticulous details of how Indiana's land was divided and surveyed into the PLSS grid. The General Land Office of the United States was responsible for setting the survey regulations. The Surveyor General of the United States provided thorough instructions to his deputies for completing the surveys which included a description of the proper use of surveying instruments, how to adjust for the variation of the compass to correctly follow the meridians, how to run and mark the

lines and establish corners to delineate sections, and how to subdivide townships into sections.

Two sets of instructions exist for surveying the rectangular system in Indiana. The first was published in 1833 and does not account for the correction lines installed to counter the convergence of the meridians. The second was published in 1850 and does consider the correction lines. The following is an excerpt from the 1850 report from the *General Instructions to His Deputies By the Surveyor General of the United States, for the States of Ohio, Indiana, and Michigan* (White, 1926). This excerpt explains how to subdivide townships and mark quarter corners:

> "Each side of a section must be made one mile in measure by the chain. Quarter section corners are to be established at every half mile, except in closing a section, when the closing line varies from eighty chains or one mile; in which case you are to place in closing a section, when closing the quarter section corner equidistant … from the corners of the section. But in running out the last section lines, to the north and west boundaries of the township, the quarter section corners are to be established at the distance of forty chains from the last section corner, and the excess or deficiency of measure (if any) carried out into the last half mile, and east upon the north and west sides of the township, as required by law.
>
> You will begin on the east boundary of the township, at the corner of sections 13 and 24, and run and measure a random line west, or parallel to the south boundary, to the west boundary of the township, and note your intersection, whether at, or north, or south of the corner of sections 18 and 19, and if not at that corner, how far from it. On this random line you will set temporary section and quarter section posts; and also set stakes, or make some other marks, at all the even tallies, or outs, between those posts. From the corner of sections 18 and 19, on the west boundary, you will then return on the true line, straight towards the corner where you commenced the random [line], blazing and marking that line, and verifying its course by means of off-sets from

the posts and stakes set, or other marks made, on the random line, and mark and establish the proper section and quarter section corners thereon.

From the corner of sections 13, 14, 23 and 24, run and measure a random line south, or parallel to the east boundary, to the south boundary of the township, and note the intersections thereof, whether at, east or west of the corner of sections 35 and 36, and if not at that corner, how far from it. On this random line, as it is run, you will set temporary section and quarter section posts, and make other marks for the even tallies, or outs, as directed on the random line through the middle of the township. From the corner of sections 35 and 36, on the south boundary, you will return on the true or direct line, blazing and marking that line, and establishing the quarter section and section corners thereon, at their average distances, or proportionate parts of the whole distance, to the corner of sections 13, 14, 23 and 24, on the middle line.

You will also run and measure a random line east from the corner of sections 25, 26, 35 and 36, to the east boundary of the township, and note its intersection, whether at, or north, or south of the section corner, and how far from it, and correct, mark and establish this line back to the corner from which you set out, in the manner before directed for the correction of random lines, establishing the quarter section corner thereon equidistant between the section corners. Proceed in like manner with each east and west section line, as you progress north, until you close at the corner of sections 13, 14, 23 and 24.

From this corner, run and measure a random line north, or parallel to the east boundary, to the north boundary of the township, and note its intersection, whether at, or east, or west of the corner of sections 35 and 36 in the township north, excepting where you close on a correction line, in which case you will note the distance east or west to the nearest section or quarter section corner, and establish a corner thereon, for sections 1 and 2, one mile west of the north-east corner of the township, according to the measure of the correction line. In running this random line, posts must be set for temporary section and quarter

> section corners, and stakes or some other marks must be left to indicate the places of all the even tallies, or outs, as before directed in similar cases. From the corner of sections 1 and 2, return on the true line, in the direction of the place of beginning the random line, to the corner of sections 1, 2, 11 and 12, blazing and marking the same as before directed for true lines, and establishing the quarter section corner so as to leave the excess or deficiency of the whole measure in the half mile next to the north boundary of the township. From the corner of sections 1, 2, 11 and 12, run and measure a random line east for its corresponding corner on the east boundary. Note its intersection, and correct back, and establish the quarter section corner on the true line at equal distances between the section corners, blazing and marking the corrected line as before directed. In like manner proceed to run, measure, mark and establish all the subdivision lines on this part of the eastern tier of sections, until you close at the corner of sections 13, 14, 23 and 24.
>
> Proceed in the same manner with each successive tier of sections, to the last, changing the order only so far as necessary, when interrupted by lakes or other interferences. From the section corners on the east side of the last tier, run random lines west for their corresponding corners on the west boundary of the township, note your intersections, correct back from those corners, as directed in other cases, before mentioned, and establish the quarter section corners on the corrected lines at the distance of forty chains from the section corners east of them, so that the excess or deficiency of measure may be thrown into the half mile next to the west boundary, as required by law."

This description was used as the basis for developing a new, more detailed PLSS grid for Indiana. Five steps were taken to convert these written descriptions to a computer model:

1. The *General Instructions* were translated to a visual model representation by tracing the surveyor's steps described in the text over a diagram of a generic

township. In this way, the section corners and boundary dimensions were determined. The visual model is shown in Figure 4.5.

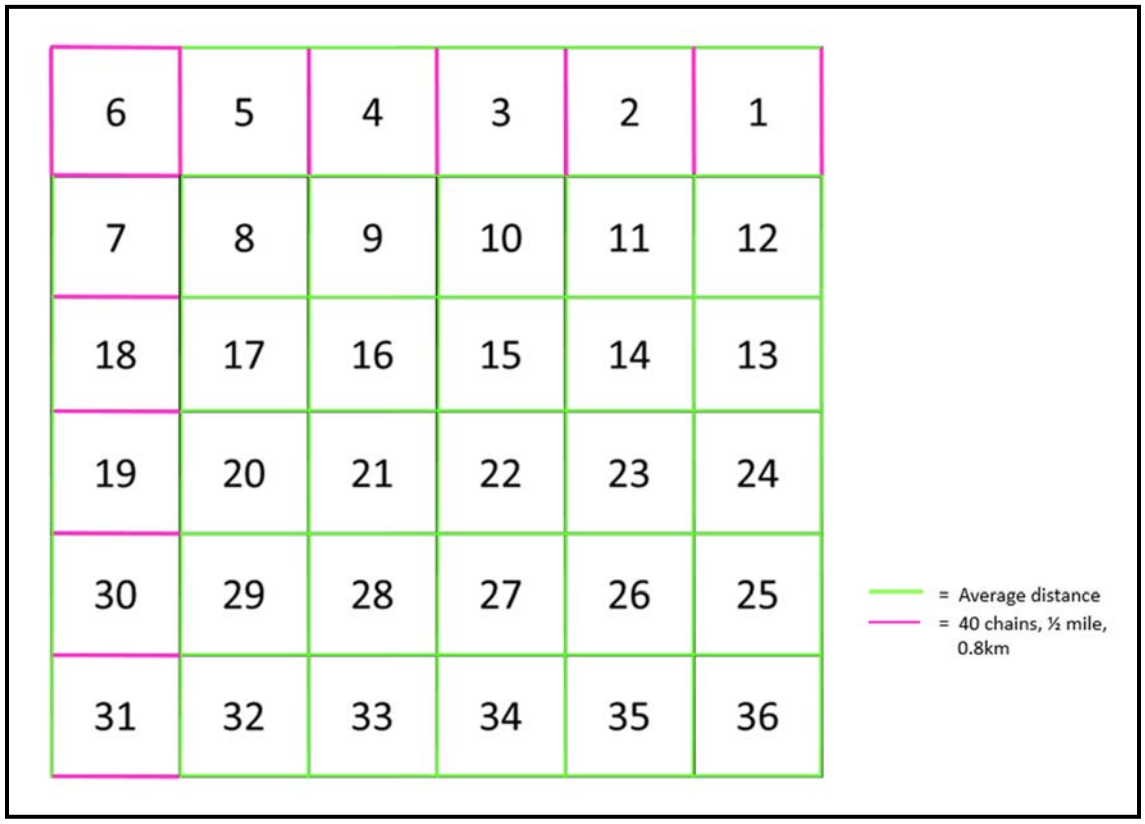


Figure 4.5. Boundary dimensions for the sections of a township of the Indiana PLSS grid as described in the General Surveyor's Instructions to His Deputies.

2. The resulting diagram from Step 1 helped to distinguish four sets of sections, each with different dimensions for their north, south, east, and west boundaries:
    - Section set 1: 8-17, 20-29, 32-36
    - Section set 2: 1-5
    - Section set 3: 7, 18, 19, 30, 31
    - Section set 4: 6

    The corners of the quarter sections in each of the sets were defined mathematically in preparation for entering them into a computer model. The

coordinates of the corners were defined in relation to the section corners and to each other. Figures 4.6 a-d show an example of a section from each set with the functions used to define each of the quarter section corners.

3. The mathematical equations derived in Step 2 were transferred to an ArcPython script to serve as the logic for an automated, computer-based model used to accurately segment the PLSS Sections polygon vector file into quarter sections. The section and quarter section centroids were calculated using the built-in Centroid function in ArcPython. The UTM Zone 16N projected coordinate system was used as the reference base to locate all section corners and centroids. The computer algorithm produced a new polygon file of quarter sections and coordinates for each corner and centroid together with full PLSS component information provided for every section and section subdivision.
4. The PLSS quarter sections produced in Step 3 were used as an input for the computer program in order to generate the X, Y coordinates of the corners of the QQ sections. Likewise, the QQ sections grid was used as in input to generate QQQ section corners. The Centroid function for ArcPython was used to compute the centroids for QQ and QQQ sections. Figures 4.7 a-d show an example of a section of the new PLSS grid generated and its corresponding attribute table. The computer algorithm used to derive the more-detailed PLSS grid for Indiana was written in such a way that every corner and centroid of a Q-, QQ-, and QQQ-section is defined through x and y coordinates in meters and contains information about all of its associated PLSS components.
5. With prior knowledge that some NRCS pedon points were only described to township and range, the vertices and centroids of the IndianaMap Township and Range file (Indiana Geological Survey, 1998) were also generated using the same computer model.

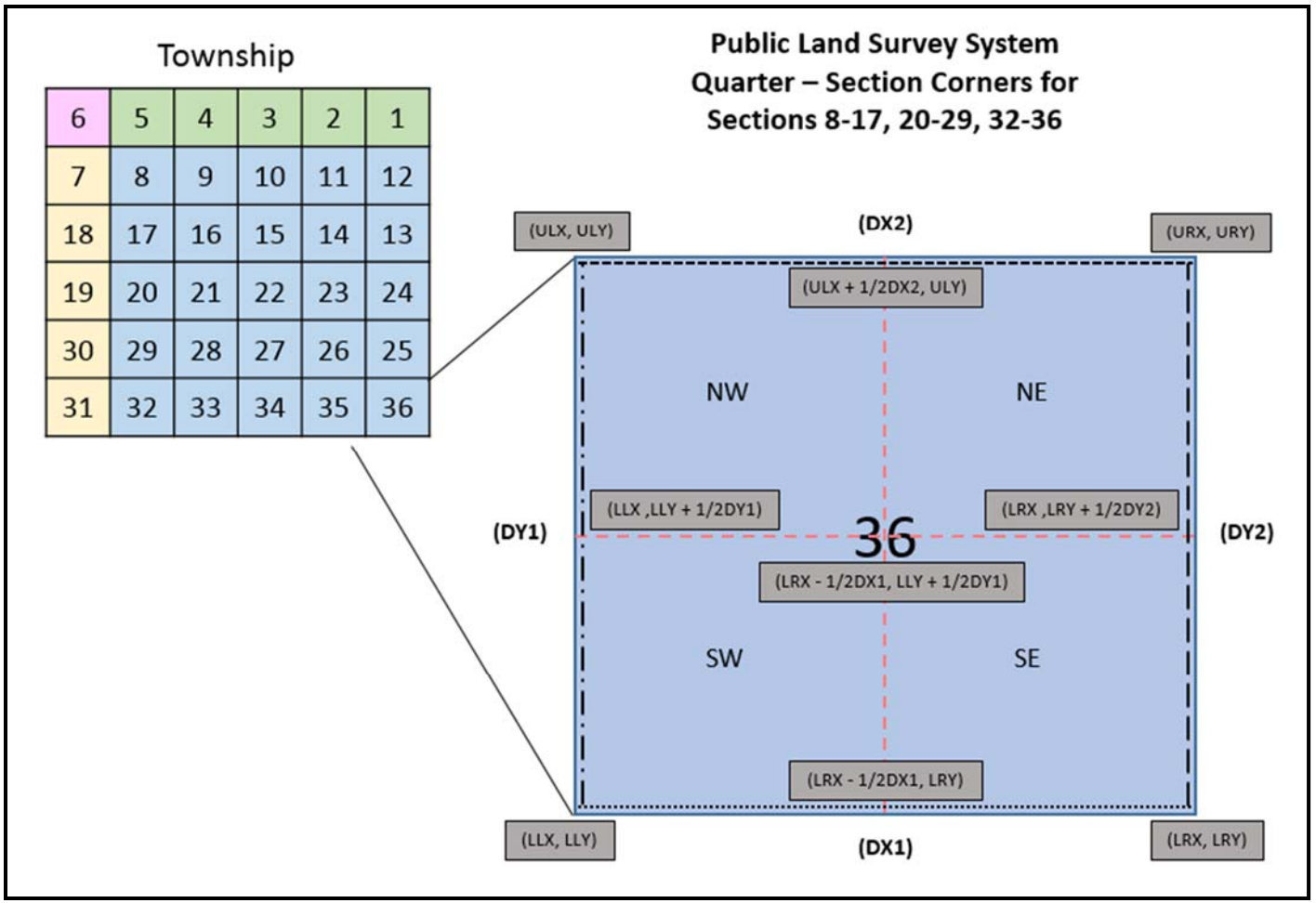


Figure 4.6 a. An example of a section from Set 1 with the formulas for deriving each of the quarter-section corners.

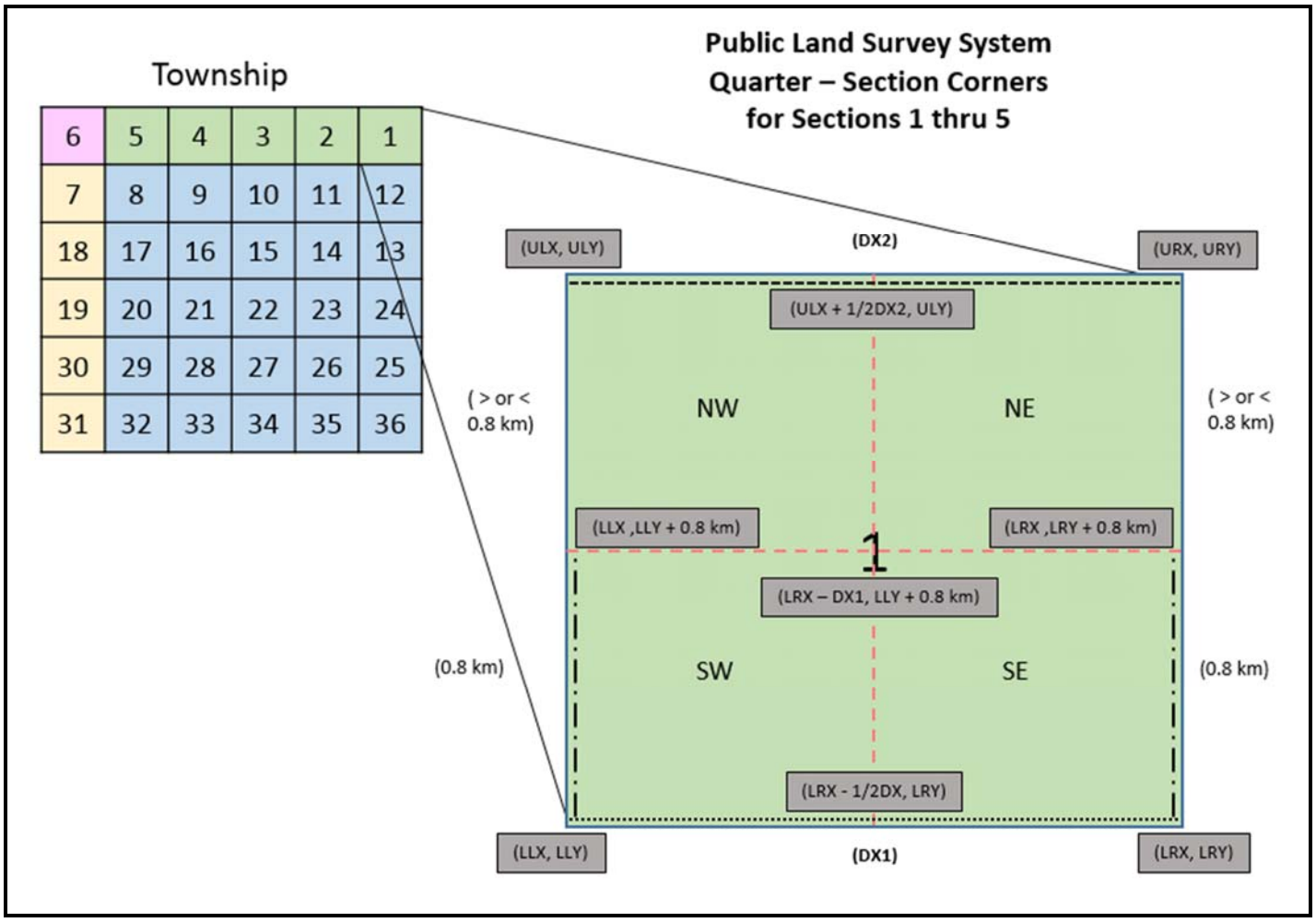


Figure 4.6 b. An example of a section from Set 2 with the formulas for deriving each of the quarter-section corners.

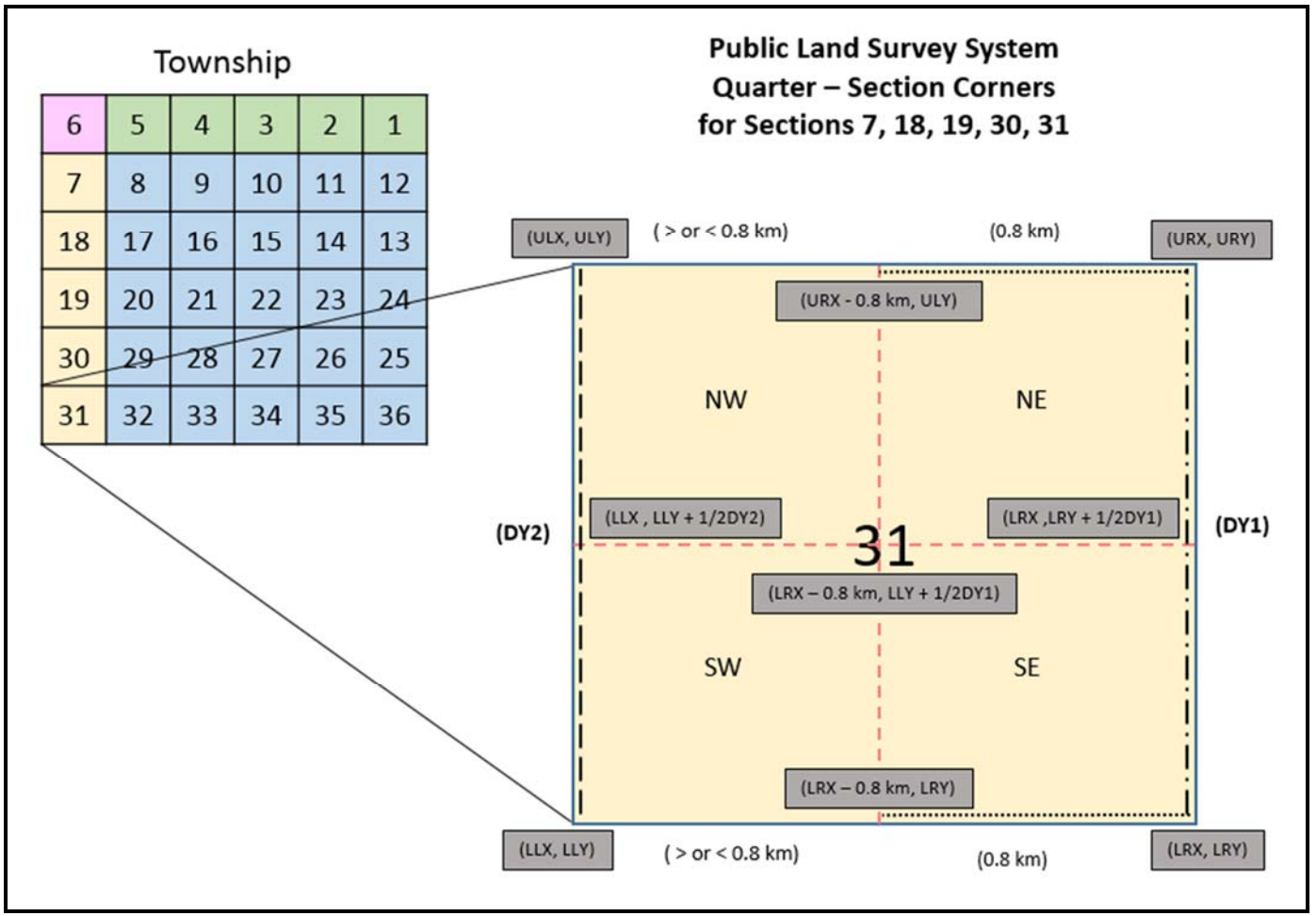


Figure 4.6 c. An example of a section from Set 3 with the formulas for deriving each of the quarter-section corners.

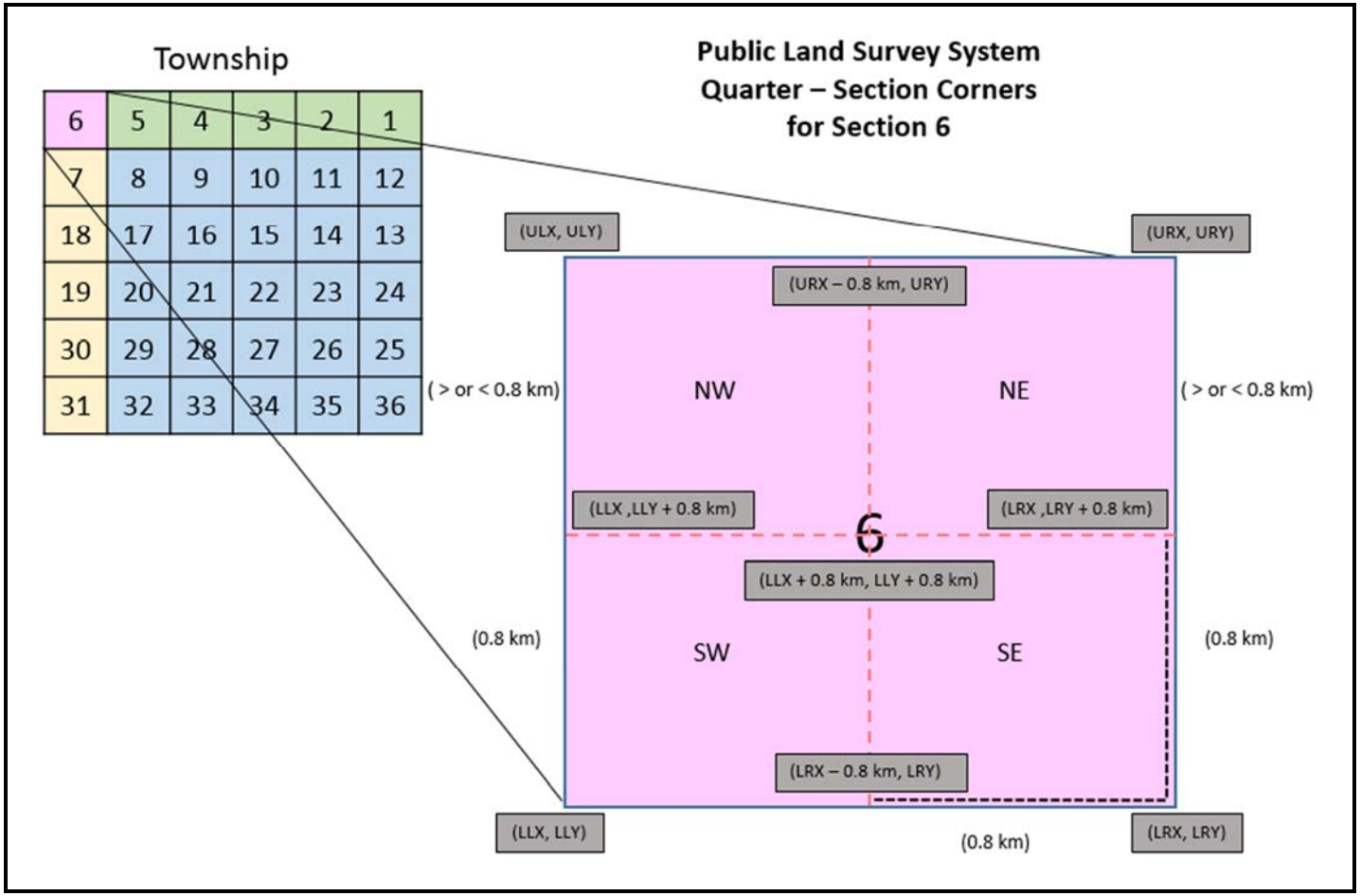
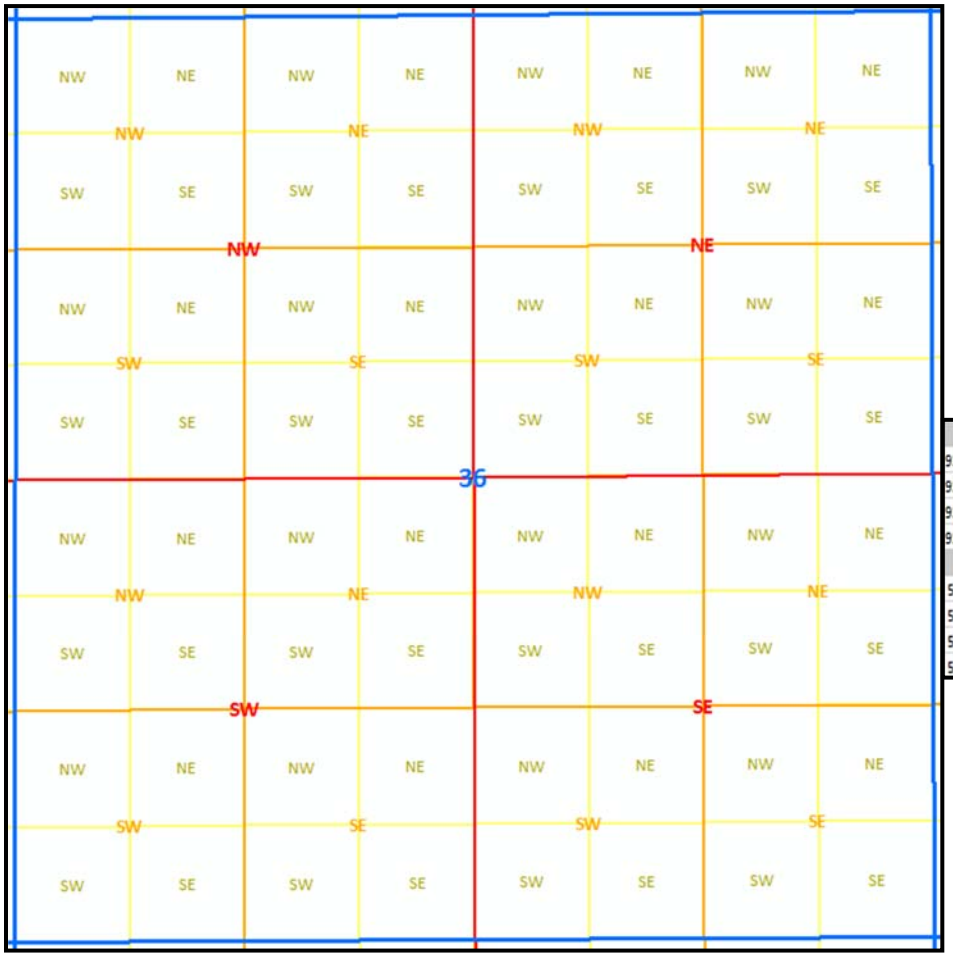


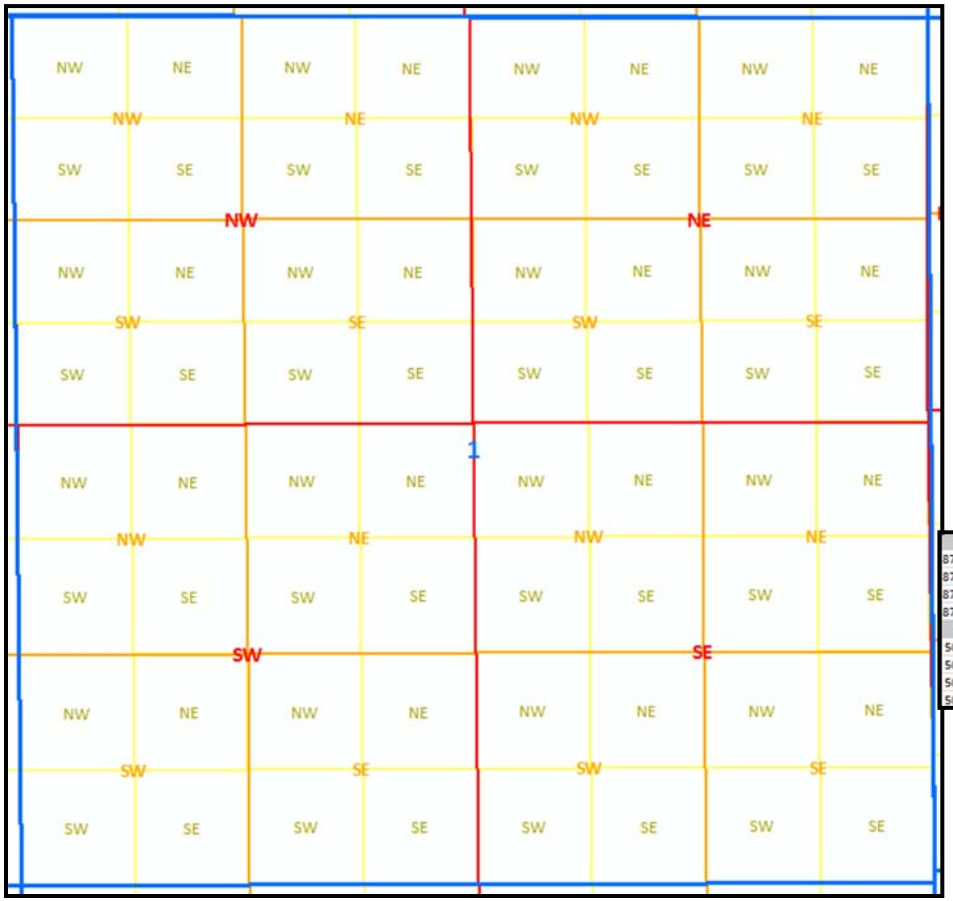
Figure 4.6 d. An example of a section from Set 4 with the formulas for deriving each of the quarter-section corners.



FID	MERIDIAN	SECTION	TWN	TWND	RNG	RNGD	QUART_ID	Q_Q_ID	Q_Q_Q_ID
923264	SECOND	36	22	N	5	W	NW	NW	NW
923265	SECOND	36	22	N	5	W	NW	NW	NE
923266	SECOND	36	22	N	5	W	NW	NW	SW
923267	SECOND	36	22	N	5	W	NW	NW	SE

NW_X	NW_Y	NE_X	NE_Y	SW_X	SW_Y	SE_X	SE_Y	CENTROID_X	CENTROID_Y
504916.49	4462818.60	505116.16	4462820.15	504916.16	4462616.40	505116.01	4462617.93	505016.21	4462718.26
505116.16	4462820.15	505315.83	4462821.70	505116.01	4462617.93	505315.85	4462619.46	505215.96	4462719.79
504916.16	4462616.40	505116.01	4462617.93	504915.84	4462414.20	505115.85	4462415.71	505015.97	4462516.05
505116.01	4462617.93	505315.85	4462619.46	505115.85	4462415.71	505315.86	4462417.21	505215.89	4462517.56

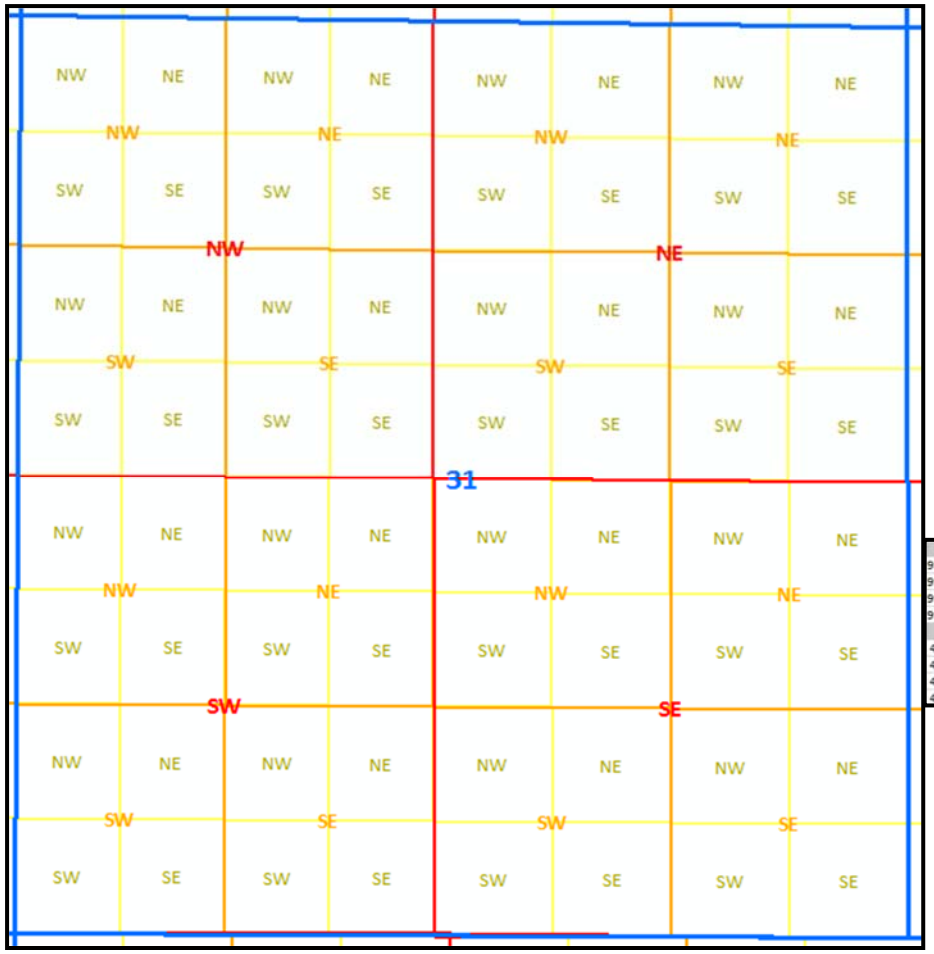
Figure 4.7 a. Section 36 of Set 1 from the new PLSS grid showing Q, QQ, and QQQ sections and the attribute table information for the NW ¼, NW ¼, NW ¼ Sec 36, T22N R5W, 2nd Principal Meridian subsection.



FID	MERIDIAN	SECTION	TWN	TWND	RNG	RNGD	QUART_ID	Q_Q_ID	Q_Q_Q_ID
877504	SECOND	1	22	N	5	W	NW	NW	NW
877505	SECOND	1	22	N	5	W	NW	NW	NE
877506	SECOND	1	22	N	5	W	NW	NW	SW
877507	SECOND	1	22	N	5	W	NW	NW	SE

NW_X	NW_Y	NE_X	NE_Y	SW_X	SW_Y	SE_X	SE_Y	CENTROID_X	CENTROID_Y
504861.22	4470796.49	505060.89	4470796.13	504863.23	4470617.72	505062.83	4470617.65	504962.02	4470707.00
505060.89	4470796.13	505260.56	4470795.78	505062.83	4470617.65	505262.44	4470617.58	505161.65	4470706.79
504863.23	4470617.72	505062.83	4470617.65	504865.25	4470438.95	505064.78	4470439.16	504964.00	4470528.38
505062.83	4470617.65	505262.44	4470617.58	505064.78	4470439.16	505264.31	4470439.38	505163.56	4470528.45

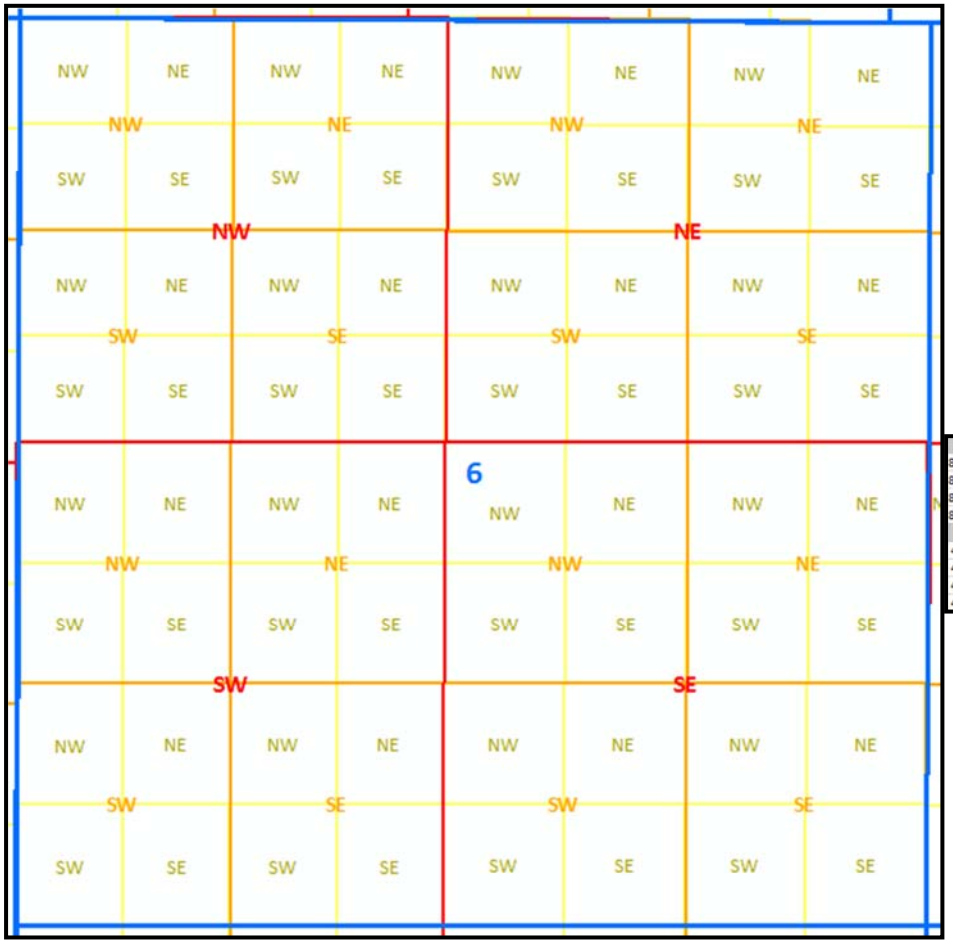
Figure 4.7 b. Section 1 of Set 2 from the new PLSS grid showing Q, QQ, and QQQ sections and the attribute table information for the NW ¼, NW ¼, NW ¼ Sec 1, T22N R5W, 2nd Principal Meridian subsection.



FID	MERIDIAN	SECTION	TWN	TWND	RNG	RNGD	QUART_ID	Q_Q_ID	Q_Q_Q_ID
923392	SECOND	31	22	N	5	W	NW	NW	NW
923393	SECOND	31	22	N	5	W	NW	NW	NE
923394	SECOND	31	22	N	5	W	NW	NW	SW
923395	SECOND	31	22	N	5	W	NW	NW	SE

NW_X	NW_Y	NE_X	NE_Y	SW_X	SW_Y	SE_X	SE_Y	CENTROID_X	CENTROID_Y
496946.05	4462826.68	497121.64	4462824.24	496944.78	4462624.08	497120.72	4462621.83	497033.28	4462724.18
497121.64	4462824.24	497297.24	4462821.80	497120.72	4462621.83	497296.66	4462619.57	497209.05	4462721.83
496944.78	4462624.08	497120.72	4462621.83	496943.51	4462421.49	497119.79	4462419.41	497032.18	4462521.67
497120.72	4462621.83	497296.66	4462619.57	497119.79	4462419.41	497296.07	4462417.34	497208.30	4462519.51

Figure 4.7 c. Section 31 of Set 3 from the new PLSS grid showing Q, QQ, and QQQ sections and the attribute table information for the NW ¼, NW ¼, NW ¼ Sec 31, T22N R5W, 2nd Principal Meridian subsection.



FID	MERIDIAN	SECTION	TWN	TWND	RNG	RNGD	QUART_ID	Q_Q_ID	Q_Q_Q_ID
877184	SECOND	6	22	N	5	W	NW	NW	NW
877185	SECOND	6	22	N	5	W	NW	NW	NE
877186	SECOND	6	22	N	5	W	NW	NW	SW
877187	SECOND	6	22	N	5	W	NW	NW	SE

NW_X	NW_Y	NE_X	NE_Y	SW_X	SW_Y	SE_X	SE_Y	CENTROID_X	CENTROID_Y
496939.26	4470817.30	497117.36	4470817.30	496937.80	4470639.60	497116.02	4470639.65	497027.61	4470728.45
497117.36	4470817.30	497295.46	4470817.30	497116.02	4470639.65	497294.25	4470639.70	497205.77	4470728.47
496937.80	4470639.60	497116.02	4470639.65	496936.33	4470461.90	497114.68	4470462.00	497026.20	4470550.78
497116.02	4470639.65	497294.25	4470639.70	497114.68	4470462.00	497293.03	4470462.09	497204.49	4470550.85

Figure 4.7 d. Section 6 of Set 4 from the new PLSS grid showing Q, QQ, and QQQ sections and the attribute table information for the NW ¼, NW ¼, NW ¼ Sec 6, T22N R5W, 2nd Principal Meridian subsection.

The IndianaMap Sections polygon file (Indiana Geological Survey, 1998) was used as the input file for the computer program described above. The IndianaMap file was chosen as the input grid because it had a well-structured attribute table with information for the townships and ranges and it was already projected to UTM Zone 16N, so no re-projection was required.

4.2.2.1 → Validating the New PLSS Grid for Indiana

The Q, QQ, and QQQ polygon vector files, collectively referred to as the "new PLSS grid," which were developed through the computer algorithm, were validated using the following procedure:

1. Layers with clearly-defined PLSS components were used as a reference for comparison to the Q section boundaries. These included the USGS topographic maps and aerial imagery. Assurance that the aerial imagery provides a ground reference for the PLSS grid lies in the fact that land in Indiana was sold in sections, half-sections, quarter sections, and even quarter-quarter section tracts, therefore county roads and property boundaries usually follow the PLSS grid. The comparison was performed in ArcMap version 10.2.2. The USGS Digital Raster Graphic topographic map was streamed from a WMS Server (http://maps.indiana.edu/arcgis/services/Imagery/Topo_Maps_24K_USGS/MapServer/WmsServer?). Bing ™ aerial imagery for Indiana was added as a base map. The Q, QQ, and QQQ section grids were added to the map as well.
2. A visual comparison was conducted comparing the Q section grid boundaries and their respective boundaries outlined in the topo maps and pictured in the aerial imagery. The same was done for QQ and QQQ sections (Figure 4.8 a-c).
    a. The new PLSS grid generated by the computer program was also compared to a sub-section grid of QQQ sections which had been

derived by the Department of Natural Resources (DNR) from a digitized PLSS sections grid through an equal-area split function (Figure 4.9).

3. To ensure that the math had translated well from the computer code to the output grids, the Ruler tool in ArcGIS was used to measure the lengths and widths of the Q, QQ, and QQQ sections. Some areas across the state were randomly checked and they were consistent with the dimensions outlined in the computer code.

Comparison with the USGS DRG topo map and Bing aerial imagery supported the statement that the new PLSS grid is an accurate representation of the PLSS features present in the reference layers. The new PLSS grid outlines the United States PLSS component features in much greater detail and more uniformly than the reference maps. Quarter sections which are only partly visible in the aerial maps or partially outlined in the topo maps are complete in the new PLSS grid. Considering the description from White (1926), it would be intuitive that the new PLSS grid that was developed, would be far more accurate than the DNR's quarter sections grid derived by splitting the section polygons into equal-area quarters. Again, the new PLSS grid mimics the surveyors' work in delineating the quarter boundaries.

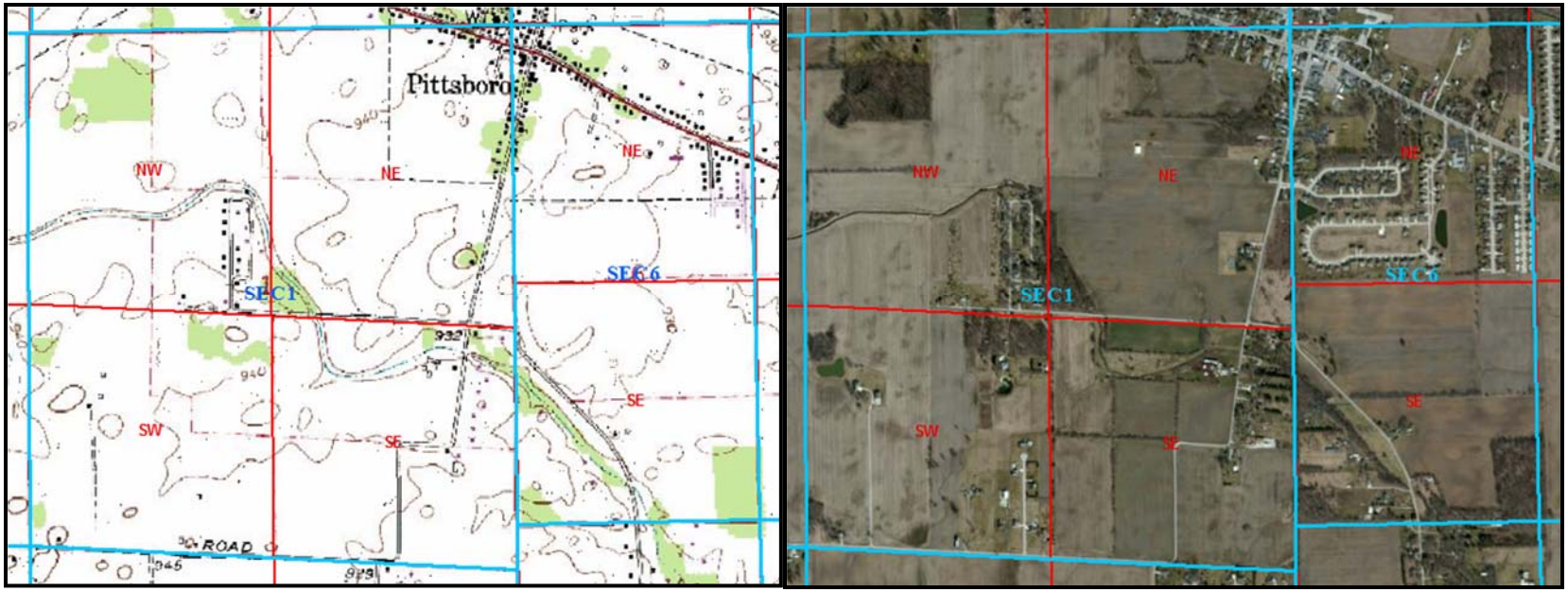


Figure 4.8 a. Quarter sections developed through the model, of Sections 1 and 6 in Pittsboro, IN overlaid on a USGS topo map (left) and aerial image (right).

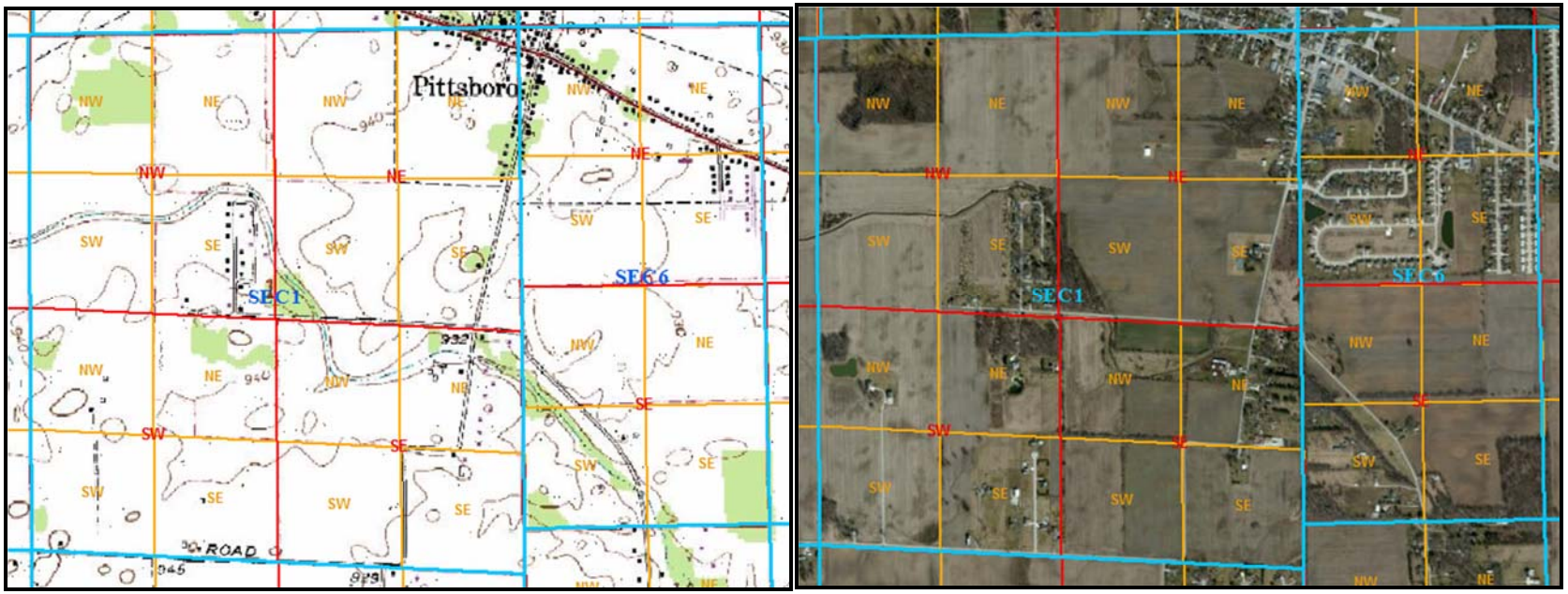


Figure 4.8 b. Quarter Quarter sections developed through the model, of Sections 1 and 6 in Pittsboro, IN overlaid on a USGS topo map (left) and aerial image (right).

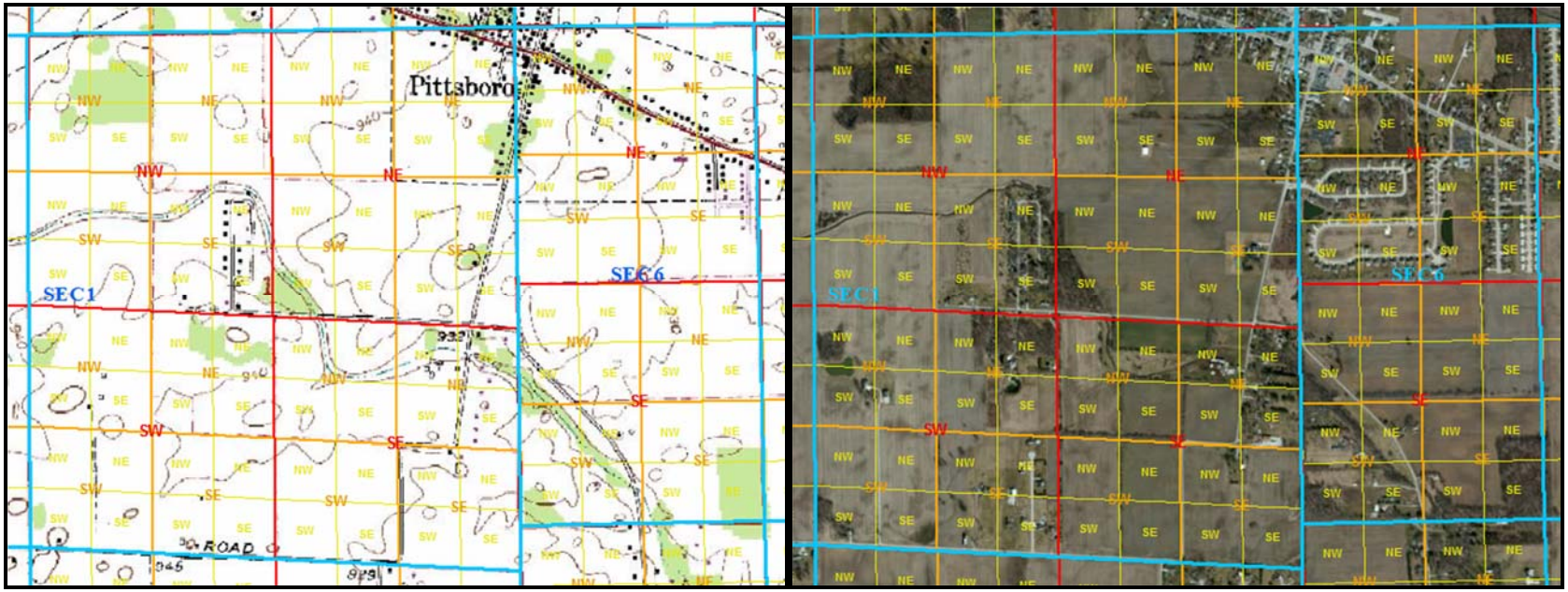


Figure 4.8 c. Quarter Quarter Quarter sections developed through the model, of Sections 1 and 6 in Pittsboro, IN overlaid on a USGS topo map (left) and aerial image (right).

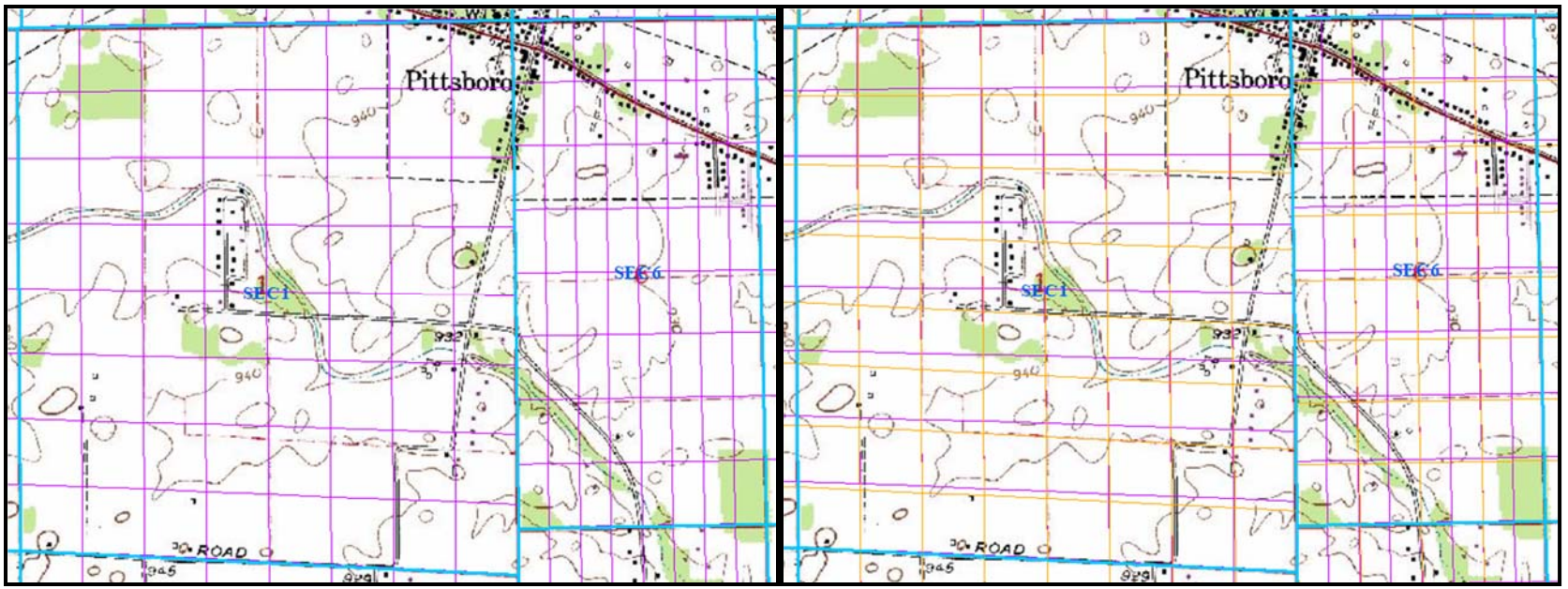


Figure 4.9. The DNR's QQQ section grid, in purple (left) and the DNR's QQQ section grid in purple and the new PLSS QQQ section grid generated, in yellow, overlaid on a USGS topo map (right).

4.2.3 Generating PLSS Model Pedon Point Locations

With an improved PLSS grid with which to locate the pedon sample points, the next step was to restructure the pedon tabular data collected for Indiana in order to link the parsed location details to the PLSS grid component details in the attribute tables of the Q grids through a one-to-one relationship. The column headings for the location descriptions in the NASIS pedon data table were edited to exactly match the column headings in the attribute tables of the new PLSS grid vector files.

A computer model was written in ArcPython to automatically place a temporary X, Y point location at the location described by the PLSS details in the pedon data table. The model read, column by column, the PLSS details of each “usiteid” and matched it to the appropriate column in the new PLSS grid. Through a series of if-then statements, a new, temporary X, Y location was created for each “usiteid” at the most detailed location described up to, and including, the center or corner of a QQQ section. Figure 4.10a is a diagram of the computer model used and Figure 4.10b. is the accompanying set of if-then statements for the model.

This new PLSS Model is a suitable replacement for the current NRCS Conversion Model because it is able to accurately locate a point from the parsed PLSS details. It improves the positional accuracy of points that rely on PLSS details in order to be georeferenced when the location is described to a subsection. The new model is more efficient than the NRCS Conversion Model because it eliminates the need for polyconic transformation equations by using a reference projected coordinate system. The ArcPython script that the new method uses is also easy to share and easy to follow. The code and an associated tool will be published to be shared with the public in the near future. Due to the complexity in the structure of the land tracts that are not part of the PLSS grid, including donations and grants, such areas were not incorporated into the new PLSS Model. Ten points from the full pedon set for Indiana which were originally ascribed to these special areas were excluded from this study.

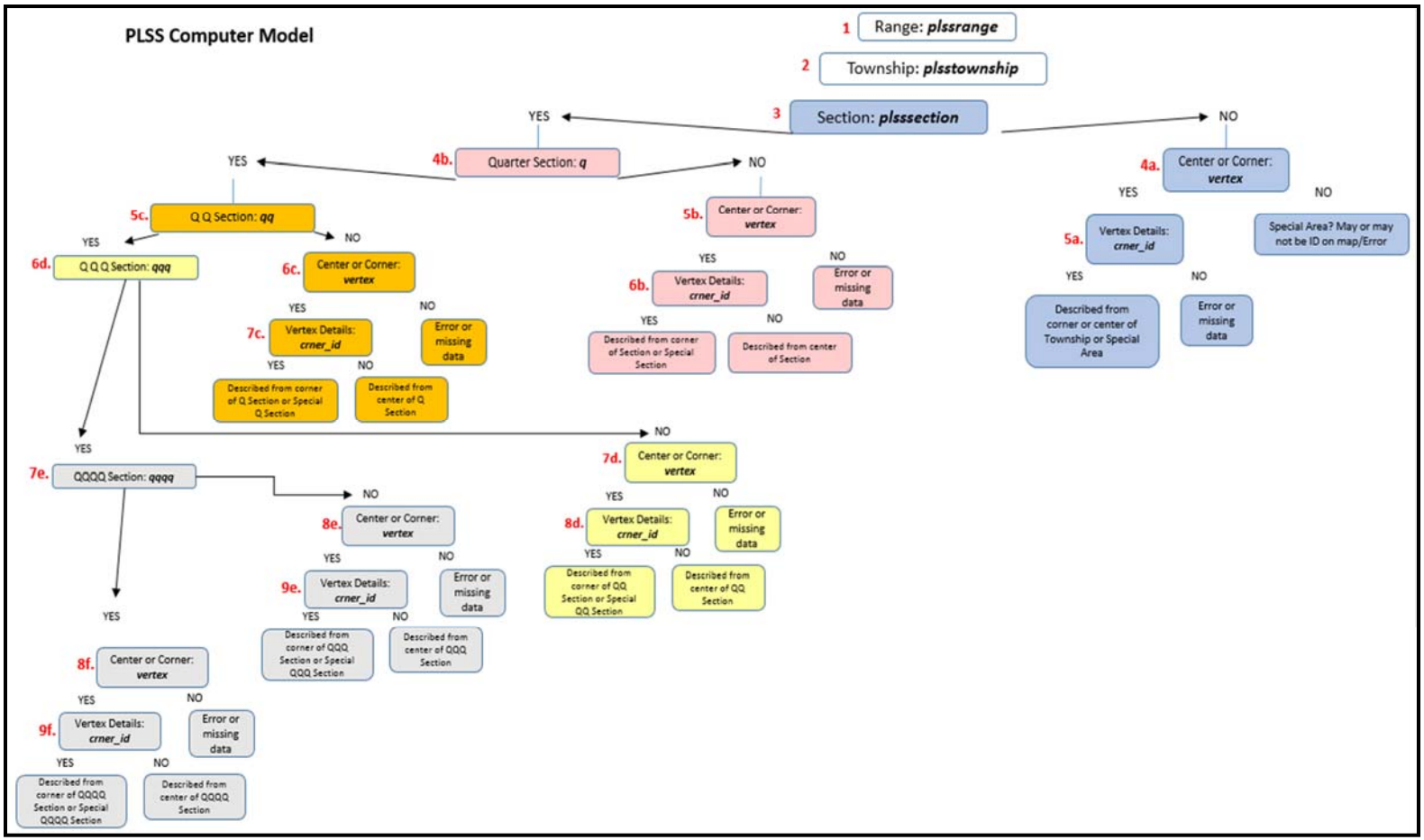


Figure 4.10 a. PLSS Model logic that shows the column data used (in bold) to determine the temporary point location for every level of detail of the location descriptions. The different section subsets are color-coded. Sections are blue, Q sections are pink, QQ sections are orange, QQQ sections are yellow, and QQQQ sections are gray.

Steps for PLSS Model:

1. Locate the Range [plssrange]
2. Locate the Township [plsstownship]
3. Locate the Section [plsssection]
3. If NO [plsssection], then [vertex]
4a. If NO [vertex], then "ERROR"
4a. If YES [vertex], then [crner_id]
5a. If NO [crner_id], then "ERROR"
5a. If YES [crner_id], then "corner"
Move [feet1] in [dir1] and [feet2] in [dir2]

3. If YES [plsssection], then [q]
4b. If NO [q], then [vertex]
5b. If NO [vertex], then "ERROR"
5b. If YES [vertex], then [crner_id]
6b. If NO [crner_id], then "center"
6b. If YES [crner_id], then "corner"
Move [feet1] in [dir1] and [feet2] in [dir2]

4b. If YES [q], then [qq]
5c. If NO [qq], then [vertex]
6c. If NO [vertex], then "ERROR"
6c. If YES [vertex], then [crner_id]
7c. If NO [crner_id], then "center"
7c. If YES [crner_id], then "corner"
Move [feet1] in [dir1] and [feet2] in [dir2]

5c. If YES [qq], then [qqq]
6d. If NO [qqq], then [vertex]
7d. If NO [vertex], then "ERROR"
7d. If YES [vertex], then [crner_id]
8d. If NO [crner_id], then "center"
8d. If YES [crner_id], then "corner"
Move [feet1] in [dir1] and [feet2] in [dir2]

6d. If YES [qqq], then [qqqq]
7e. If NO [qqqq], then [vertex]
8e. If NO [vertex], then "ERROR"
8e. If YES [vertex], then [crner_id]
9e. If NO [crner_id], then "center"
9e. If YES [crner_id], then "corner"
Move [feet1] in [dir1] and [feet2] in [dir2]

7e. If YES [qqqq], then [vertex]
8f. If NO [vertex], then "ERROR"
8f. If YES [vertex], then [crner_id]
9f. If NO [crner_id], then "center"
9f. If YES [crner_id], then "corner"
Move [feet1] in [dir1] and [feet2] in [dir2]

Figure 4.10 b. If-then statements of the PLSS Model. The numbers coincide with the logical steps outlined in Figure 4.10 a.

4.2.3.1 → Comparing the PLSS Model Point Locations to the NRCS Locations

In order to determine whether or not the new PLSS grid actually significantly changed the location of the points, the 4141 points located through the new PLSS Model were compared to the NASIS pedon points located through the NRCS Conversion Model. The decimal degree coordinates are the official locations for the pedon data in NASIS (Henry Ferguson, 2014), so these were used in the comparison. If the decimal degrees were not provided, then the latitude and longitude coordinates were used in their place. Although the assumption remained that the pedon points sampled before 1995 were located using the NRCS Conversion Model and that those sampled after 1995 were located using GPS, all NASIS pedon points were compared to their counterpart located by the new PLSS Model.

The following steps were used in ArcGIS version 10.2.2 to determine the distance between the two sets of points:

1. The Project Tool was used to project all of the points to NAD 83, UTM Zone 16N.
2. The Convert Coordinate Notation Tool was used to convert the coordinate locations to X, Y locations in meters.
3. The Distance Formula was entered into an ArcPython script to automatically tabulate the Euclidean distance between the PLSS model points and NRCS points:

$$C = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

where:

x_A = the x-coordinate of the PLSS model point

y_A = the y-coordinate of the PLSS model point

x_B = the x-coordinate of the official NRCS point

y_B = the y-coordinate of the official NRCS point

c = the distance between the two points in meters

The resulting distance in meters was populated as a new column in the NRCS pedon points table.

4.2.4 Matching the Soil Environment Described

After producing a new, improved PLSS Model that plotted the pedon points based on PLSS details, it was necessary to determine the accuracy of the point locations in relation to the original soil described at the sampling site. In order to do this, a set of pedon site description details was used to define the "soil environment." The pedon soil environment was then compared to corresponding data in the SSURGO database. This logic was transferred to an automated computer program, referred to as the "Soil Environment Match Model," which was developed for ArcGIS. The Soil Environment Match Model was used to determine whether or not a match existed between the soil environment of the pedon point and the map unit at the pedon's official NRCS location and at its PLSS Model location.

It was necessary to define a soil environment as a means of assessing the accuracy of the NASIS pedon point location. The soil environment is comprised of the drainage class, parent material, and taxonomic classification, including order, suborder, great group, subgroup, and family particle size of the pedon point. This static set of site descriptors was used to define the soil environment, because it is constant and stable over time. The columns that define the soil environment were identified in the NASIS pedon data and their counterparts were extracted from the gridded SSURGO (gSSURGO) "component" attribute table.

The gSSURGO map file is a rasterized version of the SSURGO polygon vector file, which is a digitized representation of county soil surveys conducted at the finest scale available for soil survey (Soil Survey Division Staff, 1993). Because it is in raster form, every cell of the gSSURGO grid has attribute data from the soil survey. In order to define the soil

environment for each map unit key, the “mukey” (mukey) was linked to the map unit’s component attribute data through a many-to-one relationship. Up to three major components are defined for each “mukey” and each may have a different set of soil environments. A table was created that contains mukeys and their associated soil environments. The unique values for the soil environment columns from the gSSURGO and NASIS pedon data were extracted and compared for inconsistencies in the data set. Similar parent material types were grouped and reclassified in both data sets in order to prevent mismatches with classes that were the same parent material. Table 4.2 lists the original parent material type and the reclassified parent material group adopted by the Soil Environment Match Model.

Table 4.2. Consolidated parent materials used by the Soil Environment Match Model and the original parent materials from SSURGO and NASIS.

Unique Parent Material Model	Unique Parent Material SSURGO	Unique Parent Material NASIS
Alluvium	Alluvium	Alluvium
	Slope Alluvium	Slope Alluvium
Anthropogenic	Coal extraction mine spoil	Solid rock
	Mine spoil or earthy fill	
Colluvium	Colluvium	Colluvium
Lacustrine Deposit	Lacustrine deposit	Lacustrine deposit
Loess	Loess	Loess
	Eolian deposits	Eolian deposits
		Noncalcareous Loess
Organic Deposit	Organic material	Organic material
	Herbaceous organic material	Herbaceous organic material
	Marl	Marl
	Coprogenic material	Marine deposits
Outwash	Glaciofluvial deposits	Glaciofluvial deposits
	Glaciolacustrine deposits	Glaciolacustrine deposits
	Outwash	Outwash
Pedisediment	Pedisediment	
Residuum	Residuum	Residuum
		Interbedded sedimentary
		Limestone
		Limestone-shale
		Sandstone
		Sandstone-shale
		Sandstone-siltstone
		Shale
		Shale-siltstone
		Siltstone
Sand	Beach sand	Beach sand
	Eolian sands	Eolian sands
Till	Ablation till	Ablation till
	Basal till	Basal till
	Drift	Drift
	Till	Till

In preparation for the model, the pedon data table was organized so that each row contained a single, unique pedon sample and its associated location and soil environment details. The model input was the complete pedon data table and it ran on one row at a time. The Soil Environment Match Model automatically determined whether or not the soil environment at the pedon point location matched the soil environment described by the soil scientist who described the site, using the following steps:

1. The mukey values at the new PLSS Model and NASIS points were extracted using the Extract Value to Point tool in ArcGIS version 10.2.2.
2. The column values for drainage class, parent material, taxonomic order, suborder, great group, sub group, and family particle size were compared one-by-one to the full set of values for that column in the map unit table from gSSURGO.
3. If a match was found between the pedon data column value and one or more of the gSSURGO column values, then a value for "TRUE" was assigned in a new column labeled after the soil environment column and "_match" (i.e. drainagecl_match, pmkind_match, etc.). If no match was found, then "False" was assigned.

The number of pedons with full matches to the soil environment as well as partial matches was computed for each set and the results will be presented in the following section. The pedon points generated through the new PLSS Model had only eleven full matches to the soil environment, so a decision was made to only explore the positional accuracy of the NASIS pedon point locations further.

4.2.5 Measuring Positional Accuracy

Pedon points with full matches to the soil environment were considered to be in an accurate location and their official NASIS pedon location was not changed. Points with one or more mismatches in the soil environment components were assumed to be in an inaccurate geographic location; however, to measure the extent of their misplacement, they had to be analyzed further and the radius for the match model had to be extended in the search for a matching soil environment. An automated computer algorithm called the "Nautilus Match Model" was developed to find the nearest mukey with a full match to the soil environment and place a new pedon point at that location.

The Nautilus Match Model used the same logic as the Soil Environment Match Model described above, but rather than looking at only one cell's mukey per pedon point, it expanded the search in a clockwise direction out to a neighborhood size of 210-square meters. Where it found a match, it stopped and exported the X, Y coordinates of its position to a new column in the pedon data table along with the mukey.

In preparation for the Nautilus model, the pedon data table was organized so that each row contained a single, unique pedon sample and its associated location and soil environment details. The model input was the NASIS pedon data table. The model was run in ArcGIS version 10.2.2. The Nautilus Match Model performed the following steps on one row at a time, to find the closest location with a full set of matching soil environment components, within a 100-square meter radius of the mismatched pedon point location:

1. The model computed the coordinate extent (X left, X right, Y top, Y bottom) of a 200-meter neighborhood around the NASIS pedon point location. It then clipped the gSSURGO raster to this extent using the Clip function in ArcGIS. This was done to reduce the file size that the model had to load while it performed the search for a match. Because the gSSURGO is a very large raster file, clipping it greatly reduces its size and increases the model efficiency.
2. Once the raster was clipped, the search began for an mukey with a matching soil environment. One by one, pixels surrounding the pedon point were searched in a clockwise fashion as shown in Figure 4.11 using the same steps from the Soil Environment Match Model. The search was moved from pixel to pixel by moving a certain distance + or – X and + or – Y (Figure 4.12). Once the model finished searching the pixels immediately surrounding the original pedon location, it moved up to the next tier of pixels. The number of pixels searched increased exponentially by a factor of 8 as the search moved up a tier. The model took a longer time to run the more pixels it searched.
    a. In order to increase the efficiency and speed of the model, any time that it came across a pixel with an mukey identical to the original,

mismatched mukey or any previously encountered mismatched mukey, it recognized that it would not be a match and it did not continue to compare the soil environment columns for matches; it simply moved on to the next pixel.

3. If a match to the soil environment described was found, the X, Y location was printed into the “xy_match” column and the matching mukey was printed into a column named “Nautilus_match.” If the model reached the last pixel defined by the X and Y functions (# 441), without finding a match, the original NASIS pedon coordinates and mukey were output into the “xy_match” and “Nautilus_match” columns, respectively.

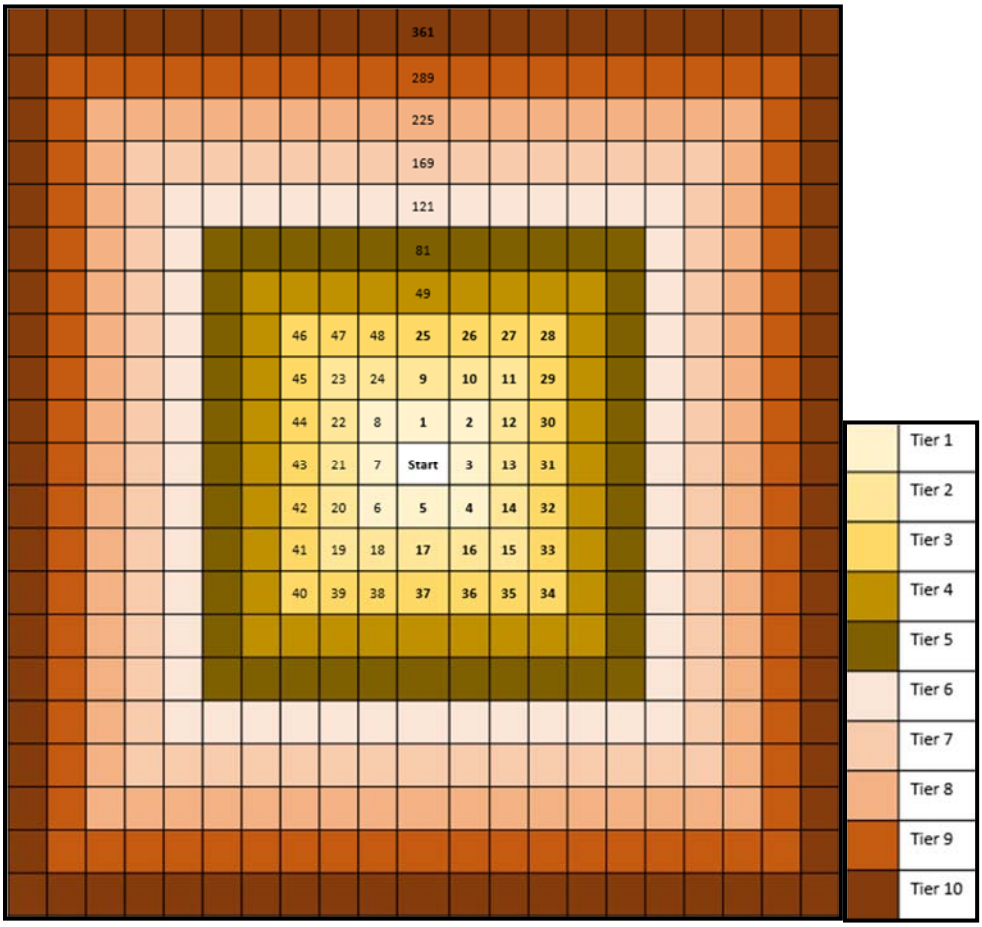


Figure 4.11. The Nautilus Match Model search logic. The numbers represent the order that the pixels were searched. Each tier is colored differently.

x-30, y+30	x-20, y+30	x-10, y+30	y+30	x+10, y+30	x=20, y+30	x+30, y+30
x-30, y+20	x-20, y+20	x-10, y+20	y+20	x+10, y+ 20	x+20, y+20	x+30, y+20
x-30, y+10	x-20, y+10	x-10, y+10	y+10	x+10, y+10	x+20, y+10	x+30, y+10
x-30	x-20	x-10	Start/ NRCS point	x+10	x+20	x+30
x-30, y-10	x-20, y-10	x-10, y-10	y-10	x+10, y-10	x+20, y-10	x+30, y-10
x-30, y-20	x-20, y-20	x-10, y-20	y-20	x+10, y-20	x+20, y-20	x+30, y-20
x-30, y-30	x-20, y-30	x-10, y-30	y-30	x+10, y-30	x+20, y-30	x+30, y-30

Figure 4.12. Mathematical equations for tiers 1-3 used to move the pedon point a certain distance in the X and Y coordinate in order to search pixels in a clockwise fashion.

4.2.6 Developing a Continuous Soil Organic Carbon Map

An objective of this study was use the updated pedon points to interpolate soil organic carbon (SOC) and produce a statewide SOC map of Indiana. Two kriged maps were generated using the VESPER™ version 1.6 spatial prediction software (Whelan, et al., 2002). Data for the total carbon of the first horizon was downloaded from NASIS and linked to the pedon data. Only data from the first horizon was used because this was the most complete carbon data for the NASIS pedon data set. One kriged map was made using the new pedon point locations derived from the Nautilus Match Model and another map was made using the original locations of those same pedons. In this way, the results would show if improving the point locations to match the soil environment impacted the spatial predictions of soil property values.

4.3 Results

Results of the PLSS Model, Soil Environment Match Model, Nautilus Match Model, and continuous soil organic carbon maps will be presented for the pedon data for Indiana. A total set of 4141 pedon points and their respective soil site descriptions were extracted from NASIS and used as the inputs for the models in this study. For analysis purposes, the full set of pedon data was stratified by the year that it was sampled. Assumptions were made, for each pedon subset, as to the source used to georeference the pedon site location in NASIS (Table 4.3). There is not sufficient information to assign an exact measure of accuracy to each georeferencing source; however, given the functional characteristics of each, it is accurate to assume that there may be a significant relative difference in the accuracy of the pedon points derived from each source.

It is important to briefly discuss the Wide Area Augmentation System (WAAS), because it improved GPS accuracy significantly. WAAS is a network of satellites and wide area ground reference stations that enable GPS signal correction. When it became available in 1999, it was widely implemented in Garmin's and other company's receivers in North America (Garmin Ltd., 2014) and provided accuracy to about 3 meters horizontally and vertically. Before 1999, most civilian GPS units relied on other systems for GPS positional accuracy, and the best accuracy was around 5 meters horizontally and vertically.

Table 4.3. List of pedon data subsets stratified by year that they were sampled and by their associated georeference source.

Pedon Data Subset	Source of Coordinates	Total Number of Points
NASIS pedon point, year < 1995	Public Land Survey System NRCS Conversion Model	2343
New PLSS pedon point, year < 1995	Public Land Survey System NRCS Conversion Model	2343
NASIS pedon point, 1995 ≤ year < 1999	GPS, No WAAS correction	101
New PLSS pedon point, 1995 ≤ year < 1999	GPS, No WAAS correction	101
NASIS pedon point, year > 1999	GPS, With WAAS correction	1697
New PLSS pedon point, year > 1999	GPS, With WAAS correction	1697

4.3.1 PLSS Model Pedon Point Locations

As a first step towards developing a methodology to improve the positional accuracy of pedon point data for Indiana, the existing model that assigns coordinate locations to pedon points that were sampled without GPS, was analyzed. Using our understanding of the structural framework of the PLSS grid for Indiana, a new, improved PLSS grid was developed. An automated model was developed to extract the location details from pedon site descriptions and plot the pedon point at the most accurate location possible on a PLSS grid. Figure 4.13 outlines the PLSS Model logic used to locate a pedon point using its PLSS details. The table embedded in the figure contains the total number of pedons described to the indicated level of detail of the PLSS grid. The new pedon locations derived from the PLSS Model are shown on the map in Figure 4.14.

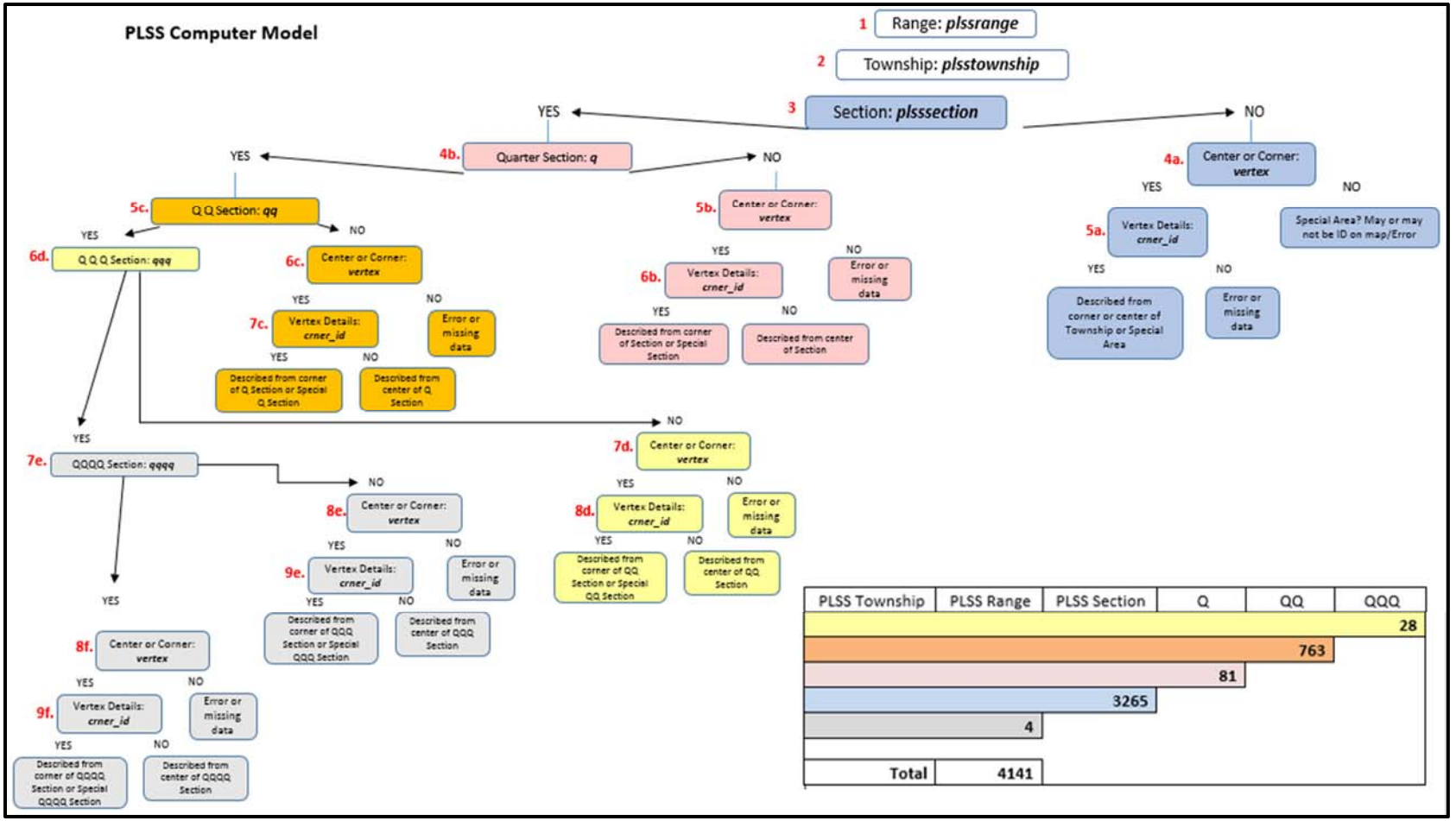


Figure 4.13. This is a diagram of the PLSS Model logic produced to locate a pedon point using its PLSS details. The table embedded in the figure contains the total number of pedons described to the indicated level of detail of the PLSS grid.

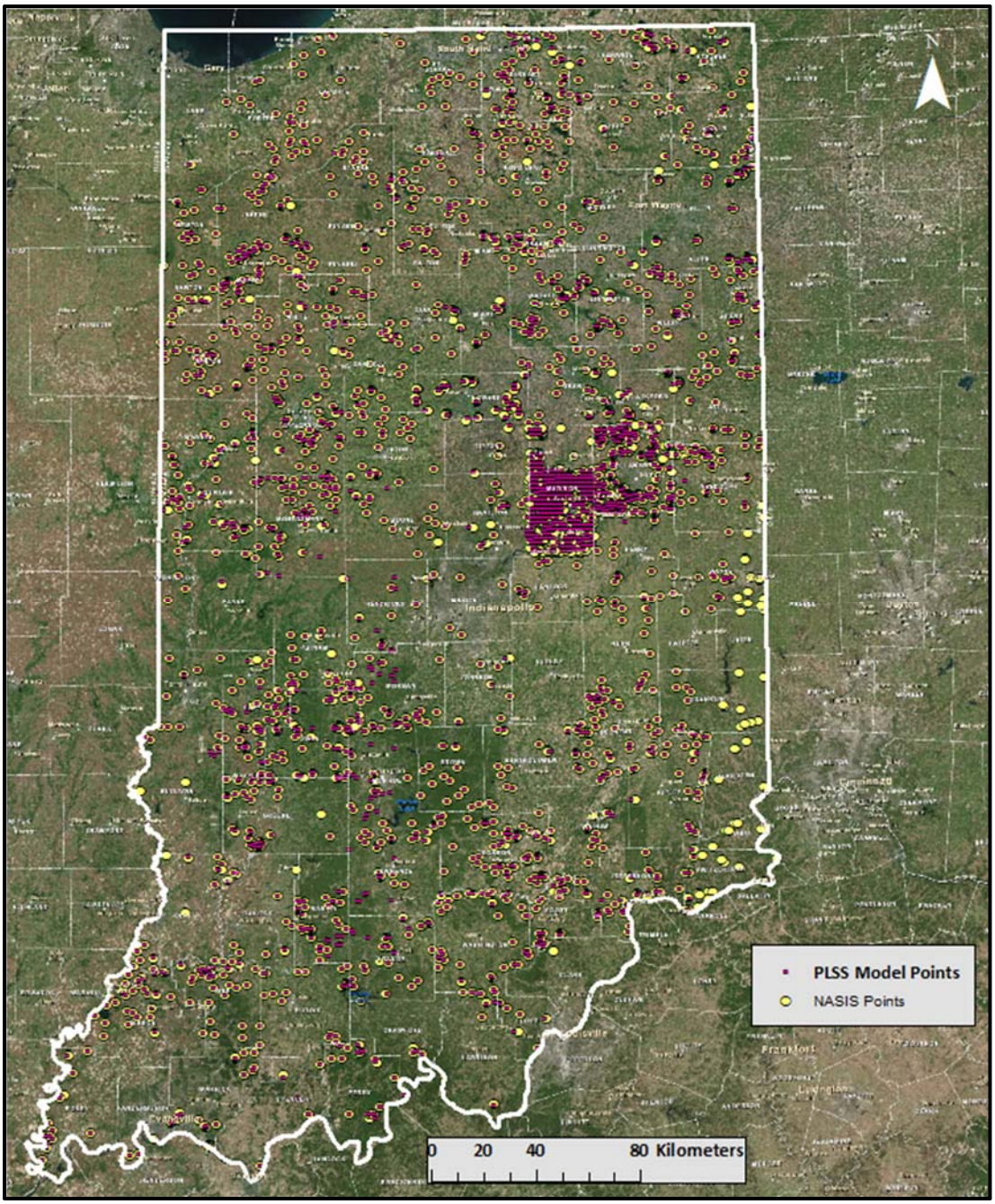


Figure 4.14. Map of the full set of new PLSS Model pedon points and the full set of NASIS pedon points.

Unlike the original NASIS pedon points, all of the pedon locations generated by the PLSS Model placed the pedons within the boundaries of the state. The structure of the PLSS Model is such that it only plots points inside of the smallest land tracts described in

its PLSS details. If the details do not match an area in Indiana, the model gives an error and does not generate a point.

A distance formula was used to compare the NASIS pedon point locations to the PLSS Model point locations. This distance provided a relative accuracy for the pedon points as well as a way to assess the spatial deviation of the NASIS pedon locations from the PLSS details for that pedon. Summary statistics, including a boxplot of the distance data, were calculated for each georeference source category, in order to understand the data distribution. These are presented as Tables 4.4 a-e, and Figure 4.15, respectively.

Tables 4.4 a-e. Summary Statistics for each category of Indiana pedon points.

a.

Statistics: All Points	
Range	4454443
Min	0.461
Max	4454443
Mean	4880
Std. Dev.	86331

b.

Statistics: All Points Sampled with GPS	
Range	4454440
Min	2.758
Max	4454443
Mean	5526
Std. Dev.	128006

c.

Statistics: Points Sampled without GPS	
Range	434510
Min	0.461
Max	434510
Mean	4383
Std. Dev.	24451

d.

Statistics: Points Sampled with GPS, with WAAS Correction	
Range	4454438
Min	4.719
Max	4454443
Mean	5731
Std. Dev.	131734

e.

Statistics: Points Sampled with GPS, No WAAS Correction	
Range	87461
Min	2.758
Max	87464
Mean	2079
Std. Dev.	10144

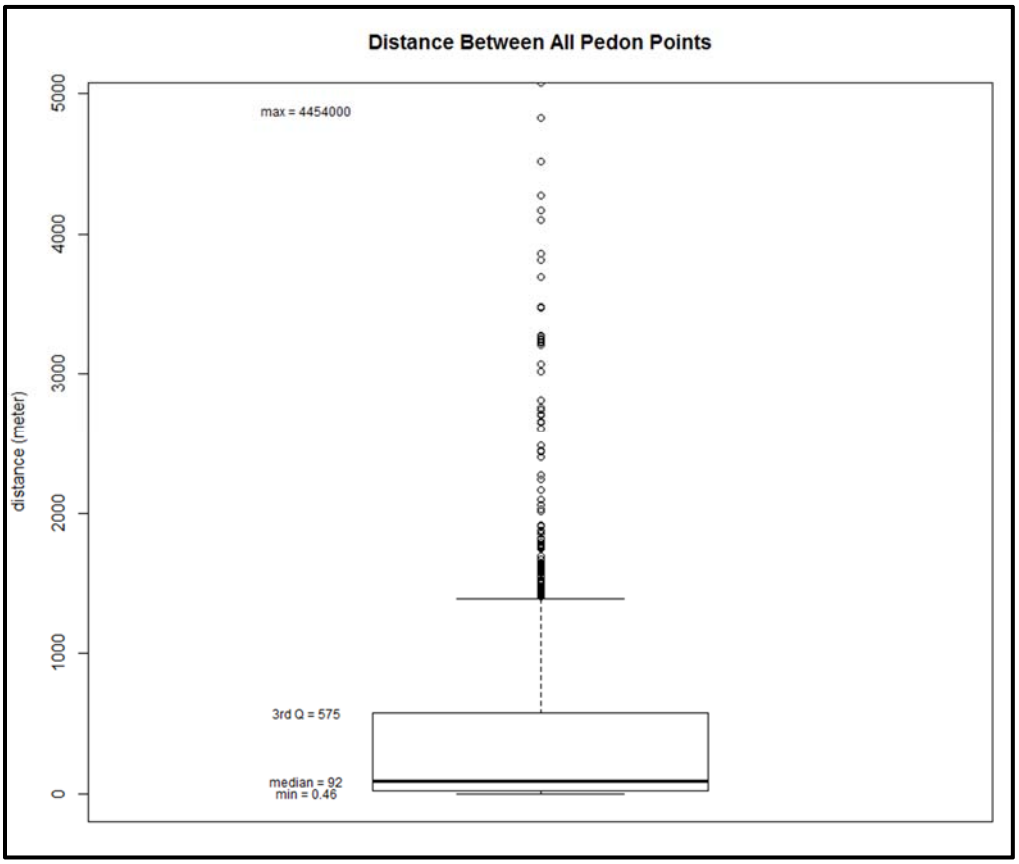


Figure 4.15. Boxplot of the distances between the NASIS pedon points and PLSS Model pedon points.

Tables 4.4 c. and 4.4 e. show that points sampled without GPS and those sampled with GPS before the WAAS corrections were implemented, have the smallest range in the distance between the NASIS-PLSS pedon pairs. Points sampled with a GPS that implemented the WAAS corrections, had the greatest standard deviation of all of the sets. The boxplot shows that most of the distances fell within a range of 0.46 meters to 575 meters or roughly 0.5 km. In addition to the summary statistics and the boxplot, a pie chart of the percentage of pedon points in each set of designated distance ranges, was created (Figure 4.16). It shows that 40 % of the pedon pairs were 10 – 100 meters away from each other.

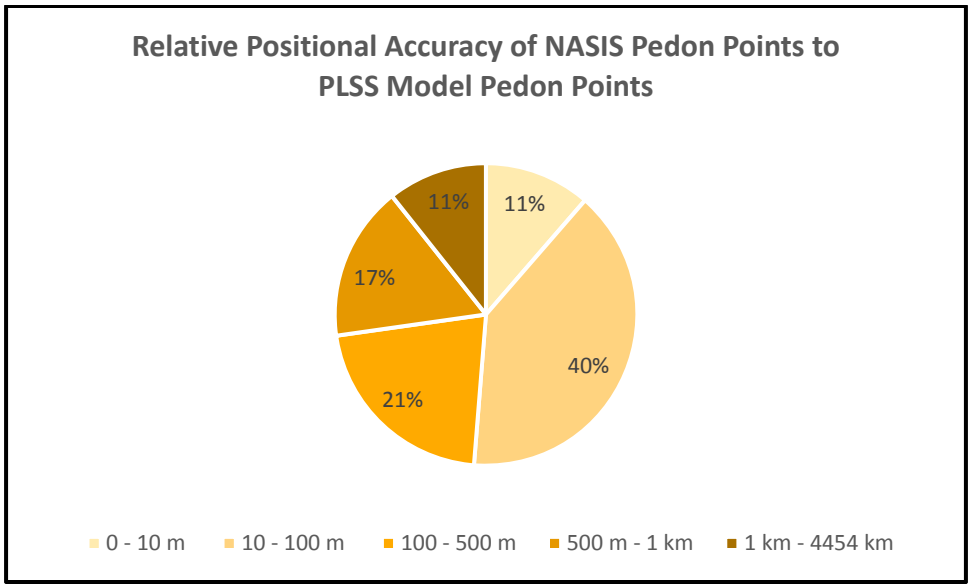


Figure 4.16. Percentage of pedon points in each of the five user-defined distance ranges.

The PLSS Model pedon points and NASIS pedon points were plotted together on a map (Figures 4.17 b,d,f,h). The NASIS pedon points and their associated distances to PLSS Model pedon points were also mapped in order to observe spatial patterns in the pedon data across the state. Distances were represented by a set of graduated circles (Figures 4.17 a,c,e,g,i). Clusters of points represent areas that were heavily surveyed. Large circles represent large differences in distances.

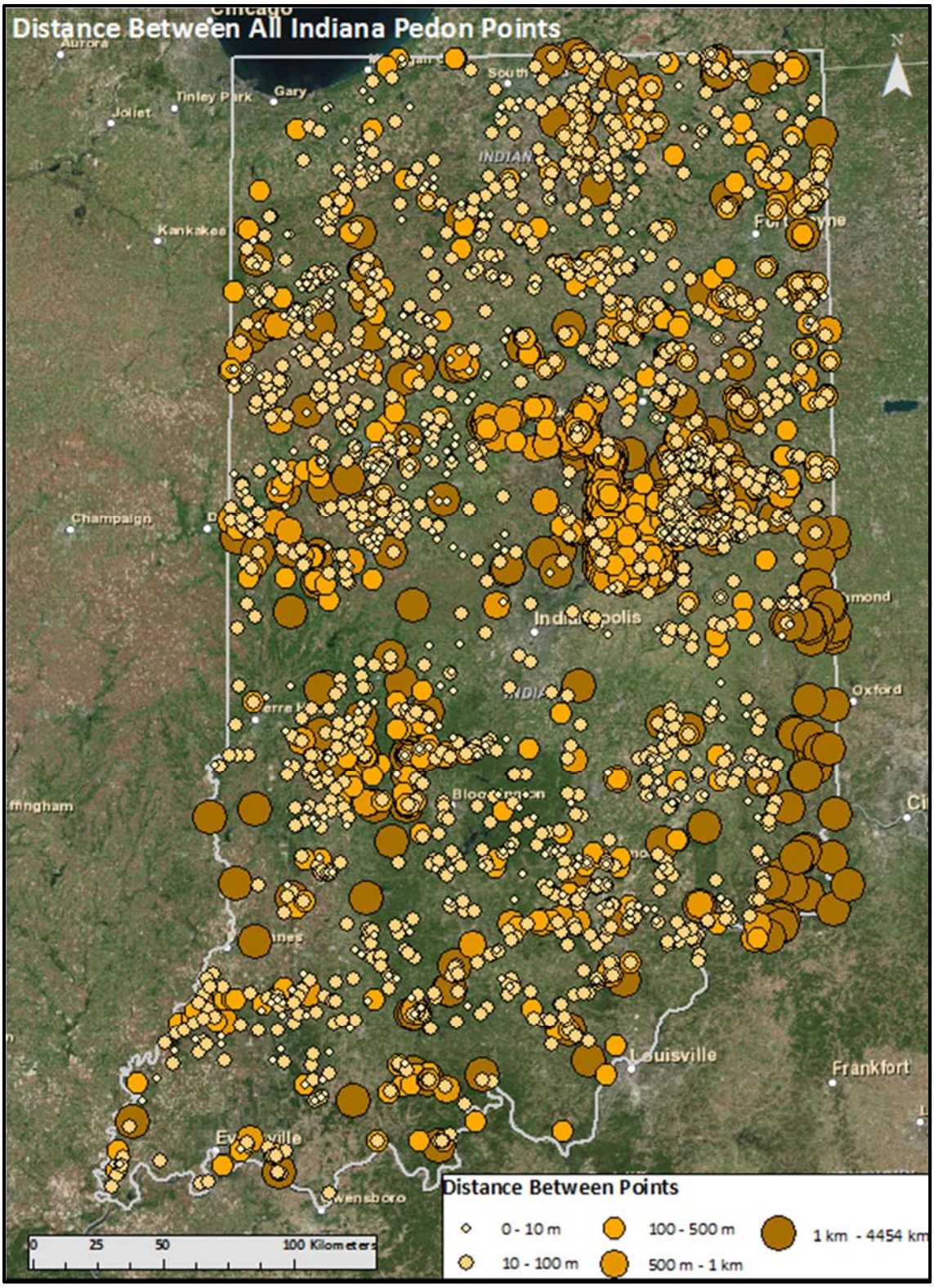


Figure 4.17 a. Distance between NASIS and PLSS Model pedon points for Indiana

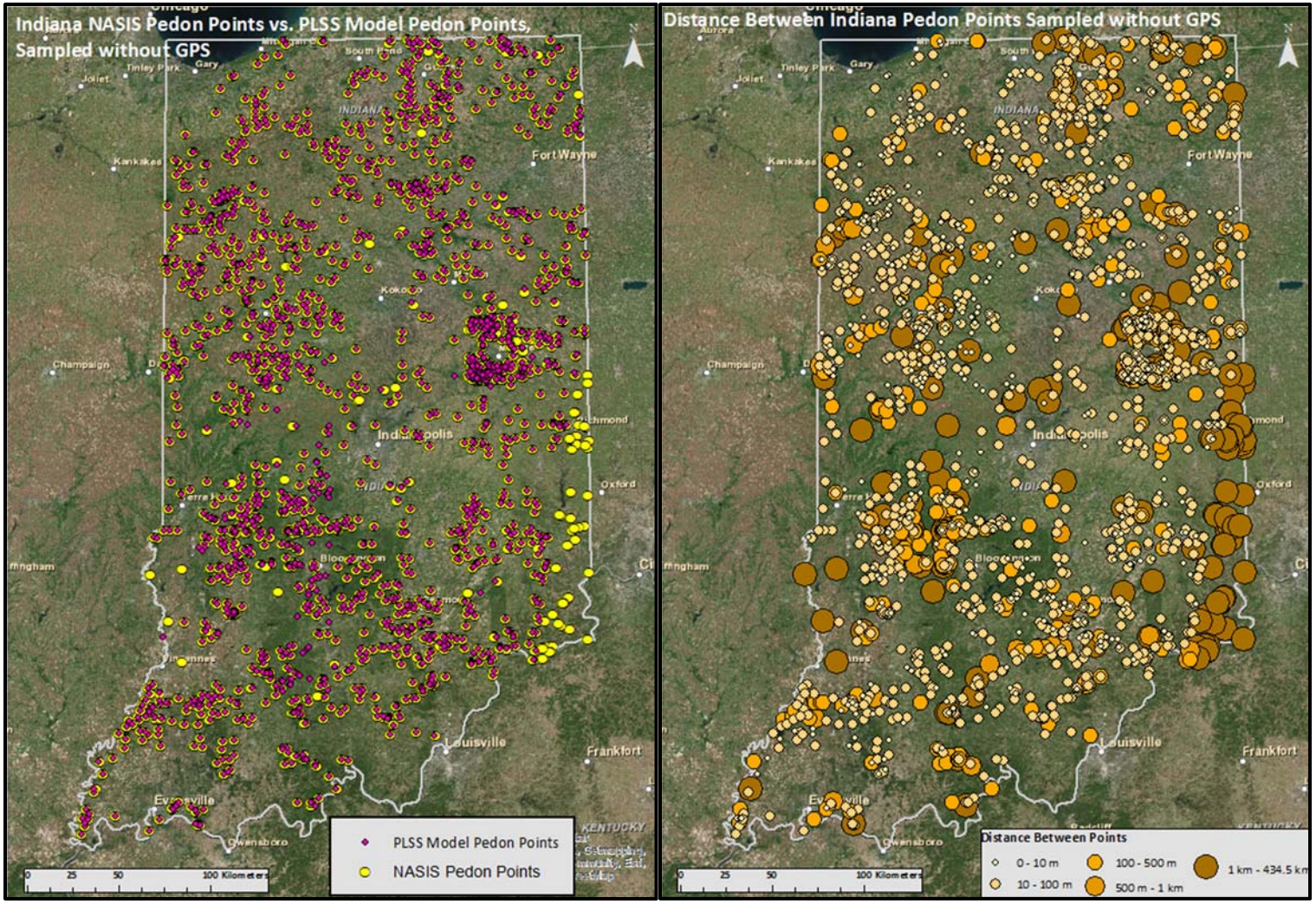


Figure 4.17 b. Indiana NASIS pedon points versus PLSS Model pedon points sampled without GPS (left) Figure 4.17 c Distance between Indiana pedon points sampled with GPS (right)

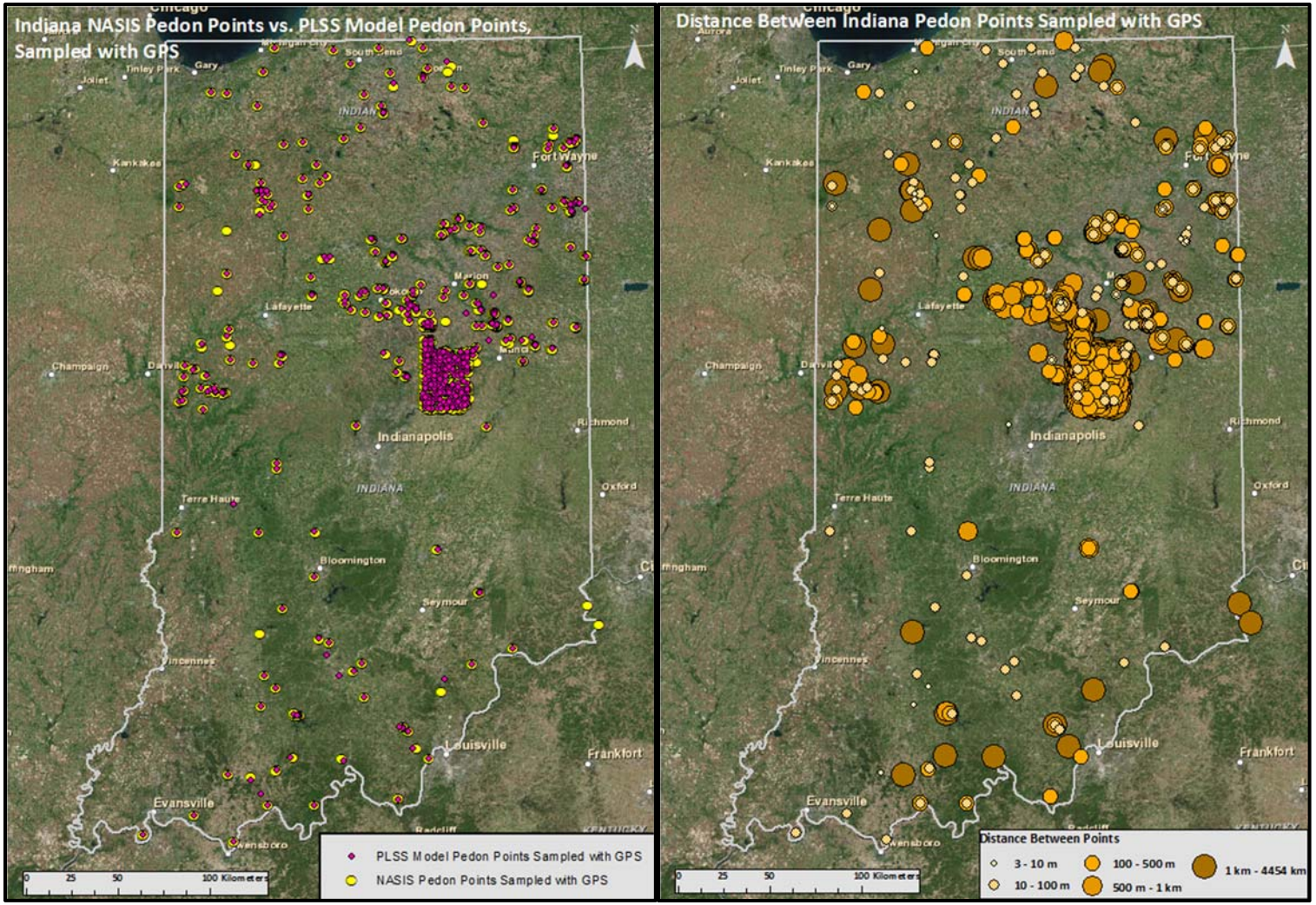


Figure 4.17 d. Indiana NASIS pedon points versus PLSS Model pedon points sampled with GPS (left) Figure 4.17 e. Distance between Indiana pedon points samples with GPS (right)

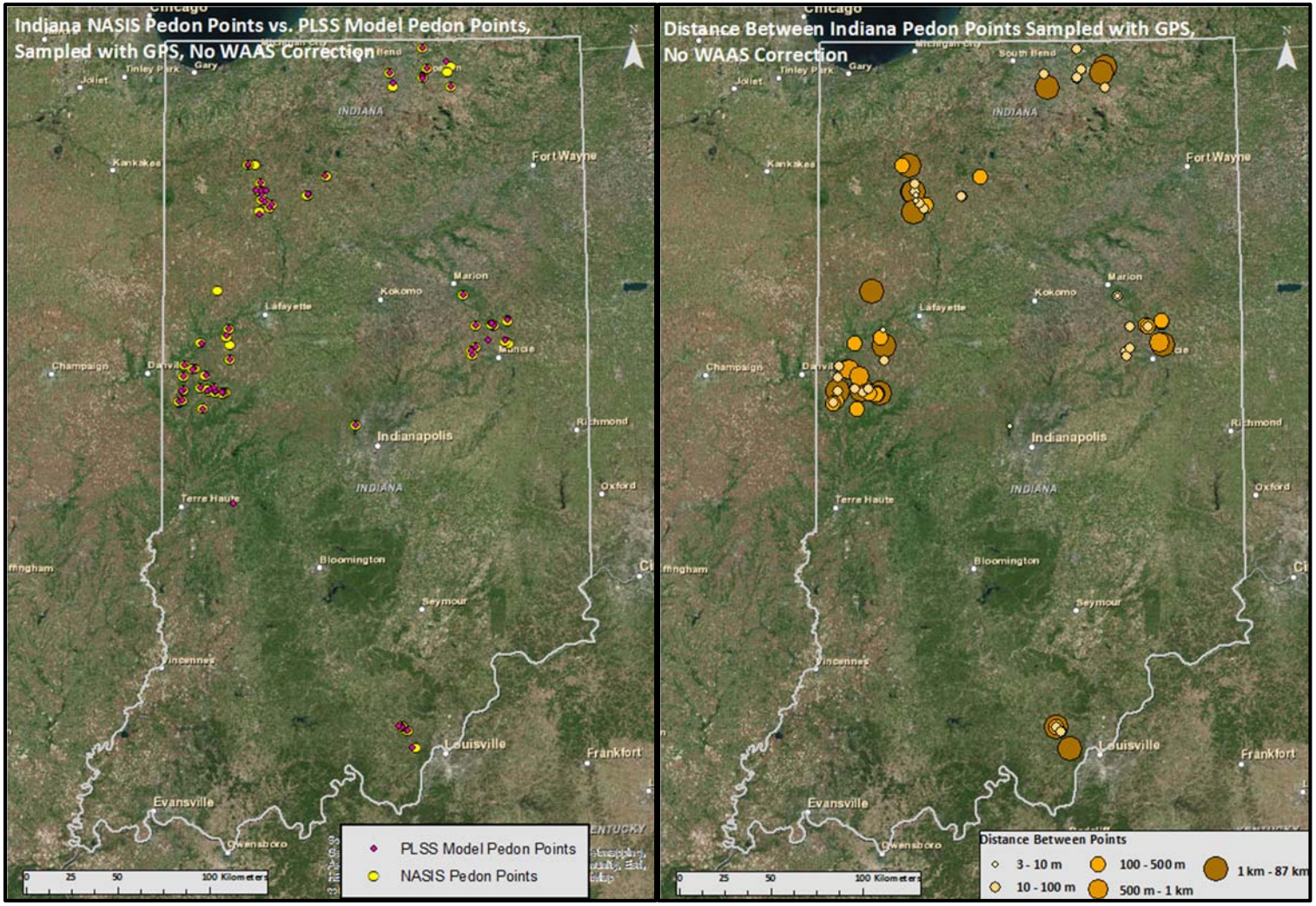


Figure 4.17 f. Indiana NASIS pedon points versus PLSS Model pedon points sampled with GPS, no WAAS correction (left) Figure 4.17 g. Distance between Indiana pedon points sampled with GPS, no WAAS correction (right)

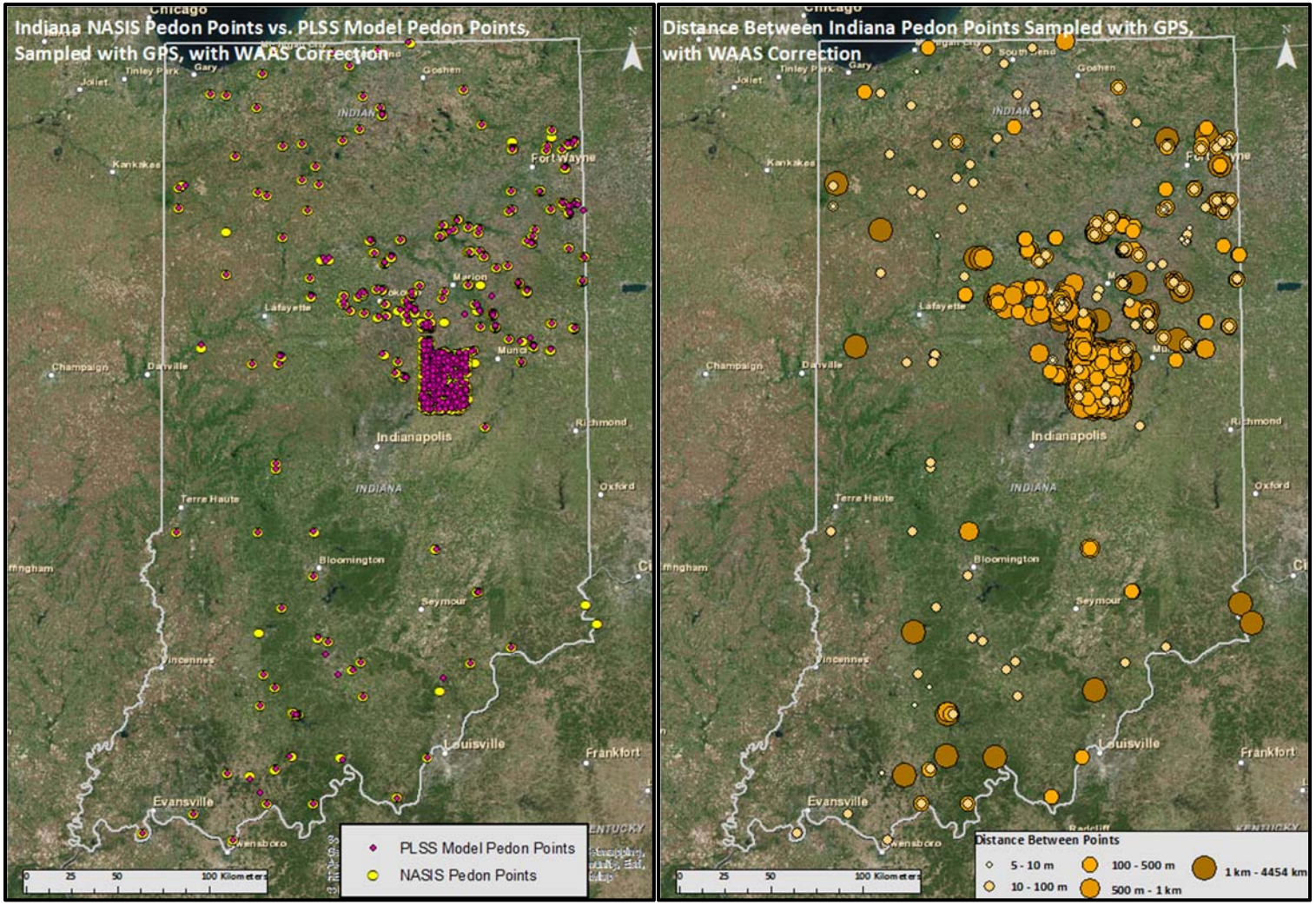


Figure 4.17 h. Indiana NASIS pedon points versus PLSS Model pedon points, sampled with GPS, with WAAS correction (left)

Figure 4.17 i. Distance between Indiana pedon points sampled with GPS, with WAAS correction (right)

4.3.2 The Soil Environment at the Pedon Point

Using distance to compare the NASIS pedon points to the PLSS Model pedon points provides a measure of relative accuracy of pedon data. However, in order to determine which set of points is more accurate or if any correction is required, it is necessary to evaluate the soil properties at those locations and compare them to the observations made by soil scientists when the pedon was described. Static details in the pedon's site description were used to define a "soil environment" for that pedon. The parent material, drainage class, and Soil Taxonomy were defined as the components of the soil environment. The gSSURGO was used as a reference base for the soil environment at the pedon location. The Soil Environment Match Model helped to identify pedon points whose soil environment was entirely represented by the gSSURGO data at that location. Results of the Soil Environment Match Model are presented for each subset of the PLSS Model and NASIS pedon points. The percent agreement between the pedons' soil environment and gSSURGO is displayed through a set of figures below (Figures 4.18 a,b,c,d,e,f).

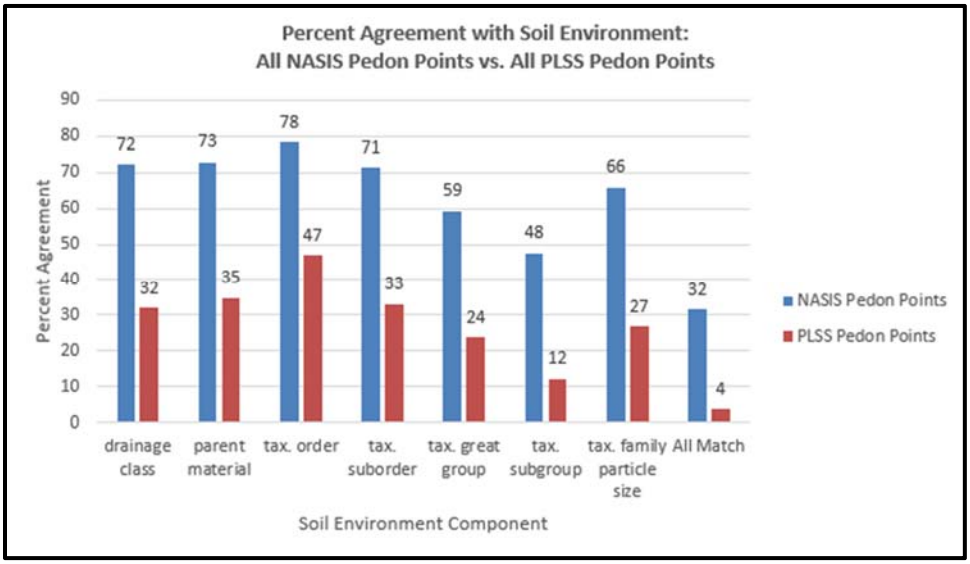


Figure 4.18 a. Percent agreement with the soil environment for all NASIS pedon points versus all PLSS pedon points

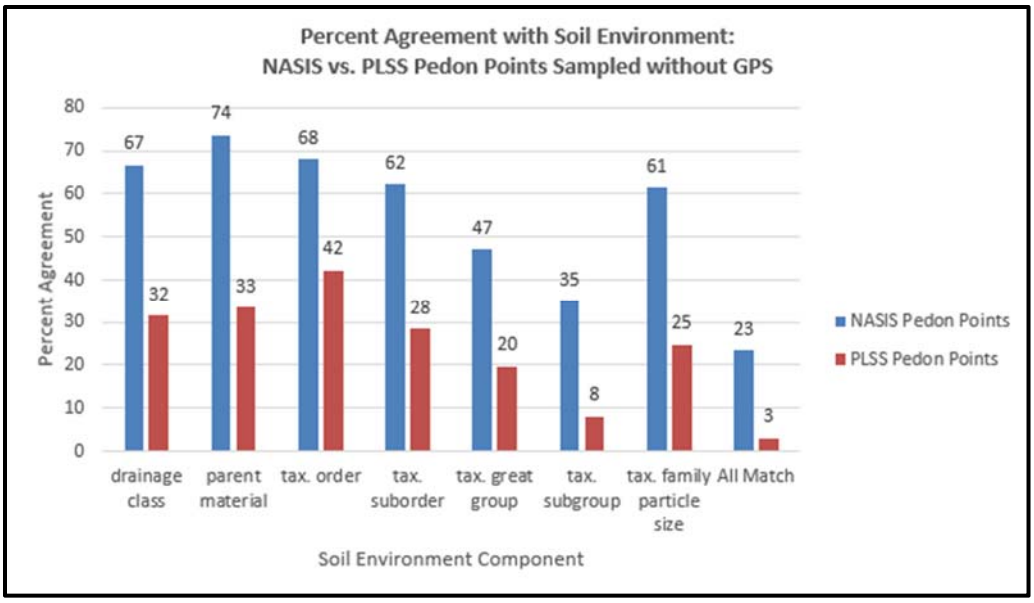


Figure 4.18 b. Percent agreement with the soil environment for the NASIS versus PLSS pedon points sampled without GPS

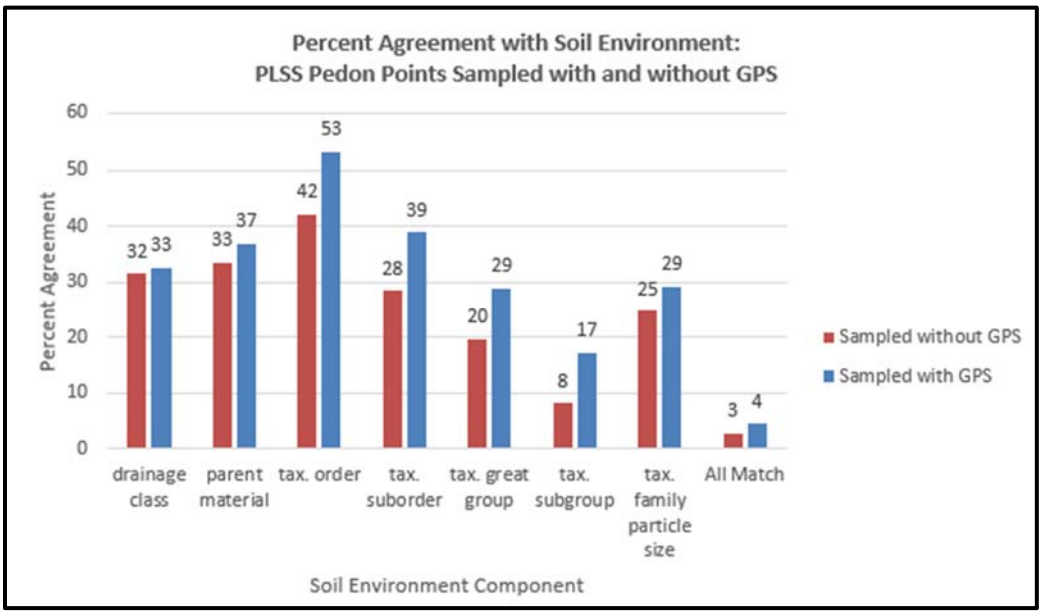


Figure 4.18 c. Percent agreement with the soil environment for the PLSS pedon points sampled with and without GPS

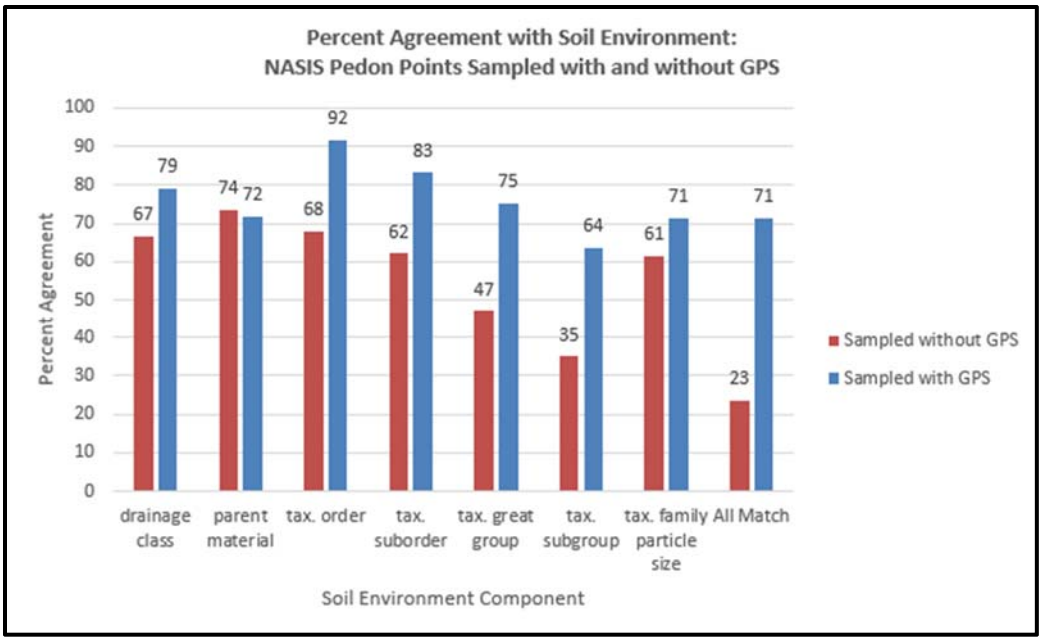


Figure 4.18 d. Percent agreement with the soil environment for the NASIS pedon points sampled with and without GPS

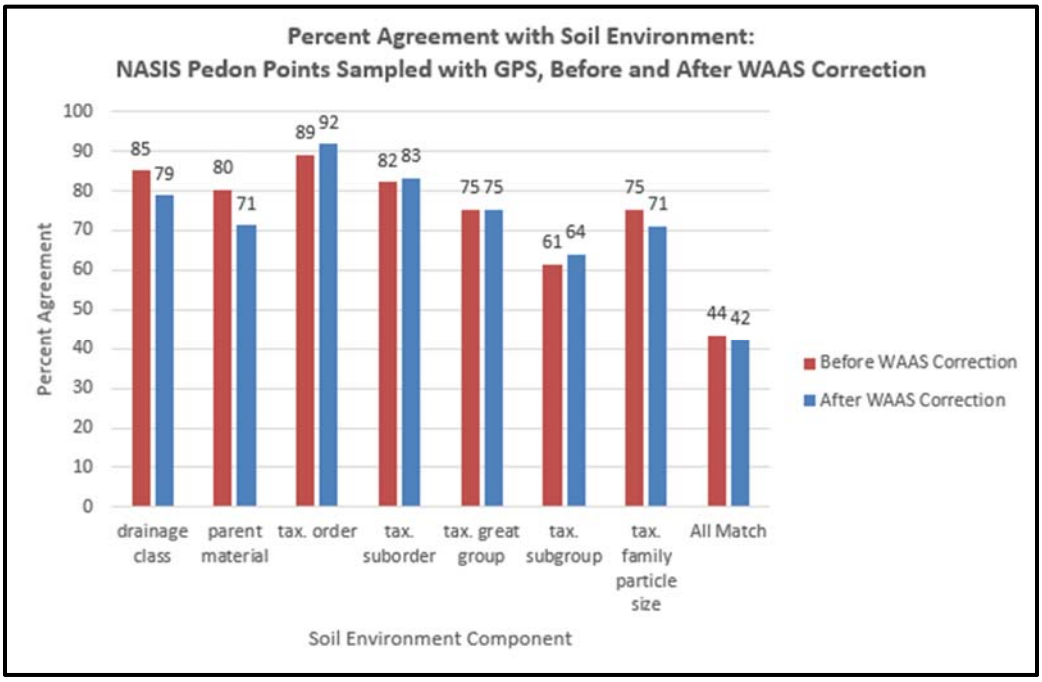


Figure 4.18 e. Percent agreement with the soil environment for the NASIS pedon points sampled with GPS, before and after WAAS correction

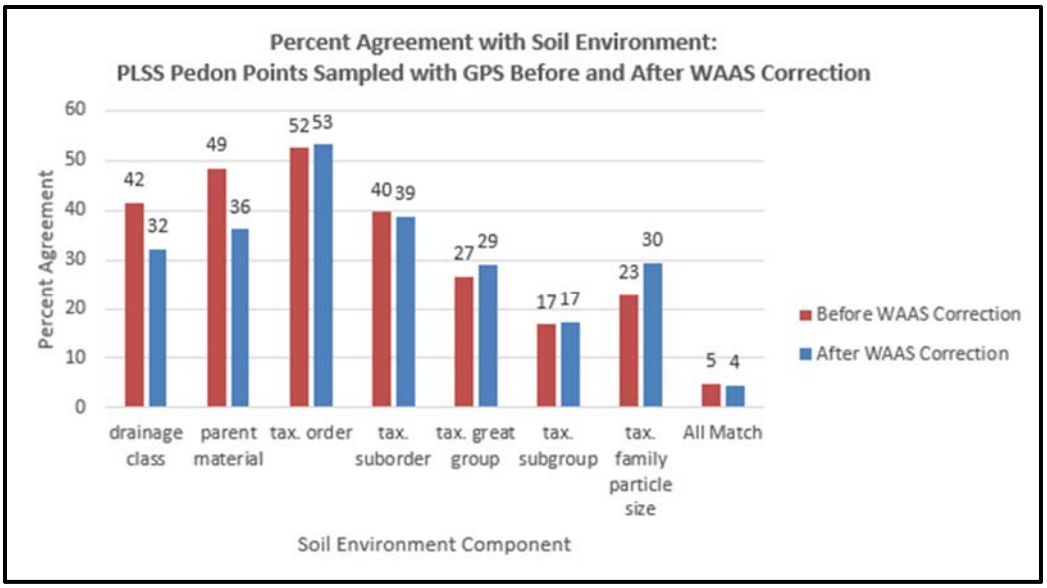


Figure 4.18 f. Percent agreement with soil environment: PLSS pedon points sampled with GPS before and after WAAS correction

Figure 4.18 a. shows that the NASIS pedon points had a significantly higher overall percent agreement with the soil environment than the PLSS Model pedon points. Figure 4.18 b. shows that even the NASIS pedon points collected without GPS had about twice as much greater percent agreement than the PLSS Model points. A comparison of PLSS pedon points sampled with and without GPS shows that those points sampled with a GPS have slightly higher percent agreement than those collected without a GPS. Figure 4.18 d. shows that NASIS pedon points collected with GPS have a much higher percent agreement than those collected without GPS. Figures 4.18 e. and 4.18 f. show that the WAAS correction did not improve the percent agreement of the pedon data for Indiana with the soil environment. Because the percentage agreement at the PLSS Model point locations was so low, it was assumed that the NASIS pedon point locations were a better estimate of the actual soil sampling location. Overall, 4% or just 61 of the 4141 total PLSS pedon points were found to have a complete match to the soil environment. The total number of pedon points from NASIS had over 23% of matches with the soil environment, that is, 2833 of the 4141 pedon points did not match.

For this reason, only the NASIS pedon points were used as the inputs for the next model, the Nautilus Match Model.

4.3.3 Positional Accuracy of Pedon Point Locations

The Nautilus Match Model was developed to find the nearest mukey with a full match to the soil environment and place a new pedon point at that location.

Table 4.5 shows a summary of the positional accuracy of the NASIS pedon points as determined by the different models.

Model Used	Number of Pedons	Positional Accuracy
Soil Environment Match Model	1308	Exact match to soil environment, 0 meters
Nautilus Match Model – No Match	2550	No match to soil environment, > 140 meters
Nautilus Match Model	283	Match to soil environment found, < 140 meters

Of the 4141 total soil pedon points from NASIS for Indiana, 1308 entirely matched the soil environment and were considered to be in the best location possible. The remaining 2833 pedon points that did not match were run through the Nautilus Match Model to find their closest matching program file. As a result of the Nautilus Match Model, 283 new matches were found and those pedon points were moved to that new location. A measure of positional accuracy can be assigned to each of these NASIS pedon points as the distance between their original location and their final location from the Nautilus Match Model. The Nautilus Match Model only searched a neighborhood of 21x21 cells around each pedon point, or a maximum distance of 140 meters from the NASIS pedon point. Points for which no full match to the soil environment was found through the Nautilus Model, are considered to have a positional

accuracy of more than 140 meters. Appendix C lists the NASIS pedon points used in this study and their positional accuracy.

4.4 Discussion

The three models developed for this study: PLSS Model, Soil Environment Match Model, and Nautilus Match Model, used separately or together, provide a way to assess the positional accuracy of the existing pedon point location in NASIS. The PLSS Model algorithm provides a more efficient method to parse the soil scientist's location descriptions and plot the pedon point on a map. If certain details are included in the location description, then the PLSS Model is a more accurate method of assigning a coordinate location to a pedon point than Reinsch's PLSS Conversion Model used by the NRCS. Additionally, the PLSS grid that was developed for the PLSS Model can be a useful tool for making management decisions that rely on detailed land boundaries and can be used for spatial analysis in a GIS software. The Soil Environment Match Model provides a consistent way of comparing tabular data from SSURGO and NASIS in order to determine whether or not the NASIS pedon point is in a location with similar site characteristics to that described by the soil scientist. The Soil Environment Match Model algorithm design allows for its application to any study that requires a comparison of tabular data extracted for an X, Y location. The Nautilus Model serves a dual purpose. It finds the closest, matching soil environment to the NASIS pedon point location and it prints the X, Y coordinates of that location along with the new mukey that is a match. It is also a way to derive a measure of positional accuracy to the pedon points. The positional accuracy is computed as the distance in meters between the original NASIS pedon location and the new pedon location from the Nautilus Model. Moreover, the Nautilus Model can be adopted for use in a similar study that requires a pixel-by-pixel search for data in a GIS software.

Although all three of the automated models developed provide an effective, consistent way of assessing the level of positional accuracy for the NASIS pedon points, they are limited by the deficiencies in the input pedon data. As the embedded table in Figure 4.13 shows, pedon sample sites were described to different levels of detail. The level of detail contained in the descriptions greatly affects the positional accuracy of the resulting PLSS Model point location. Positional accuracy of the pedon point should increase going from descriptions with a high level of abstraction, which reference only a township, to descriptions with a low level of abstractions, which cite one of the five points of a quarter-quarter-quarter section.

The results of the Soil Environment Match Model show that points sampled with a GPS without WAAS corrections were not necessarily more accurate than points sampled with GPS with the WAAS corrections. This is contrary to what was expected because WAAS corrections were meant to improve the accuracy of GPS readings. It is possible that the pedon data was not subset correctly by different georeference sources. Additional information is necessary to understand the discrepancy in these results.

The Nautilus Model is the best-available method known to find the closest matching cell with the soil environment described for the NASIS pedon. However, it is limited by the user's computer processing capability. In order to run 2833 pedon points through the model, the data had to be split into sets to run it through five systems and it took over 3 weeks to complete. Future work may include increasing the search radius of the Nautilus Match Model and running multiple iterations until the full set of pedon points is matched.

The U.S. Soil Survey pedon data is the cumulative effort of several generations of soil scientists in collecting, describing and analyzing soil pedons. Today, this data is in high demand as inputs into soil predictive models and models for precision agriculture. For these reasons, it is imperative that pedon data is scrutinized before it is introduced into these models to reduce model error and improve estimates.

REFERENCES

REFERENCES

Aksoy, E., Özsoy, G., & Dirim, M. S. (2009). Soil mapping approach in GIS using Landsat satellite imagery and DEM data, *4*(11), 1295–1302.

Ali, R. R., & Kotb, M. M. (2010). Use of Satellite Data and GIS for Soil Mapping and Capability Assessment. *Nature and Science*, *8*(8), 104–115.

Ashtekar, J. M., & Owens, P. R. (2013). Remembering Knowledge: An Expert Knowledge Based Approach to Digital Soil Mapping. *Soil Horizons*, *54*(5), 1–6. doi:10.2136/sh13-01-0007

Bock, M., & Kothe, R. (2008). Predicting the Depth of Hydromorphic Soil Characteristics Influenced by Ground Water. *Hamburger Beiträge Zur Physischen Geographie Und Landschaftsökologie*, (19), 13–22.

Boettinger, J. L., Ramsey, R. D., Bodily, J. M., Cole, N. J., Nield, S. J., Saunders, A. M., & Stum, A. K. (2008). Landsat Spectral Data for Digital Soil Mapping. In A. E. Hartemink, A. B. Mcbratney, & L. Mendonça-Santos (Eds.), *Digital Soil Mapping with Limited Data* (1st ed., pp. 193–202). London: Springer Science & Business Media.

Bureau of Land Management. (2002). *Surveying Our Public Lands* (pp. 1–21). Retrieved from http://www.blm.gov/pgdata/etc/medialib/blm/wo/MINERALS__REALTY__AND_RESOURCE_PROTECTION_/W0350/cadastral_pdfs.Par.11507.File.dat/SurveyingPL2002.pdf

Butler, K. (2013). Band Combinations for Landsat 8. Retrieved June 26, 2014, from http://blogs.esri.com/esri/arcgis/2013/07/24/band-combinations-for-landsat-8/
Cook, S. E., Jarvis, A., & Gonzalez, J. P. (2008). A New Global Demand for Digital Soil Information. In A. E. Hartemink, A. McBratney, & M. de L. Mendonca-Santos (Eds.), *Digital Soil Mapping with Limited Data* (1st ed., pp. 31–41). Springer Science & Business Media.

Crum, J.R., Steinhardt, G.C., Franzmeier, D. P. (1977). Soil Characterization in Indiana: 1 - Field and Laboratory Procedures. In *Purdue University Agricultural Experiment Station Research Bulletin, 943.* (Vol. 943, p. 35). West Lafayette, IN: Government Publication. doi:632049040

De Alba, S., Lindstrom, M., Schumacher, T. E., & Malo, D. D. (2004). Soil landscape evolution due to soil redistribution by tillage: a new conceptual model of soil catena evolution in agricultural landscapes. *Catena*, *58*(1), 77–100. doi:10.1016/j.catena.2003.12.004

Demattê, J. A. M., Huete, A. R., Guimarães, L., Jr, F., Nanni, M. R., Alves, M. C., & Fiorio, P. R. (2009). Methodology for Bare Soil Detection and Discrimination by Landsat TM Image. *The Open Remote Sensing Journal*, *2*, 24–35.

Dewitte, O., Jones, A., Elbelrhiti, H., Horion, S., & Montanarella, L. (2012). Satellite remote sensing for soil mapping in Africa: An overview. *Progress in Physical Geography*, *36*(4), 514–538. doi:10.1177/0309133312446981

Dobos, E., Bliss, N., Worstell, B., Montanarella, L., Johannsen, C., & Micheli, E. (2002). Paper no. 649. The use of DEM and satellite data for regional scale soil databases. In *17th World Congress of Soil Science, 14-21 August 2002, Thailand* (pp. 1–12).

Dobos, E., Montanarella, L., Negre, T., & Micheli, E. (2001). A regional scale soil mapping approach using integrated AVHRR and DEM data. *International Journal of Applied Earth Observation and Geoinformation*, *3*(1), 30–42.

Du-blayo, L. Le, Gou, P., Corpetti, T., Michel, K., Lemercier, B., & Walter, C. (2008). Enhancing the Use of Remotely-Sensed Data and Information for Digital Soilscape Mapping. In A. E. Hartemink, A. B. Mcbratney, & L. Mendonça-Santos (Eds.), *Digital Soil Mapping with Limited Data* (1st ed., pp. 337–348). London: Springer Science & Business Media.

Estes, J. E., & Mooneyhan, D. W. (1994). Of Maps and Myths. *Photogrammetric Engineering and Remote Sensing*, *60*(5), 517–524.

Evans, I. S. (2013). Land surface derivatives : history , calculation and further development . *Geomorphometry.org*, 1–4. Retrieved from http://geomorphometry.org/system/files/Evans2013geomorphometry.pdf

Exelis Visual Information Solutions. (2009). *ENVI EX User's Guide*. Boulder, Colorado: Exelis Visual Information Solutions.

Exelis Visual Information Solutions. (2014). FLAASH Background FLAASH ® Background. Retrieved from http://www.exelisvis.com/docs/BackgroundFLAASH.html

Exelis Visual Information Solutions Inc. (2014). *Calibrate Landsat 8 Imagery For Image Processing | ENVI 5 . 1*. United States: youtube.com. Retrieved from https://www.youtube.com/watch?v=M6IqEysywtw

Fleming, A. (2013). Ice Age in Indiana. Retrieved from http://igs.indiana.edu/Surficial/IceAge.cfm

Food and Agriculture Organization. (2014). The multiple dimensions of food security. Retrieved August 10, 2013, from http://www.fao.org/publications/sofi/2013/en/

Franzmeier, D. P., Steinhardt, G. C., & Schulze, D. G. (2004). *Indiana Soil and Landscape Evaluation Manual* (Vol. 1, p. 74). Retrieved from https://www.extension.purdue.edu/extmedia/AY/AY-323.pdf

Gallant, J. C., & Dowling, T. I. (2003). A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research*, *39*(12), n/a–n/a. doi:10.1029/2002WR001426

Gardner, D. R. (1998). *The National Cooperative Soil Survey of the United States* (p. 270). Washington D.C.

Grunwald, S. (2014, May). Soil-Information Systems in the U.S. Retrieved from http://soils.ifas.ufl.edu/faculty/grunwald/teaching/eSoilScience/soildata.shtml

Gu, Y., Hunt, E., Wardlow, B., Basara, J. B., Brown, J. F., & Verdin, J. P. (2008). Evaluation of MODIS NDVI and NDWI for vegetation drought monitoring using Oklahoma Mesonet soil moisture data. *Geophysical Research Letters*, *35*(22), 1–5. doi:10.1029/2008GL035772

Hudson, B. D. (1992). The Soil Survey as Paradigm-based Science. *Soil Science Society of America Journal*, *56*, 836–841.

Huete, A. R. (1988). A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment*, *25*, 295–309.

Indiana Geological Survey. (1998). LANDSURVEY _ SECTIONS _ POLY _ IN : Land Survey Lines of Indiana (Indiana Geological Survey, 1:24,000, Polygon Shapefile) Metadata. Retrieved from http://maps.indiana.edu/metadata/Reference/PLSS_Sections.html

Jasiewicz, J., & Stepinkski, T. (2013). GRASS GIS manual: r.geomorphon. Retrieved from http://grass.osgeo.org/grass70/manuals/addons/r.geomorphon.html

Jasiewicz, J., & Stepinski, T. F. (2013). Geomorphons — a pattern recognition approach to classification and mapping of landforms. *Geomorphology*, *182*(null), 147–156. doi:10.1016/j.geomorph.2012.11.005

Jenny, H., & Amundson, R. (1994). *FACTORS OF SOIL FORMATION A System of Quantitative Pedology* (p. 191). Mineola, NY: Dover Publications, Inc. Retrieved from http://www.soilandhealth.org/01aglibrary/010159.Jenny.pdf

Lozano-Garcia, D. F., Fernandez, R. N., Gallo, K. P., & Johannsen, C. J. (1995). Monitoring the 1988 severe drought in Indiana, U.S.A. using AVHRR data. *International Journal of Remote Sensing*, *16*(7), 1327–1340. doi:10.1080/01431169508954479

Lytle, D. J. (1999). United States Soil Survey Database. In M. E. Sumner (Ed.), *Handbook of Soil Sciences* (pp. H53–67). Boca Raton: CRC Press.

MacMillan, R. a, Martin, T. C., Earle, T. J., & McNabb, D. H. (2014). Automated analysis and classification of landforms using high-resolution digital elevation data: applications and issues. *Canadian Journal of Remote Sensing*, *29*(5), 592–606. doi:10.5589/m03-031

Maselli, F. (2004). Monitoring forest conditions in a protected Mediterranean coastal area by the analysis of multiyear NDVI data. *Remote Sensing of Environment*, *89*(4), 423–433. doi:10.1016/j.rse.2003.10.020

McBratney, a. ., Mendonça Santos, M. ., & Minasny, B. (2003). *On digital soil mapping*. *Geoderma* (Vol. 117, pp. 3–52). doi:10.1016/S0016-7061(03)00223-4

McSweeney, K., Gessler, P. E., Slater, B. K., Hammer, D. R., Petersen, G. W., & Bell, J. C. (1994). Towards a New Framework for Modeling the Soil-Landscape Continuum. In R. R. Amundson, J. Harden, & M. Singer (Eds.), *Factors of Soil Formation: A Fiftieth Anniversary Retrospective* (SSSA Speci., pp. 127–145). Soil Science Society of America. doi:10.2136/sssaspecpub33.c8

Milne, G. (1936). *A PROVISIONAL SOIL MAP OF EAST AFRICA* (p. 33). Amani, Tanganyika Territory.

Nachtergaele, F. O. (1999). From the Soil Map of the World to Digital Global Soils and Terrain Database: 1960-2002. In M. E. Sumner (Ed.), *Handbook of Soil Sciences* (pp. H5–17). Boca Raton: CRC Press.

National Oceanic & Atmospheric Administration. (2004). *Climatography of the United States No. 20 Lafayette 8S, IN 1971-2000* (Vol. 2001, p. 031–A – 031–E).

Peel, M. C., Finlayson, B. L., & Mcmahon, T. A. (2007). Updated world map of the Koppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*, 1633–1644.

Peterson, J. B. (1990). The prairie point. Retrieved from https://ag.purdue.edu/agry/acre/Documents/Prairie-point.pdf

Ray, T. (2014). A FAQ on Vegetation in Remote Sensing. Retrieved June 26, 1014, from http://www.yale.edu/ceo/Documentation/rsvegfaq.html

Reinsch, T. G. (National S. S. C. (1994). PLSS Conversion Access Database. doi:http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_051345.pdf

Rossel, R. A. V., McBratney, A. B., & Minansy, B. (Eds.). (2010). *Proximal Soil Sensing* (p. 446). London: Springer Science & Business Media.

Ruhe, R. V. (1960). Elements of the soil landscape. In *Transactions of the 7th International Congress of Soil Science* (pp. 165–170).

Sanchez, P. A., Ahamed, S., Carré, F., Hartemink, A. E., Hempel, J., Huising, J., … Zhang, G. (2009). Digital Soil Map of the World. *Science Magazine*, *325*(5941), 680–681. doi:10.1126/science.1175084

Senay, G. B., & Elliott, R. L. (2000). Combining AVHRR-NDVI and landuse data to describe temporal and spatial dynamics of vegetation. *Forest Ecology and Management*, *128*(1-2), 83–91. doi:10.1016/S0378-1127(99)00275-3

Singh, D., Meirelles, M. S. P., Costa, G. a., Herlin, I., Berroir, J. P., & Silva, E. F. (2006). Environmental degradation analysis using NOAA/AVHRR data. *Advances in Space Research*, *37*(4), 720–727. doi:10.1016/j.asr.2004.12.052

Snyder, J. P. (U. S. G. S. (1987). *Map Projections--A Working Manual, U.S. Geological Survey Professional Paper 1395* (p. 383). Washington D.C.: United States Government Printing Office. Retrieved from http://pubs.usgs.gov/pp/1395/report.pdf

Soil Survey Division Staff. (1993). *Soil Survey Manual. USDA- NRCS Agric. Handb. 18.* Washington, D.C.: US Government Printing Office.

Soil Survey Staff. (2011). *Soil Survey Laboratory Information Manual. Report No. 45, Ver. 2.0* (Vol. 2.0). Washington, D.C. Retrieved from http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_052226.pdf

Soil Survey Staff. (2014). NASIS 7.0.4 Table Column Descriptions. Retrieved from file:///C:/Users/mdorante/Downloads/NASIS_Table_Column_Descriptions (2).pdf

Steinhardt, G., Owens, P., & Schulze, D. (2013). *Indiana Land Surveys : Their Development and Uses*. *Purdue Extension* (Vol. AY-237-W, pp. 1–10).

Tsvetsinskaya, E. A., Schaaf, C. B., Gao, F., Strahler, A. H., Dickinson, R. E., Zeng, X., & Lucht, W. (2002). Relating MODIS-derived surface albedo to soils and rock types over Northern Africa and the Arabian peninsula. *Geophysical Research Letters*, *29*(9), 3–6.

Tucker, C. J. (1979). Red and Photographic Infrared l , lnear Combinations for Monitoring Vegetation. *Remote Sensing of Environment*, *8*, 127–150.

U.S. Geological Survey. (2014a). Frequently Asked Questions about the Landsat Missions. Retrieved June 26, 2014, from http://landsat.usgs.gov/band_designations_landsat_satellites.php

U.S. Geological Survey. (2014b). Landsat 8. Retrieved June 25, 2014, from http://landsat.usgs.gov

United States Department of Agriculture - NRCS. (2008). *Analysis PC 1.0 User's Guide*. Retrieved from http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_053346.pdf

USGS & USDA-FS. (2003). Part 5 Public Land Survey System Standards for USGS and USDA Forest Service Single Edition Quadrangle Maps. In *National Mapping Program Technical Instructions* (pp. i–13). Retrieved from http://nationalmap.gov/standards/pdf/5seqm503.pdf

Wang, X., Xie, H., Guan, H., & Zhou, X. (2007). Different responses of MODIS-derived NDVI to root-zone soil moisture in semi-arid and humid regions. *Journal of Hydrology*, *340*(1-2), 12–24. doi:10.1016/j.jhydrol.2007.03.022

Whelan, B. M., McBratney, A. B., & Minansy, B. (2002). VESPER 1.5 – SPATIAL PREDICTION SOFTWARE FOR PRECISION AGRICULTURE. In P. C. Robert, R. H. Rust, & W. E. Larson (Eds.), *Precision Agriculture, Proceedings of the 6th International Conference on Precision Agriculture* (p. 14p). Madison, Wisconsin. Retrieved from http://sydney.edu.au/agriculture/pal/documents/vesperpaper.pdf

White, C. A. (1926). *A History of the Rectangular Survey System* (p. 776). United States Department of Interior, Bureau of Land Management. Retrieved from http://www.blm.gov/cadastral/Manual/pdffiles/histrect.pdf

Wysocki, D., Schoeneberger, P., Hirmas, D., & LaGarry, H. (2011). Geomorphology of Soil Landscapes. In *Handbook of Soil Sciences* (pp. 1–26). CRC Press. doi:doi:10.1201/b11267-35

Yale Center for Earth Observation. (2013). Yale Guide to Landsat 8 Image Processing. Retrieved June 15, 2014, from http://www.yale.edu/ceo/Documentation/Landsat 8 image processing.pdf

Yokoyama, R., Shlrasawa, M., & Pike, R. J. (2002). Visualizing Topography by Openness : A New Application of Image Processing to Digital Elevation Models, *68*(3), 257–265.

Zeng, X., Dickinson, R.E., Walker, A., Shaikh, M., De Fries, R.S. and Qi, J. 2000. Derivation and evaluation of global 1-km fractional vegetation cover data for land modeling. Journal of Applied Meteorology 39(6): 826–839.

APPENDICES

Appendix A Soil Pedon Database Entry Form from

Purdue Agricultural Experiment Station – Soil Profile Description

SOIL PROFILE DESCRIPTION

SERIES: CRIDER SLOPE: 7%
EROSION: MODERATE DRAINAGE: WELL
LANDFORM: SIDESLOPE, SHOULDER VEGETATION: HAY
PARENT MATERIAL(S) :LOESS, WIS. :LIMESTONE
LOCATION: 920 FT E AND 250 FT N OF SW CORNER SEC.17 T.8N R.1W
SOIL NO.: S72IN53-1-(1- 7) COUNTY: MONROE FILE NO.: M07201
DATE DESCRIBED: 01/17/73 DESCRIBED BY: FROEHLE, WINGARD

DEPTH INCHES	HORIZON	BN- DRY	--DOMINANT-- MOIST	COLOR	-----MOTTLES----- MOIST	COLOR	A	S	C	TEXTURE MD	CLASS	---STRUCTURE--- -PRIM- G	SI	SH	SECOND- G	SI	SH
0- 9	AP	AS	10YR4/3								SIL	2	M	GR			
9- 12	B1T	CW	7.5YR5/4							LT	SICL	2	M	SBK			
12- 23	B21T	CW	7.5YR5/6								SICL	2	M	SBK			
23- 28	B22T	CW	5YR5/6								SICL	2	M	SBK			
28- 33	B23T	CW	5YR5/6								SICL	2	M	SBK			
33- 67 2	B24T		2.5YR4/6								C	3	M	ABK			
67- 0	R																

DEPTH	-----OTHER FEATURES----- KND	LOC	A	COLOR	KND	LOC	A	COLOR	CONSISTENCE DRY	MOI	WET	REACTION PH	EFF	COARSE FRAG LL	VX
0- 9										FR		6.3			
9- 12	CL	PE	C	7.5YR4/4						FR		5.8			
12- 23	CL	PE	C	5YR4/4						FR		4.8			
23- 28	CL	PE	C	5YR4/4	SI	PP	C	7.5YR6/4		FR		4.8			
28- 33	CL	PE	C	5YR4/4	SI	PP	M	10YR6/4		FR		4.8			
33- 67	CL	PE	M		CO		M			FI		4.8			
67- 0															

NOTES:

Extracted from Soil Characterization in Indiana: II. 1967-1973 Data. Station Bulletin No. 174. December 1977.

Appedix A continued → Soil Pedon Database Entry Form from

Purdue Agricultural Experiment Station – Sample Analysis Data

SERIES: CROSBY SOIL NO.: S69IN6-1-(1- 8)
COUNTY: BOONE FILE NO.: B06901

DEPTH INCHES	HORIZON	PARTICLE SIZE DISTRIBUTION (% <2MM) VCS	CS	MS	FS	VFS	SAND	SILT	FI SILT	CLAY	TEXT CLASS
0- 4	A1	1.8	4.0	4.6	9.3	7.0	26.7	56.3		17.0	SIL
4- 9	A2	3.0	4.5	4.7	8.6	6.6	27.4	56.1		16.5	SIL
9- 12	B1	1.3	2.7	3.8	8.4	6.6	22.8	49.4		27.8	CL
12- 19	2 B21T	0.8	2.5	4.1	8.9	7.3	23.6	40.2		36.2	CL
19- 28	2 B22T	0.9	2.3	4.0	9.4	8.7	25.3	41.2		33.5	CL
28- 36	2 B3	1.5	2.6	4.3	10.4	8.6	27.4	42.7		29.9	CL
36- 40	2 C1	3.5	4.3	5.5	12.4	10.9	36.6	44.5		18.9	L
40- 48	2 C2	5.2	4.8	5.6	12.0	11.1	38.7	45.2		16.1	L

DEPTH INCHES	PH 1:1	PH CA	PH K	LB/AC SMP	LB/AC P	LB/AC K	EXTRACTABLE (MEQ/100GM) CA	MG	NA	K	H	CEC SUM	%BASE SAT
0- 4	6.6				4		10.8	1.6	0.0	0.4	6.9	19.7	65.0
4- 9	6.7				10		7.1	1.2	0.1	0.2	4.8	13.4	64.2
9- 12	5.8			6.7	2		10.4	2.7	0.1	0.2	7.6	21.0	63.8
12- 19	5.3			6.3	1		10.6	5.2	0.1	0.2	10.1	26.2	61.5
19- 28	6.7				0		15.2	6.8	0.1	0.2	4.5	26.8	83.2
28- 36	7.5				0								
36- 40	8.2				2								
40- 48	8.4				0								

DEPTH INCHES	% ORG C	%CACO3 EQ	CO FRAG UL	CO FRAG WT%	Db-CORES-CLODS	
0- 4	3.17				1.20	
4- 9	0.51				1.48	
9- 12	0.53			4	1.54	
12- 19				3	1.48	
19- 28				3	1.48	1.64
28- 36		9.2		3	1.58	
36- 40		27.9		10	1.78	
40- 48		34.4		10		1.91

TAXADJUNCT, LOW CLAY CONTENT IN B HORIZONS

Extracted from Soil Characterization in Indiana: II. 1967-1973 Data. Station Bulletin No. 174. December 1977.

Appendix B Abbreviations Used in Soil Descriptions

Table 4. Abbreviations Used in Soil Descriptions (Terms defined in the Soil Survey Manual).

Slope
- A - 0-2%
- B - 2-6%
- C - 6-12%
- D - 12-18%
- E - 18-25%
- F - 25-35%
- G - >35%

Boundary (BNDRY)
- Distinctness
  - A - Abrupt
  - C - Clear
  - G - Gradual
  - D - Diffuse
- Topography
  - S - Smooth
  - W - Wavy
  - I - Irregular
  - B - Broken

Color
- Munsell notation

Mottles
- Abundance (A)
  - F - Few
  - C - Common
  - M - Many
  - V - Very many
- Size (S)
  - 1 - Fine
  - 2 - Medium
  - 3 - Coarse
- Contrast
  - F - Faint
  - D - Distinct
  - P - Prominent

Texture
- Modifier (MD)
  - CH - Channery
  - GR - Gravelly
  - GT - Gritty
  - HV - Heavy (high clay for texture class)
  - LT - Light (low clay for texture class)
  - MU - Mucky
  - SH - Shaly
  - ST - Stony
- Class (Components of abbreviation)
  - S - Sand(y)
  - SI - Silt(y)
  - C - Clay(ey)
  - L - Loam(y)
  - G - Gravel(ly)
  - V - Very
  - F - Fine
  - CO - Coarse
  - Eg., VCOS - Very coarse sand

Structure
- Grade (G)
  - 0 - Structureless
  - 1 - Weak
  - 2 - Moderate
  - 3 - Strong
- Size (SI)
  - VF - Very fine
  - F - Fine
  - M - Medium
  - C - Coarse
  - VC - Very Coarse
- Shape (SH)
  - ABK - Angular blocky
  - BL - Blocky
  - CPR - Prismatic columnar
  - CR - Crumb
  - GR - Granular
  - M - Massive
  - PL - Platey
  - PR - Prismatic
  - SBK - Subangular blocky
  - SG - Single grain

Other features
- Kind (KND)
  - Surface
    - CA - Carbonate coat
    - CL - Clay skin
    - MN - Manganese coat
    - OM - Organic matter coat
    - SA - Sand coat
    - SI - Silt coat
    - SL - Slickensides
    - ST - Stress surface
  - Interior
    - CO - Concretions
    - CR - Crystals
    - DC - Dry color
    - MO - Mottles
    - NO - Nodules
    - PN - Plinthite
    - PS - Pseudo-rock frag.
    - SK - Streaks
    - SO - Soft accumulation
- Location (LOC)
  - CH - Channels
  - GR - Grains
  - IN - Internal planes
  - PE - Peds
  - PO - Pores
  - PP - Peds and pores
  - RO - Rock fragments
- Amount (A)
  - F - Few
  - C - Common
  - M - Many
  - V - Very many
- Color
  - Munsell notation

Consistence
- Dry
  - L - Loose
  - S - Soft
  - SH - Slightly hard
  - H - Hard
  - VH - Very hard
  - EH - Extremely hard
- Moist
  - L - Loose
  - VFR - Very friable
  - FR - Friable
  - FI - Firm
  - VFI - Very firm
  - EFI - Extremely firm
- Wet
  - SO - Nonsticky
  - SS - Slightly sticky
  - S - Sticky
  - VS - Very sticky
  - PO - Nonplastic
  - PS - Slightly plastic
  - P - Plastic
  - VP - Very plastic

Reaction
- PH - Actual value, (Eg., 5.6)
- Effervescence (EFF)
  - E - Slight
  - ES - Strong
  - EV - Violent

Coarse fragments
- LL - Lower size limit, mm
- V% - Volume % coarser than LL (19 mm), or between indicated limits

Appendix C Positional Accuracy of NASIS Pedon Points

usersiteID	new_X	new_Y	nasis_X	nasis_Y	Position_accuracy	nasis_mukey	new_mukey
1977IN033067	675562.9128	4572581.186	675502.9128	4572631.186	78.10249676	160347	160321
1977IN033065	678596.6862	4576734.053	678606.6862	4576724.053	14.14213562	160320	160320
2013IN13500110	666463.9202	4455245.156	666473.9202	4455235.156	14.14213562	161308	161308
2013IN13500109	666515.8696	4455249.357	666525.8696	4455239.357	14.14213562	161308	161308
2013IN13500108	666563.136	4455250.4	666573.136	4455240.4	14.14213562	161308	161308
2013IN13500107	666610.4025	4455251.444	666620.4025	4455241.444	14.14213562	161308	161308
2013IN13500106	666660.0086	4455252.54	666670.0086	4455242.54	14.14213562	161308	161308
2013IN13500105	666704.8666	4455256.647	666714.8666	4455246.647	14.14213562	161308	161308
2013IN13500104	666752.133	4455257.691	666762.133	4455247.691	14.14213562	161308	161308
2013IN13500103	666797.0362	4455258.684	666807.0362	4455248.684	14.14213562	161308	161308
2013IN13500102	666842.052	4455253.51	666852.052	4455243.51	14.14213562	161308	161308
2013IN00300110	679595.7356	4569799.647	679605.7356	4569789.647	14.14213562	164721	164721
2013IN00300109	679580.1834	4569768.404	679590.1834	4569758.404	14.14213562	164721	164721
2013IN00300108	679635.0073	4569724.825	679605.0073	4569724.825	30	164721	164667
2013IN00300107	679600.4767	4569700.984	679610.4767	4569690.984	14.14213562	164721	164721
2013IN00300106	679579.9878	4569681.994	679589.9878	4569671.994	14.14213562	164721	164721
2013IN00300105	679552.3136	4569672.022	679562.3136	4569662.022	14.14213562	164721	164721
2013IN00300104	679553.2263	4569635.045	679563.2263	4569625.045	14.14213562	164721	164721
2013IN00300103	679553.836	4569610.342	679563.836	4569600.342	14.14213562	164721	164721
2013IN00300102	679557.0526	4569573.421	679567.0526	4569563.421	14.14213562	164721	164721
2013IN00300101	679560.2933	4569536.471	679570.2933	4569526.471	14.14213562	164721	164721
2012IN0698A03	622181.571	4509240.963	622161.571	4509250.963	22.36067977	161409	2567860