Changes in Land Use from 2005 to 2012 in Williston, North Dakota

FR 3262 - Remote Sensing
Heidi Eaves (ESPM-CRM) and Jennifer Schumann (CE)
Fall 2013

ABSTRACT

Since oil production took off on the Bakken Formation in 2006, the once quiet, farming community of Williston, North Dakota has rapidly changed in land use composition. The increasing population has also prompted high demands for housing, commerce, and transportation. The result of which is a rapidly change land use composition. The purpose of this study was to quantify these changes in land use and identify relationships between the changes by comparing data from 2005 (pre-oil boom) to data from 2012. NAIP and LandSat 7 images were obtained and composited for each year. The composite images were analyzed using a supervised classification in ERDAS Imagine. Then, the classified images were compared against each other in a change detection also processed using ERDAS Imagine.

Due to the scan line error of the LandSat 7 data, the study produced mixed results. While trends in changes in land use could be observed, they could not be quantified. The change detection process did not provide accurate data. Therefore, it was not possible to use to draw conclusions about relationships between land use class changes. In other words, it is unknown whether an increase in one class can be linked to a decrease in another class based solely on these data.

What can be derived from the data in this study are trends in the land use composition. By comparing the data from 2005 classified image and the 2012 classified image, it was possible to identify which land use classes were trending up and which were trending down. As was expected, there is an observable decrease in crop fields and native prairies and an increase in developed land and roads. Swales and wetlands demonstrated an unexpected increasing trend.

This study had to draw from secondary information to address these trends and determined what could be likely causes for their observation.

In the conclusion of this study, suggestions are discussed for further analysis of these data which could address the scan line error issue. If these suggestions were carried out, the subsequent change detection would provide useful information in quantifying relationships for the observed changes in land use.
INTRODUCTION

In 1950 the first oil well in North Dakota was erected in Antelope field in McKenzie County. Though the well did not produce enough to be commercially viable, it did put North Dakota on the maps for oil exploration. It wasn’t until 1995 that the U.S. Geological Survey assessed the oil reserves of North Dakota and discovered what is known as the Bakken Formation. The Bakken Formation consists of three layers: (1) the lower shale, (2) the middle sandstone, and (3) the upper shale. What oil producers are interested in are the lower and upper shale layers where as much as 35% of the layer’s composition is made up of marine organic matter. These layers are the petroleum source rocks from which the oil is extracted (Pollastro, 1995). As technologies in oil production advanced, the potential for extracting these oil reserves came to fruition.

EOG Resources were the first producers to establish wells on the Bakken Formation in 2006 using hydraulic fracturing (fracking) technologies. Other major oil producing companies quickly moved in and the number of well installations has grown exponentially. Along with the wells has come an influx of workers to install, operate, and maintain the wells. Additional people were needed work in the transportation and distribution sector of oil production. With the immense, growing work force came the families they belonged to and a need for bigger schools, more housing, retail centers, industry, and infrastructure development.

At the heart of all this growth, in the geographical center of the Bakken Formation, sits the small city of Williston, North Dakota (see Image 1). Boasting to have the only Wal-mart for over a 100 miles in any direction, this community is the largest city on “the [oil] patch”. The landscape has changed dramatically to meet the ever increasing demand for development. Prior to 2006, the area surrounding Williston was largely agricultural; crop production and range land. Since then, innumerable hotels, “man camps”, industrial buildings, and commercial buildings have sprung up in their place, changing land use composition.

![Image 1. Communities with Substantial Population in the Bakken Formation.](https://www.prweb.com)
The changes in land use composition are of particular interest to those who are tasked with planning urban development and managing natural resources. The purpose of this study was to quantify these changes in land use and identify relationships between the changes by comparing data from 2005 (pre-oil boom) to data from 2012.

The study needed to be conducted using remote sensing techniques. The work had to be completed in the time constraint of seven 5-hour time blocks with a budget of $0.00 (and a government shutdown). Analyses of raster data were processed using ERDAS Image 2013 software at the University of Minnesota.

MATERIALS
National Agriculture Imagery Program (NAIP) images were obtained from the USDA NRCS website. A pre-oil boom image was obtained from 2005 and the most recent post-oil boom image available was taken in 2012. These images were available from the US Department of Agriculture website with three spectral bands; blue, green, and red (USDA/NRCS, 2013).

Accuracy of classifications is greatly improved when infrared (IR) bands are included in an image classification. NAIP images are largely natural color only and did not include IR bands until 2007 (USDA/FSA, 2013). The availability of images with IR bands varies from state to state and are not uniformly available. LandSat images provide IR image data and are available for free download from the USGS website (USGS/EROS, 2013). Therefore, a combination of these data was required for improved classification accuracy.

To avoid temporal errors it was important to obtain LandSat images that coincided with NAIP capture dates; however, this information was not readily available. Therefore, assumptions were made about the capture date based on foreknowledge of the area and the amount of “greenness” of the photo. The only images available from the acceptable time frame with minimal cloud cover over the area of interest were LandSat 7 images obtained on 22 June 2005 and 27 July 2012.

LandSat 7 images have eight spectral bands. Bands 5, 6, and 7 were omitted because their bandwidths did not provide information that would contribute to the study. Retaining these layers would have resulted in longer processing times for the raster analyses.

NAIP images have a resolution of approximately one meter, LandSat 7 images have a resolution of approximately thirty meters. The discrepancy in resolutions was addressed by resampling the LandSat 7 images to one-meter using a nearest neighbor kerneling function in ERDAS Imagine.

The three layers of the NAIP images were stacked with the LandSat 7 images containing bands 1,2,3,4 and 8. The composite image provided the spectral layer information outline in Table 1 (USDA/FSA, 2013 and USDA, 2013).
### TABLE 1. ERDAS IMAGINE STACKED IMAGE SPECTRAL BAND LAYER ASSIGNMENTS

<table>
<thead>
<tr>
<th>Stacked Image Layer</th>
<th>Layer Source Image</th>
<th>Spectral Bandwidth</th>
<th>Pre-Processed Layer Spatial Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer 1</td>
<td>NAIP</td>
<td>“Natural Blue”</td>
<td>1 meter</td>
</tr>
<tr>
<td>Layer 2</td>
<td>NAIP</td>
<td>“Natural Green”</td>
<td>1 meter</td>
</tr>
<tr>
<td>Layer 3</td>
<td>NAIP</td>
<td>“Natural Red”</td>
<td>1 meter</td>
</tr>
<tr>
<td>Layer 4</td>
<td>LandSat 7 Band 1</td>
<td>0.45-0.50 µm</td>
<td>30 meter</td>
</tr>
<tr>
<td>Layer 5</td>
<td>LandSat 7 Band 2</td>
<td>0.52-0.60 µm</td>
<td>30 meter</td>
</tr>
<tr>
<td>Layer 6</td>
<td>LandSat 7 Band 3</td>
<td>0.63-0.69 µm</td>
<td>30 meter</td>
</tr>
<tr>
<td>Layer 7</td>
<td>LandSat 7 Band 4</td>
<td>0.77-0.90 µm</td>
<td>30 meter</td>
</tr>
<tr>
<td>Layer 8</td>
<td>LandSat 7 Band 8</td>
<td>0.52-0.90 µm</td>
<td>15 meter</td>
</tr>
</tbody>
</table>

The final product was two images; one for 2005 (Image 2) and one for 2012 (Image 3) with 1-meter resolution and eight spectral bands.

**Image 2. 2005 Composition Image for Area of Interest Showing Layer 1, Layer 2, and Layer 4.**
METHODS

The first attempt to categorize the pixels was through an unsupervised classification with 20 classes and 25 iterations. The area of interest covered nearly 100,000 acres and the classification took two days to process. The results were mixed and the natural break categories were not very useful. There were too many categories that were dedicated to one land use and not enough dedicated to another land use which resulted in muddled understanding of the overall land use composition.

A supervised classification resolved the categorical issue. It provided a much clearer understanding of the land use composition and was determined to be the best option for this project; however, prior to completing a supervised classification, the area of interest was reduced to 15,360 acres on the west side of Williston, north of the Missouri River. The processing time was reduced from two days to approximately 10 minutes.

The supervised classification process began with the selection of well-defined, land use categories. The following is a brief description of each category. These criteria were employed in selecting appropriate training sites for the supervised classification.

DEVELOPED. Initially, the goal was to differentiate between industrial/commercial, residential, and temporary residential (e.g. “man camps”, RV parks, etc). Attempts to
isolate these classes were unsuccessful. Rather than addressing the need for more distinct training sites, it was determined that combining these into one category, Developed, would be sufficient to extrapolate the changes in land use data.

**NATIVE PRAIRIE.** Western North Dakota is largely range land. Locally this often referred to as “native prairie”. It represents unmanaged vegetation that is native to the northern prairie region. Cattle and horses graze these prairies and contribute to the ecosystem much like the buffalo did before the area was settled by the Europeans.

**CROP FIELD.** This category identifies agricultural land used for crop production. Wheat is the primary crop in this area, but other grains such as oat, flax, canola, and sunflowers are also common. Differentiating between the types of crops in the agricultural fields may have caused temporal errors because most farmers raise crops in rotations. In other words, what is a canola field in 2005 may be a wheat field in 2012. It was sufficient for this study that the area be identified simply as crop field.

**BARE AG SOIL.** This could be considered a subclass of crop field. A change in category from crop field to bare ag soil, or vice versa, would not truly represent a change in land use. However, the spectral signature of this category would be unique from that of a crop field in production.

**WOODY.** Williston, North Dakota is an arid region; average precipitation is 14 inches annually. Trees and shrubs grow in small, dense pockets where water collects at high enough concentrations to support them. Woody plants includes any plant categorized as arborescent, shrubby, and suffrutescent. These species can be found along rivers and streams, stock ponds, or planted in rows as windbreaks in crop fields.

**SWALES/WETLANDS.** The area of interest is located on the northern fringe of the North Dakota Badlands. There are large plateaus with steep slopes that cause gullies to develop. These gullies are locally referred to as swales. Some swales are wide and expansive (15-20 feet) or narrow and channeled (1-5 feet). Consequently these areas have more silt and clay in the soil, receive a high proportion of hydration (due to runoff during snowmelt or rainfall), and have denser, greener vegetation. Some swales develop hydric soils and can be categorized as wetlands; some do not and are referred to as dry swales. It is not possible to differentiate between dry swales and wetlands from aerial imagery alone. For this study, the differentiation is not necessary.

**ROAD.** Roads through the area of interest have a lot of variables and were a difficult category to isolate in the classification process. Roads include county, state, and federal highways; city streets; rural (dirt) roads; two-tracks (between fields on section lines); and well-pad access roads. The spectral signature for the road depends on the material with which it was built: asphalt (black), concrete (white), dirt (light brown), or scoria (pinkish-red).
WATER. Water was another difficult category to isolate due to the variations in spectral signature. Flowing water in this area appears brown due to high turbidity. Ponded water has a much lower turbidity and absorbs more radiation as would be expected of water. Therefore, ponded water had to be isolated from rivers and streams. Most of the ponded water in this area are water treatment ponds for municipal waste or stock ponds used for raising cattle.

RIVER. This category represents all streams and rivers. As mentioned before, the rivers and streams have a lot of silt and clay suspended in the water causing a high turbidity which, in turn, affects the spectral signature of the water. It has a higher radiance than clear water and appears to be brown in images.

Before training sites could be selected for these categories, another issue had to be addressed. LandSat 7 had a mechanical failure in May of 2003 (Graham, 2004). The Scan Line Corrector failed resulting in images with fully intact data in the center of the image and scan lines that are void of data on the eastern and western thirds of the image. LandSat 7 images for the area of interest were subject to the scan line error. To compensate for this during the classification process, a set of training sites were selected for each category for pixels within the scan lines and for pixels outside scan lines (e.g. Native Prairie and Native Prairie Scan Line).

The number of training sites for each category varied depending on the size of the areas available for a category. For large crop fields, fewer training sites were used. For small tree rows, more training sites were used. The minimum number of training sites for any category was five and the maximum number was ten.

The supervised classification was run using maximum likelihood algorithms. The results of this process were a classified image for 2005 (Image 4) and a classified image for 2012 (Image 5).

Upon completion of the supervised classification, the 2005 and 2012 images were analyzed in ERDAS using a change detection.

RESULTS

Supervised classifications were performed on the 2005 composite image. Image 4 shows is the results of the classification process for the 2005 composite image.
The large, magenta stripes demonstrate the effect of the scan line error on the image classification. Major land uses in 2005 were Developed, Native Prairie, and crop field. Table 2 outlines the number of pixels assigned to each classification and the percentage of pixels within the image assigned a classification.

<table>
<thead>
<tr>
<th>2005</th>
<th>Number of Pixels</th>
<th>Percentage of Pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclassified</td>
<td>37,047</td>
<td>0.1%</td>
</tr>
<tr>
<td>Water</td>
<td>1,617,285</td>
<td>2.6%</td>
</tr>
<tr>
<td>Swale/Wetland</td>
<td>4,618,132</td>
<td>7.5%</td>
</tr>
<tr>
<td>Crop Field</td>
<td>11,161,154</td>
<td>18.2%</td>
</tr>
<tr>
<td>Road</td>
<td>1,473,673</td>
<td>2.4%</td>
</tr>
<tr>
<td>Woody</td>
<td>4,440,497</td>
<td>7.2%</td>
</tr>
<tr>
<td>Native Prairie</td>
<td>8,258,854</td>
<td>13.4%</td>
</tr>
<tr>
<td>Developed</td>
<td>27,618,561</td>
<td>44.9%</td>
</tr>
<tr>
<td>Bare Ag Soil</td>
<td>2,239,398</td>
<td>3.6%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>61,464,601</strong></td>
<td></td>
</tr>
</tbody>
</table>

Supervised classifications were performed on the 2012 composite image. Image 5 shows is the results of the classification process for the 2012 composite image.
As in the 2005 image, the large, magenta stripes demonstrate the effect of the scan line error on the image classification. The major land uses in 2012 were also Developed, Native Prairie, and crop field. Table 3 outlines the number of pixels assigned to each classification and the percentage of pixels within the image assigned a classification.

To quantify the changes in land use from 2005 to 2012 a change detection was performed. Table 4 outlines the changes in land use between the two years.
The columns with red titles identify the number of pixels assigned that class in 2005, the rows with blue titles identify the number of pixels assigned to that class in 2012. Where classes intersect, the number was printed in bold. This represents the number of pixels that remained in the same category from 2005 to 2012. Numbers not in bold print represent number of pixels that changed from one category to another.

For example, there were 1,617,285 pixels assigned to the Water class in 2005. That number changed to 1,504,573 in 2012. According to this chart, there were 732,036 pixels identified as Water that did not change between 2005 and 2012. However, there were 24,934 Swale/Wetland pixels, 28,341 Crop Field pixels, 14,749 Woody pixels, 17,823 Native Prairie pixels, 664,012 Developed pixels, and 1,471 Bare Ag Soil pixels from the 2005 classified image that were added to the Water category in 2012. Likewise, there were 23,904 Water pixels in the 2005 classified image that were changed to Swale/Wetland pixels in 2012. There were 3,500 Water pixels in the 2005 image that were changed to Crop Field pixels in 2012, and so on.

From this information the percentage of pixels for any particular class can be calculated. Determine the number of pixels that changed categories by subtracting the 2005 categorical total from the 2012 categorical total. Divide the number of changed pixels by the 2005 categorical total. The resulting number indicates the percent growth (positive number) or reduction (negative number) of a class.

**DISCUSSION**

This project had several significant obstacles to overcome and despite the researchers best efforts, not all issues were effectively addressed. As a result, the data quantifying the changes in land use and establishing relationships between land use changes from 2005 to 2012 are not valid. While the expected decrease in categories such as Crop Field and Native Prairie were observed, along with the expected increase in categories such as Roads and Developed; the change relationships and measurements were negatively affected by the Scan Line Error in the LandSat 7 satellite.
Training sites were designated for land use classes in the scan lines of the LandSat 7 images in an attempt to compensate for the lack of spectral data. Unfortunately, that did not resolve the issue. Pixels within the scan line had digital numbers (DN) for only three of the eight layers, i.e. NAIP “natural blue”, NAIP “natural green”, and NAIP “natural red”. Pixels outside of the scan line had DN for all eight layers. During the classification process, the pixels with only three DN contributing to its spectral signature were statistically similar to the eight-DN pixels of the Developed class. Therefore, all pixels with only three DN were classified as Developed (see Image 4 and Image 5).

Had the scan lines for the 2005 image and the 2012 image lined up, the inaccuracy of the Developed classification would have been mitigated. Unfortunately, this was not the case. Image 6 is mosaic of the 2005 composite image (left) and the 2012 composite image (right). Due to the offset of the scan lines, only a nominal amount of the error caused by the scan line was mitigated. Instead approximately 35% of the total area of interest, area not overlapping scan lines from either 2005 or 2012 images, produced statistically accurate results in the change detection process.

Image 6. Comparison of Position of Scan Line Errors in LandSat 7 Images Between 2005 and 2012 Images
In comparing Image 6 to the information in change detection matrix (Table 4), it is easy to conclude that while the data in the matrix is precise, it is not accurate, especially with regard to the Developed class. The matrix indicates only a 5% change in Developed pixels; however, Images 2 and 3 indicate that percent change should be significantly higher.

Despite the inaccurate calculations in the error matrix, it can still provide some useful data. Disregarding the Developed pixel figures allows the matrix to be used to identify trends in the data, especially when compared to the classification data in Tables 2 and 3. It is possible to conclude that up until the oil boom, the dominant land uses were agricultural; managed land for crop production or native prairies for cattle grazing. As expected, these land uses declined in the 2012 image.

Crop field pixels showed the greatest decline and it is reasonable to assume the pixels were largely changed to Developed pixels based on secondary information. Native Prairie land use is indirectly regulated by the Threatened and Endangered Species Act of 1973 (US Govt, 1988). The Fish and Wildlife Service define critical habitat for these species and areas of land fitting these descriptions are difficult to obtain permits to develop it. Wetlands and swales are protected by the Clean Water Act (CWA, 1972). The U.S. Army Corp of Engineers provides oversight on protecting all navigable waters and adjacent wetlands. They provide parameters to enforce the protection of these landscape features that often impede development of land on and/or near them. Crop fields are privately owned land. Unless the land owner has volunteered for conservation programs, such as the USDA’s Conservation Reserve Program, there are relatively few limitations on development. As a result, crop fields are the most desirable for expanding urban development.

An initially surprising trend was the increase in the Swales/Wetlands category. Historical data was reviewed to determine if this change could be the result of an increase or decrease in annual precipitation on the region. Records obtained from the Williston Airport for the six months prior to data acquisition for 2005 and 2012 reported a total precipitation of 10.53 inches and 9.00 inches respectively (NCDC, 2013). These data are not consistent with the observed increase in Swales/Wetlands; however, it would explain the decrease for Water. Instead, the increase in Swales/Wetlands could be explained by the effects of environmental policy. As part of Section 404 of the Clean Water Act, the Compensatory Mitigation clause requires entities to replace wetlands that are removed due to land development with a in-kind wetland somewhere else within the watershed (CWA, 1972). It is reasonable to believe this could be a driving force behind the increase because the programs guiding the Compensatory Mitigation clause would favor development of wetlands in Native Prairie landscapes or by expanding or restoring wetlands along streams and rivers; both of which are plentiful near Williston.

The North Dakota Public Service Commission’s Reclamation Division establishes revegetation standards for the state. Woodland areas, which North Dakota has a legal definition of, are included in the revegetation standards. These standards require that trees and shrubs, classified as Woody in this study, removed during land development must be replenished. The
replacement trees and shrubs must be at 80% of the original tree’s or shrub’s growth development within six years of their removal (NDPSC, 2003). It is reasonable to assume that the decrease in Woody areas for this study could be attributed to development. A large portion of the pixels designated as Developed are still under construction. Therefore, replacement trees and shrubs would not be planted yet. Furthermore, newly developed areas with completed construction would not be likely to reached the six-year deadline yet; therefore, the replacement trees and shrubs may be too small to be distinguishable from Native Prairie or Swale/Wetland classes.

Developed land trends can only be derived from the classification tables (Tables 2 and 3). This class showed a strong increasing trend as was expected. The population in Williston is growing at an accelerated rate. According to recent North Dakota State University study, the population of Williston has doubled since the 2010 census (Associated Press, 2013). This leads to significant demands for development; housing, retail outlets, etc. From 1997 to 2005, the average number of new buildings was approximately 15 per year. From 2010 to 2012 the average was 237 new buildings per year (City-Data, 2013). This supports the observed increasing trend.

It also to practical to assume the increase in the Developed land use class correlates with the decrease in the Crop Field class. From a civil perspective, public infrastructure (roads, buildings, etc) would be built in the most practical locations, meaning the cheapest locations possible. Private land that doesn’t require the demolition or relocation of some existing structure would be the most economical. Also, land that is subject to environmental regulations that restricts “buildable” areas would be less desirable because of the administrative costs of filing the required environmental reports, surveys, and permits.

The increasing trend in the “roads” category is consistent with what would be expected in an area that is experiencing significant growth and development. Generally speaking, significant growth cannot occur in sparsely populated areas unless roads are constructed. This is particularly true when the growth is caused by the extraction and distribution of a resource such as oil. The only way it can go to market is by road, rail, or pipeline. In addition to the need for additional road capacity for product distribution, there is also need for new and improved roads to serve communities that the growing population reside in. For every gas station, hotel, RV park, and home that is built, a driveway and additional road capacity must also be built. This increase in “roads” and “developed” could also be inversely related to the decreasing trend in “Crop Field” areas over the same time period, consistent with the idea that new roads and development are likely being built on what has historically been agricultural land. The trending increase in “road” and “developed” areas and the trending decrease in “native prairie” and “crop field” would be expected to continue, as long as the extraction of oil continues to be profitable.

The increase in “bare ag soil” could be explained by several factors. One possibility is that the digital numbers associated with “bare ag soil” are statistically very similar to that of “roads” or “developed”. It makes sense that a road having an aggregate surface may be indistinguishable from an aggregate surfaced parking lot, or a rocky soil with low moisture that has not been
planted, or possibly a well pad.

Though this study failed to deliver a quantified explanation of land use changes and the relationships between those changes in Williston, North Dakota, it did provide reliable trend data. These data are still useful for civil and environmental planning.

The trends seen in this study can be used by civil engineers to make informed decisions regarding city and infrastructure planning. Economic development is often stifled because of insufficient infrastructure, and untimely policy changes, to compensate for it. By monitoring the changes in land use in the Williston area, the current and future need for infrastructure improvements can be assessed and subsequent planning initiated.

The cornerstone of good environmental practice is summed up in one word: sustainability. The trends in this study would aid in actively monitoring changes in land use in the Williston area. Researchers and policy makers could use it to gage the effectiveness of environmental protection, mitigation, and restoration efforts. Too much expansion or diminishment of any particular land use class would alert experts to reevaluate how they perceive the protection of ecosystems. From there, quality decisions could be made for policy changes.

**SUGGESTIONS FOR FURTHER DATA ANALYSIS**

As previously discussed, this study fell short in quantifying the changes in land use and in drawing quantitative correlations between observed changes in the area of interest. However, if time constraints has not been a limiting factor there were possible steps to take to address the issue. Rather than attempting to classify areas within the scan line error as part of the composite NAIP-LandSat 7, these areas should be omitted (i.e. clipped out) using the AOI feature in ERDAS Image resulting in, what could be called, Full Spectrum Stripes. The same process should be conducted on the NAIP imagery only omitting the areas that overlapped with the Full Spectrum Stripes, resulting in what could be called Scan Line Error Stripes.

Each image, Full Spectrum Stripes and Scan Line Error Stripes, should then be subjected to the same supervised classification as the composite image was from this study. An accuracy assessment at this stage in the process would be advisable; especially for the Scan Line Error Stripes image because it lacks the infrared band. Adjustments to training sites should be made to improve accuracy.

Once the classification process is completed and accuracy is acceptable, the classified images should be mosaiced resulting in a composite image of Full Spectrum Stripes and Scan Line Error Stripes for both 2005 and 2012. Conducting a change detection from these images would provide a statistically credible quantification of changes in land use for this area.
ACKNOWLEDGMENTS (OPTIONAL)
Special thanks to Don Kilberg for his assistance in overcoming data limitations and providing guidance for which analyses to conduct and which analyses to omit.

LITERATURE CITED


Reclamation Division . July 2003. Standards for Evaluation of Revegetation Success and Recommended Procedures for Pre- and Postmining Vegetation Assessments [Internet]. North


APPENDIX A
Lab Notes from Raster Analysis
10/25/2013
Procedure
* Downloaded NAIP imagery from the NRCS website
(http://datagateway.nrcs.usda.gov/GDGOrder.aspx)

Using a shapefile, we created an AOI and clipped the original NAIP to fit our AOI - this was due to file size restrictions from the computers in the labs

- First pass was an unsupervised classification of 20 classes and 50 iterations; this gives us an idea of how easy it will be to differentiate land uses based on NAIP data without IR options.

CLASSES
1. Ag fields
2. Native prairie/Pasture land
3. Wetlands?
4. Industrial/Commercial
5. Well pads
6. Residential - Permanent
7. Residential - Temporary
8. roads - paved
9. Roads - unpaved (scoria)
10. Roads - two-tracks
11. Woody Draws
12. Rivers/Coulees/Stockponds

11/1/2013 - We started an unsupervised classification of each dataset (2005, 2009, 2012) in the early afternoon. The AOI we were using was roughly a 5 mile radius around Williston. The first pass at an unsupervised classification (20 classes, 25 iterations) took almost 2 days to complete. The issues is that with 1-meter resolution, a 5 mile radius gave us too many pixels to categorize.

11/8/2013 - Because we were concerned about the ability to differentiate various types of vegetation, we decided to explore the possibility of using Landsat data (for the IR bands) instead of or in addition to the NAIP data we had already acquired. After retrieving and clipping the data, stacking the layers and performing an unsupervised classification of the data, we determined that the resolution was not fine enough to use with our defined AOI.

The city of Williston has a radius of approximately 2km. The nearest sizable community (over 500 people) is Watford City which is nearly 50km away (as the crow flies). The point is that a 30m resolution would work if the community we were studying was larger, i.e. Twin Cities Metro Area. But because our region is smaller, we need to use a fine-scale image to locate those changes.
We determined that we need to choose one data source or the other, because it would be too difficult to merge data and use both. We saved the Landsat data, but tabled the idea for the time being. We will pursue the NAIP data, and perform a supervised classification.

11/8/2013 - supervised classification
Concern - could we differentiate between vegetative cover types with the NAIP data knowing we didn’t have and NIR band to work with?

The answer is mostly.

We first tried doing an unsupervised classification (25 classes, 10 iterations) to see if there were natural breaks we could use. Found out that that wasn’t enough classes for the computer to differentiate between shades of green to the degree that we needed.

We decided to try the supervised classification to see if training sites would help differentiate between shades of green...more success but accuracy is largely unknown right now.

We did two supervised classifications of the 2012 image. The first was without simplifying the training sites; the second was with simplifying the training sites. The results were mixed. Most important take home from this is that some classes can be simplified and still be accurate and others cannot. E.g. Native Prairie can and Streams cannot.

Classes used:
Bare Ag Soil - very difficult to tell the difference between this and industrial areas
Native Prairie - Mostly consistent
Crop - Mostly consistent but wet swales were often called crop field in the simplified
Woody - in the simplified these were often confused with stockpond, in the unsimplified more accurate
Industrial-Residential - very difficult for differentiation between the two in both supervised classification runs
Roads - Streams called roads in simplified
wetlands - UNKNOWN

REPLACING WETLANDS WITH SWALES

11/11/2013 - We reviewed our project timeline this morning, and decided that we will not have enough time to include the 2009 data in our project. After some discussion, we agreed that it is redundant data anyhow (since we are doing change assessment from 2005 to 2012). By excluding the 2009 data, we will free up some time that can instead be used on merging / integrating Landsat data (for the IR band info) with NAIP data. After consulting with Don, it sounds like it may be possible to re-sample the Landsat data from a 30 m resolution, into a 1m resolution to be compatible with the NAOP data we have already acquired.
After consulting with Don, we decided to create a stacked image of the NAIP with the LandSat. The only LandSat data we could obtain that was temporally compatible with the NAIP was LandSat 7 and it has the strip issues.

We pulled bands 1, 2, 3, 4, and 8 from LS7.

Final image band layers in clip below:

To classify we’re going to have to address the stripe issues. We propose to do two classes for each landcover type.

i.e. cropfield A - no stripe, cropfield B - stripe, etc.

We also adjusted the resolution on the LandSat to 1m using Resample Pixel with nearest neighbor kerneling.
11/15/2013
We are now using only the 2005 and 2012 data sets.

We created training sites that were for the stripes and training sites that were not for the stripes. Categories:
Native Prairie
Native Prairie Stripe
Developed (industrial/residential)
Developed (industrial/residential) Stripe
Bare Ag Soil
Bare Ag Soil Stripe
Swales/Wetlands
Swales/Wetlands Stripe
crop field
crop field Stripe
Woody
Woody Stripe
Road
Road Stripe
Water 1-6
Water 1-6 Stripe

When we ran the classification, all pixels in the stripes were classified as Developed. We suspect it’s because all pixels have DNs from 8 layers contributing to the total DN “value” of the pixel. The total DN “value” for pixels in the stripes only have 3 layers contributing to the DN. Therefore the 3-layer value must have most closely matched the 8-layer value for Developed; hence, all pixels in the stripes were categorized as Developed.

Other complications:
2005 Landsat Stripes are not in the same place as 2012 Landsat Stripes
NAIP 3-band classifications have significant (though uncalculated) margins of error

These complications will affect our CHANGE DETECTION

Possible Solutions:
1.) Omit the strip areas from the change detection
   Pros
   - Quick solution
   - Accuracy will be greatly improved
   Cons
   - We would only be classifying small strips of land; i.e. information would not be very useful
2.) Run two classes - one for Landsat Stripes and one for NAIP stripes, merge the two thematic images

Pros
- Can do a change detection over the whole area of interest
- We already have the training sites selected for the specific portions of the NAIP (areas in the stripes) that we are interested in

Cons
- Is it reasonable to merge two clips? Is it possible?
- How will this affect the accuracy of the change detection?

11/18/2013
Started an accuracy assessment (per lab 11 instructions) on the 2005 classification. Having multiple water classes is a problem, because we only have “water” (or maybe “water” and “river”??) for an option in our reference data. We should have the following classes to use for accuracy assessment:
1. Developed (magenta)
2. Native Prairie (orange)
3. crop field (green)
4. Bare Ag Soil (sienna)
5. Woody (dk green)
6. Swales/Wetlands (blue)
7. Road (red)
8. Water (cyan)
9. River (turquoise) *** we could include this in “Water”, but they seem like 2 different categories

I saved my work, but stopped, because we need to figure out the water problem - JLS

11/20/2013
After discussing the problem of multiple water classes, we decided to combine all water into one class before performing a change detection or accuracy assessment. We did the change detection (2005 - 2012) and exported an ASCII file, for use in Excel to analyze the data.

Now that the bulk of the work is complete, we decided to attempt an accuracy assessment for our original NAIP classification, and an accuracy assessment for the hybrid (Naip + Landsat 7), to quantify our earlier realization that NAIP was not going to be accurate enough for a change detection.

11/25/2013
Don advised us that we did not have enough time to complete an accuracy assessment for this study. We decided to follow his advice and work on writing our final report and prepare our presentation.