Corn and Soybean Differentiation Using Multi-Spectral Landsat Data, Freeborn County Minnesota, 2008-2013

Matthew McGuire and Andrew Munsch

Forest Resources 3362 / 5262

Remote Sensing of Natural Resources and Environment

Fall, 2013
In a time of slow economic growth, any sign of good news is welcome. That the United States is the world's largest producer and exporter of corn and soybeans is encouraging. Minnesota’s contribution to the nation’s overall agricultural production consistently ranks high. According to figures just released from Minnesota Department of Agriculture, Minnesota’s agricultural exports reached a record $8.2 billion in 2012. Top exported commodities for the year were soybeans at $2.2 billion and corn at $941 million. That volume places Minnesota third in soybeans and fourth in corn against all states.

**Fig 1: Top corn producing counties in Minnesota, 2011.**

Many of the highest producing counties are located in southern Minnesota. Corn and soybean are not only a high percentage of agricultural acreage in these counties, they also make up the majority of the total land area. These crops are a powerful economic force with an impact across the globe. Regionally, corn is a driving force behind the ethanol industry and ancillary corn derivatives. At the same time global demand for soy products has been increasing.

Because of the importance of these crops, it makes sense that we would want to keep track of how much of each of these crops that farmers are planting for aid in short-term
policy making and to predict economic effects. Remotely sensed multi-spectral data gives us a potential tool to assess what crops farmers are growing before they are harvested to see if early agricultural forecasts are correct, and to make policy changes if they are not.

In this investigation, we experimented with several procedures using remotely sensed data, as well as with already existing products derived from similar data to see if corn and soybean fields could be easily differentiated from each other using their spectral properties part-way through the growing season in order to predict current-year trends in proportion of these two crops being grown.

Study Area and Data Used:

Our study area was Freeborn County Minnesota, located along Minnesota's southern border with Iowa in a region almost totally devoted to corn and soybean agriculture (see figure 1, below).

![Fig 2: Location of Freeborn County in Minnesota.](image)

For data, we used to remotely sensed Landsat images, one from 2008 taken by the Landsat 5 Thematic Mapper, and one from 2013 taken by the Landsat 8 Operational Land Imager. Both images were taken in mid July (7/16/2008, and 7/14/20013 respectively) and were coincidentally the most cloud-free available of the study area during the growing season. The 2008 image is virtually cloud-free, while the 2013 has a smattering of cumulous and high cirrus (see figure 2 below).
Part 1 - Using Pre-Clased Images to Determine Recent Trends in Corn / Soy Rotation

Although crop data for Freeborn county is readily available for the last few years, this data does not give information about where these crops are grown within the study area, nor the history of the fields crop rotations. To determine this information, we used the Cropland Data Layer (CDL) produced yearly by the U.S. Department of Agriculture's National Agricultural Statistics Service. This data layer is a land-cover classification focused on crop types, and advertised to be highly accurate for corn and soybean fields.

The CDL Images for the years 2008 through 2012 were acquired and clipped to a polygon shapefile of the Freeborn County study area. The files for 2008 and 2009 have a spatial resolution of 56 meters, while the newer images from 2010 through 2012 have a 30 meter spatial resolution. The images come color mapped, with a golden yellow representing corn, and a dark
green representing Soy (see figure 4 below).

Fig 4: 2008 CDL Image of Freeborn County.

By examining the histograms of the clipped images, the total areas classified as corn or soy could be determined for each year. The results are given in table 1 and chart 1 below.

Table 1: Corn and Soy As Classified in the CDL, Freeborn County, 2008-2012

<table>
<thead>
<tr>
<th></th>
<th>2008(56m)</th>
<th>2009(56m)</th>
<th>2010(30m)</th>
<th>2011(30m)</th>
<th>2012(30m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn Pixels</td>
<td>240255</td>
<td>222150</td>
<td>835060</td>
<td>888529</td>
<td>935077</td>
</tr>
<tr>
<td>Corn (Hectares)</td>
<td>75343.968</td>
<td>69666.24</td>
<td>75155.4</td>
<td>79967.61</td>
<td>84156.93</td>
</tr>
<tr>
<td>Soy Pixels</td>
<td>182799</td>
<td>198793</td>
<td>636293</td>
<td>606028</td>
<td>556096</td>
</tr>
<tr>
<td>Soy (Hectares)</td>
<td>57325.7664</td>
<td>62341.4848</td>
<td>57266.37</td>
<td>54542.52</td>
<td>50048.64</td>
</tr>
</tbody>
</table>
It appears that since 2009, the total acreage of corn fields has been increasing steadily, while the acreage of soy fields has been decreasing at nearly the same rate, with corn always being the dominant crop planted in the landscape.

Next, pairs of CDL images from adjacent years were entered as input to the "Matrix Union" dialog in the ERDAS Imagine software package. The "Matrix Union" function compares the pixels between two images and outputs an image consisting of thematic classes for each possible combination of classes between the two images. The pixel counts from these classes tell us how many pixels transitioned between corn and soy, stayed the same classification, or transitioned to or from corn and soy from some other classification category. The results are given in table 2 below:
Table 2: CDL Class Changes, Corn and Soy, Freeborn County, 2009-2012

<table>
<thead>
<tr>
<th>Values in Hectares</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn to Corn</td>
<td>18763.315</td>
<td>18812.34</td>
<td>23880.24</td>
<td>30494.79</td>
</tr>
<tr>
<td>Soy to Corn</td>
<td>48021.882</td>
<td>49235.49</td>
<td>51904.44</td>
<td>49835.97</td>
</tr>
<tr>
<td>Other to Corn</td>
<td>2881.043</td>
<td>7107.57</td>
<td>4182.93</td>
<td>3826.17</td>
</tr>
<tr>
<td>Soy to Soy</td>
<td>6251.616</td>
<td>6312.24</td>
<td>3124.62</td>
<td>2269.08</td>
</tr>
<tr>
<td>Corn to Soy</td>
<td>52267.085</td>
<td>45882</td>
<td>48948.03</td>
<td>46097.01</td>
</tr>
<tr>
<td>Other to Soy</td>
<td>3822.7838</td>
<td>5072.13</td>
<td>2469.87</td>
<td>1682.55</td>
</tr>
<tr>
<td>Corn to Other</td>
<td>4313.568</td>
<td>4971.9</td>
<td>2327.13</td>
<td>3375.81</td>
</tr>
<tr>
<td>Soy to Other</td>
<td>3052.2684</td>
<td>6793.755</td>
<td>2237.31</td>
<td>2437.47</td>
</tr>
</tbody>
</table>

Figure 6: Corn and Soy Rotation Trends Derived From CDL Images of Freeborn County, 2008-2012

From the rotation graph above, we can see that the amount of land being maintained as corn
from year to year was increasing, while the amount of land remaining soy year to year was decreasing. At the same time, the amount of corn being rotated to soy was decreasing, and the soy fields being rotated to corn was increasing, although the amount of rotation remained high overall (see chart 2). All of this information points to a greater emphasis on the growing of corn over soy between 2008 and 2012. In fact, 2012 was the 5-year high in hectares of corn grown, while it was also the 5-year low for soy.

Part 2 - Quick-and-Dirty 2-Class Unsupervised Classification of 2008 LandSat 5 Thematic Mapper Data

For the next part of our investigation, we did an unsupervised 2-class classification of Landsat 5 Thematic Mapper imagery taken of the study area on July 2008 to see how well corn and soy could be differentiated spectrally, and use the 2008 CDL image as "psuedo" ground truth for comparison to see how well we did.

Processing Steps (All Done in ERDAS Imagine):

1. The 2008 CDL image was thematically recoded into a mask so only pixels representing corn and soy were included; all other pixels were recoded to zero.
2. The CDL mask was applied to a 2005 Landsat 5 image (luckily cloudless!) taken of Freeborn County on July 16, 2008. The TM image was stacked with all bands except for Band 6 (the thermal infrared band).
3. An unsupervised classification was run on the masked Landsat image using the K-Means method with 2 classes.
Results:

Figure 7: Classified Image with 2 Classes from 2008 masked imagery

A quick look at the histogram of the resulting 2-class thematic raster image showed that “Class One” had a significantly greater incidence (pixel-wise) than the other. Our previous analysis would lead us to assume that corn fields should occupy more area. Upon making this assumption, we named class 1 “corn”, and class 2 “soy”. We then ran a "Matrix Union" operation in ERDAS to compare the CDL image with the results of our "quick and dirty" classification. The results are summarized in the following error matrix (table 3):
Table 3: Accuracy of 2-Class Unsupervised Classification of 2005 Landsat Image Compared with 2008 CDL Image

<table>
<thead>
<tr>
<th>Reference Class</th>
<th>Pixels Classified As:</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Corn</td>
<td>Soy</td>
<td>Total</td>
<td>Producer's Acc</td>
</tr>
<tr>
<td>Corn</td>
<td>777038</td>
<td>59587</td>
<td>836625</td>
<td>92.88%</td>
</tr>
<tr>
<td>Soy</td>
<td>81344</td>
<td>555095</td>
<td>636439</td>
<td>87.22%</td>
</tr>
<tr>
<td>Total</td>
<td>858382</td>
<td>614682</td>
<td>1473064</td>
<td></td>
</tr>
<tr>
<td>User Accuracy</td>
<td>90.52%</td>
<td>90.31%</td>
<td></td>
<td>90.43% Total Accuracy</td>
</tr>
</tbody>
</table>

The results of this most basic of classifications were surprisingly good, although this shouldn't come as too much of a surprise due to how this experiment was set up. To investigate this further, we created an NDVI (Normalized Difference Vegetative Index) image from the Landsat 5 imaged used in the classification, and visually compared it to the CDL image (see figure below). The NDVI value is essentially the difference between the near-infrared and red returns normalized by their sum. Actively growing plants absorb energy in the red part of the spectrum, but reflect energy in the near infrared, resulting in high NDVI values for growing vegetation, and low values for bare soil or other substances that absorb red and near infrared energy nearly equally. Different NDVI responses from different crops can arise from differences in their relative stages.
of growth (Campell and Wynne, 2011).

Figure 8: Comparison of NDVI of Landsat 5 Image Compared with 2008 CDL

The NDVI image is on the left, and the CDL image is on the right. The brighter fields in the NDVI image seem to correspond with the fields classed as "corn" in the CDL image, while the soybean fields are distinctly darker. Higher NDVI values indicate greater returns from live vegetation, so it would seem that at this point in the two crop's growing cycles (in mid July), that the corn's leaves cover a greater portion of the underlying soil than soybeans. The soy plants are likely not as far along, and thus the radiance from the soil has a large effect on the spectral signature of the field. This allows the corn and soy fields to have very different spectral signatures from each other at this point in the growing season, allowing us to easily differentiate them. Further evidence can be gleaned from the “Signature Mean Plot”, which gives the mean digital number value of each of the bands in the multi-spectral image of the pixels included in a class (see figure 9 below).
From this plot, we can see that “Class 1”, or corn, has a definite spike in band 4, which is the near-infrared band, with a minimum in band 3 (red). “Class 2”, or soy, also exhibits a minimum at band 3, but much less pronounced, and it’s peak is in band 5, or the the short wave infrared band.

**Part 3 - Attempting to Estimate Corn and Soybean Acreage for Freeborn County for 2013 Using Landsat 8 Operational Land Imager Data**

For the next part of this investigation, we attempted to use an unsupervised classification method on 2013 multi-spectral imagery from the new Landsat 8 Operational Land Imager to see if we could get comparable results to our 2008 image classification, and if not, why not. The big
difference with this iteration is that we had no pre-classed image to use as "pseudo" ground truth. Instead, we had final acreage statistics for corn and soybeans published in November 2013 by the USDA.

**Processing Steps:**

1. The 2012 CDL image of Freeborn County was used to create a mask of corn and soy fields using a thematic recode operation. This mask was created to limit the study area in Freeborn County to areas that had been planted as corn and soy the year before under the assumption that the corn/soy growing area was not changing by more than 10% year-to-year based on our earlier Matrix Union operations.

2. A band stack was done on the Landsat 8 image using bands 2 through 7 (visual spectrum and all near-infrared bands). This image was then masked to the 2012 CDL corn and soy crop areas.

3. An unsupervised classification was done on the masked Landsat 8 data in ERDAS Imagine using the Isodata Algorithm, with between 2 and 8 classes allowed using all other setting at default. An 8-class classification was done instead of the "quick-and-dirty" 2-class classification done on the 2008 Landsat data because of the unknown nature of this year's image, as well as to allow more analysis before doing aggregation of the output classes.
Results:

Figure 10: Results of 8-Class Unsupervised Classification of Masked 2013 Landsat Imagery

The unsupervised classifier worked as advertised and gave us 8 classes to work with. The appearance of the classified image when viewed at a large scale was very mottled, with multiple classes occurring in distinct fields. (see figure 11 below).
From looking at the signature-means plot of the classes, it became clear that several of the classes were related (see figure 12 below).
Several of the classes had distinct peaks in the near-infrared (band 5), with very low red (band 4) response, seeming to indicate growing leafy vegetation. Many of the other classes had their peaks in the short-wave infrared (band 6), with relatively flat response in the visual spectrum. This would seem to indicate some vegetation, but with varying effects from bare soil. Finally, there was an obvious class for the clouds, where the response in the visual bands was much greater than the others.

Comparing the class's signature means plots to the corn and soy class plots from our initial 2-class analysis, the non-cloud classes were recoded into "corn" (classes 3, 4, 5, and 7) and "soy" categories (1, 2, and 6). Class 8 remained independent as “clouds”. After the recode, the
pixels totals were converted to acres and compared to initial USDA crop planting estimates released in November, 2013. The results of this initial analysis are below:

Table 4

<table>
<thead>
<tr>
<th>2013 Initial Crop Estimates</th>
<th>Pixels (Classified Image)</th>
<th>Estimated Acres</th>
<th>Actual Acreage Planted</th>
<th>Error%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn</td>
<td>772472</td>
<td>171793.5242</td>
<td>151705.71</td>
<td>13.20%</td>
</tr>
<tr>
<td>Soy</td>
<td>714161</td>
<td>158825.4785</td>
<td>84240.07</td>
<td>88.50%</td>
</tr>
</tbody>
</table>

As you can see, our classification over-classified corn, and almost doubled the amount of soy planted! This seems to indicate that our original assumptions underlying our ability to differentiate corn and soy fields based on corn's greater spectral response in the near infrared at this point in the growing season is not valid in this case. Further investigation revealed that the spring of 2013 was exceptionally wet in southern Minnesota, preventing normal planting schedules due to wet fields. A USDA crops and weather report form July 15 (the same week as our image) stated that in Minnesota corn was 15 inches shorter than normal, and that 21% of soy was blooming compared to the 5-year average of 40%. Crop data from the USDA also stated that 77,191.65 acres of corn were prevented from being planted in 2013, while 27,738.38 acres of soybeans were prevented. It appears likely that the bare fields we classified as "soy" could have been either corn or soy fields, but because of stunted growth could not be differentiated from each other.

Another way to see the difference in the crop conditions in 2013 in comparison to the same time in 2008 is to compare the histograms of NDVI images derived from each Landsat image after it
has been masked to the corn-soy study area for each year.

Figure 13: NDVI Histograms of 2008 (left) and 2013 Landsat Study Area Images

In the above histograms, notice how different the distributions are - the 2008 NDVI image has a very high skewed peak, while the 2013 histogram shows two peaks, and a much lower maximum NDVI value (note the different horizontal scales of the NDVI values). The mean NDVI value for 2008 was .610, while the mean for 2013 was .331.

In response to this information, we re-classified class 1 from the original 8-class unsupervised classification to “unplanted field due to its over-all low spectral response.” Our new results were as follows:

Table 5

<table>
<thead>
<tr>
<th>2013 Initial Crop Estimates</th>
<th>Pixels (Classified Image)</th>
<th>Estimated Acres</th>
<th>Actual Acreage Planted</th>
<th>Error%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn</td>
<td>772472</td>
<td>171793.5242</td>
<td>151705.71</td>
<td>13.20%</td>
</tr>
<tr>
<td>Soy</td>
<td>501850</td>
<td>111608.6798</td>
<td>84240.07</td>
<td>32.50%</td>
</tr>
</tbody>
</table>
Our soy estimate was now lower, but still very high compared to the actual value. It is likely that we were not able to differentiate soy fields at this point in 2013 due to the stunted growing season.

**Conclusions**

From these two analyses, it appears that in some years corn and soy are easily differentiated from each other by their spectral signatures. However, any disruption to the normal growing process can throw this off. A more robust approach would be to use several images over time to plot changing NDVI levels for fields over the growing season, and then match these profiles to known crop growing profiles. Delineated field data would also be helpful to allow for finding areas that should have a homogenous crop make-up. Lastly, a digital elevation model could be used in wet years to determine low-lying areas in fields that would be unable to be planted, and could thus be set aside during the classification process. A large obstacle to the multi-temporal approach is the relatively low temporal resolution of medium-resolution multi-spectral data, and the tendency for croplands to be covered by clouds often during the growing season in wet years. Despite these limitations, basic vegetative indices could be carried out on Landsat data as it is released to evaluate the overall health of agricultural areas to determine whether or not early-year planting forecasts were accurate, and allow for better short-term agricultural policy decisions.

**Sources:**


