*Presentation at the International Geoscience and Remote Sensing Symposium*
*July 19 – 27, 2007, Barcelona, Spain (In Presss)*

# Crop Classification in the U.S. Corn Belt Using MODIS Imagery

Paul C. Doraiswamy, Alan J. Stern
ARS–Hydrology and Remote Sensing Lab
US Department of Agriculture
Beltsville, MD, USA

Bakhyt Akhmedov
Science Systems and Application, Inc.
Lanham, MD, USA

***Abstract*—Landcover classification is essential in studies of landcover change, climate, hydrology, carbon sequestration, and yield prediction. The potential for using NASA's MODIS sensor at 250-meter resolution was investigated for USDA's operational programs. This research was conducted over Iowa and Illinois to classify corn and soybean crops. Multitemporal 8-day composite 250-meter-resolution surface reflectance product time series were used to generate the NDVI data, which were used to differential between corn and soybean crops in the U.S. Corn Belt. The results of the MODIS-based classification were compared with the Landsat-based classification for the 2-year period. The overall classification accuracy for Iowa was 82%, and for Illinois 75%. In conclusion, this method has been used successively during the 2002–2006 years to develop crop classifications and products for crop conditions and potential yield maps for Iowa and Illinois.**

***Keywords—MODIS Classification; Data filtering; Crop Classification***

## I. INTRODUCTION

Landcover classification uses both supervised and unsupervised approaches and the decision tree technique. The decision tree approach is increasingly being used for classification purposes. The detailed methodology of this approach is described in [1]. Comparison of the two methods for landcover classification in global, continental, and regional scales was performed by [2]. These studies show that decision tree algorithms consistently outperform the maximum likelihood and linear discriminate classifiers with regard to landcover classification accuracy. At the same time, different techniques are developed for improving the decision tree method. Other researchers suggested the stochastic gradient boosting method, which refined the standard decision tree method and increased accuracy of IKONOS classification by 11%, and Probe–1 hyperspectral classification by 10% [3].

Vegetation indices (VI) for classification reduce errors caused by atmospheric conditions that strongly influenced visible and near-infrared reflectance. Knowledge of the vegetation phenology facilitates development of landcover classification using vegetation indices as a normalized difference vegetation index (NDVI). Time series (or multitemporal) VI data describe phenology of the vegetation, and have been used widely in landcover classification [4], [5], [6], [7]. In this research, we applied a simple decision tree approach to select corn and soybean pixels from Moderate Resolution Imaging Spectrometer (MODIS) images and developed the method to discriminate them from noncrop vegetation. The results of the classification can be used in operational yield prediction.

## II. MATERIALS AND METHODS

### *A. Study Area*

Fig. 1 shows the Corn Belt in the central U.S., with the States of Iowa and Illinois, where the classification study was conducted. About 70% of Iowa and Illinois is primarily cropland; corn and soybeans are the predominant crops. The field sizes average 64 hectares (0.5x0.5 miles). Crops are grown under rain-fed conditions, and crop planting is completed by the end of May, with corn planted about 2 weeks earlier than soybeans. Crop maturity occurred by late September.

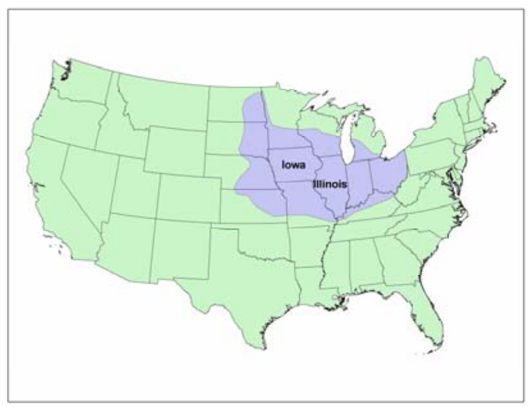


Figure 1. Study area of Iowa and Illinois, part of the U.S. Corn Belt

### *B. MODIS Data and Processing*

The 8-day composite 250-meter-resolution MODIS surface reflectance product was acquired through the NASA Distributed Active Archive Center (DAAC), EROS Data Center, Sioux Fall, SD, for the years 2002–2006 and for day of year (DOY) from 121–305 (MOD09Q1 collection 4). The images were reprojected from integrated sinusoidal (ISIN) projection to the standard universal transverse Mercator (UTM) projection for zone 15 (Iowa) and zone 16 (Illinois) using the MODIS reprojection tool. The images were imported in ERDAS image format, and NDVI was calculated for each image to develop the time series.

*C. MODIS Data Filtering*

Besides cloud cover, there were three kinds of errors associated with the standard MODIS data product—georeferencing, atmospheric correction, and bidirectional reflectance function (BRDF) effect. Because image processing cannot fully correct these errors, the 8-day composite images frequently are discontinuous (patches) caused by reflectance differences between single-day images used. Atmospheric correction errors are partly reduced using a normalized vegetation index such as NDVI. Errors in reflectance caused by atmospheric conditions always decrease the NDVI value. To eliminate these decreases, we used multiple steps in the Savitzky–Golay filtering technique [8] adapted to develop the upper envelope of the NDVI profile. This filter was applied to smooth every pixel's time series profile through the crop season [9]. The Savitzky–Golay method uses a moving window; in each window, noisy values are approximated by polynomial.

*D. Crop Classification*

The initial step in the classification algorithm is separating the noncrop area using the decision tree algorithm shown in Fig. 2. This algorithm is applied to each pixel in NDVI time series.

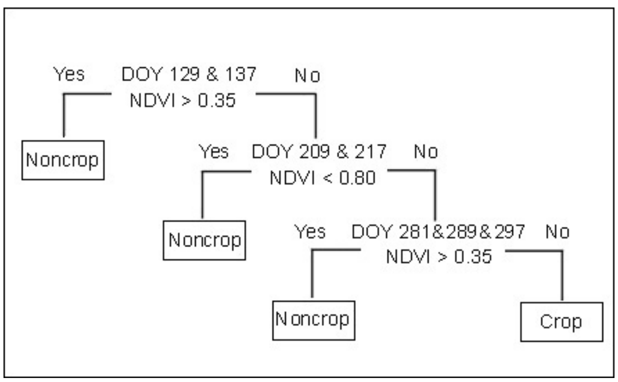


Figure 2. Decision tree algorithm to extract crop pixels

In the second step, corn and soybean pixels are separated using their difference in the NDVI profile from DOY 137 to DOY 209. The maximum difference between corn and soybean NDVI occurs on DOY 169 and 177. To decrease the errors in discriminating corn and soybean, it is better to take into account the adjoining DOYs by approximating data with curve. In [10], the authors used a piecewise logistic function to describe the phenology during the vegetative phase of crop development. Here we used a third-power polynomial

$$f(t) = c_1 + c_2 t + c_3 t^2 + c_4 t^3 \qquad (1)$$

Where $c_1$ , $c_2$ , $c_3$ , $c_4$ are coefficients and t is DOY.

Using a polynomial has an advantage in determining the coefficients. To find them, we will have a set of linear equations that can be easily solved. This is important when the NDVI time series for pixels have errors. In the case of nonlinear equations, we have to provide initial values for coefficients; errors in data may result in failure in the least squares method.

The first derivative of this polynomial represents the green-up rate during the vegetative phase of the crop.

$$f'(t) = c_2 + 2c_3 t + 3c_4 t^2 \qquad (2)$$

The second derivative of the polynomial represents the green up acceleration.

$$f''(t) = 2c_3 + 6c_4 t \qquad (3)$$

The green-up rate for corn pixels begins early and the NDVI profile between DOY 166 and 177 starts to level off at the end of this period; the trajectory is convex. For soybean pixels, the green-up rate still increases in progress during this period; the NDVI profile trajectory is concave. The second derivative of mixed corn and soybean pixels is between them, and according to their percentage in the pixel.

In the second step, we divide the pixels that passed the first step into two groups that represent corn and soybean classes. Second-derivative values between DOY 166 and 177 have large spatial variations within a region (county, State) because of differences in sowing date, growing conditions (soil type, soil moisture), and errors in remote sensing data. Distribution of the second-derivative value is continuous; there is no exact value that distinguished corn and soybean pixels. The exact value separating corn from soybean pixels was determined by the following procedure. The study region was divided into 50x50 km sections, and the pixels were separated into two classes to represent corn and soybean crops based on the second-derivative value. State-level reports of crop acreage are prepared each spring by the USDA. A breakpoint to separate corn and soybean crops is assessed from this report.

Values less than a breakpoint number represented corn, and greater values were soybeans. The breakpoint within each section of the study region was calculated based on reported distribution of crops within the State. The corn area in Iowa is reported to be 54%, and in Illinois 53% of the total corn–soybean areas in each State.

## III. RESULTS AND DISCUSSIONS

The methodology described above was applied for classification of corn and soybean crops in Iowa and Illinois for a 5-year period (2002–2006). NDVI images at 250-meter resolution for the period DOY 121–DOY 305 were developed from reflectance channels 1 and 2.

*A. MODIS Data Filtering*

A Savitzky–Golay filter was applied to smooth NDVI profiles for each pixel, using a five-point moving window in which NDVI values are approximated by a second-order polynomial. Fig. 3 is an example of how the filtering algorithm performed for a single MODIS pixel.

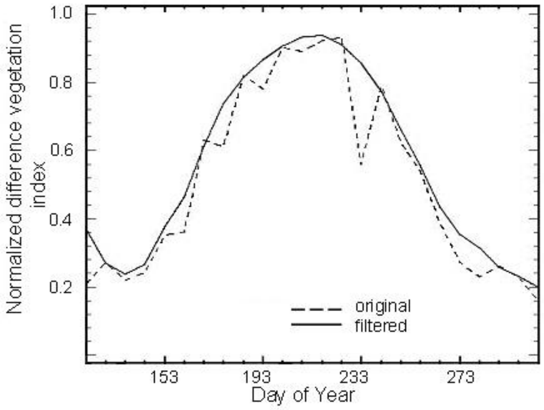


Figure 3. NDVI profile of a sample pixel in its original form and after it is processed with the Savitzky–Golay filter algorithm.

*B. Classification of Crop Area*

The initial step in developing the crop classification was achieved using the decision tree algorithm to separate cropland from other landcover. The thresholds to select crop pixels were chosen using the Landsat classification developed for the 2002 crop season for Iowa and Illinois. Landsat 30-meter classification pixels were aggregated into 250-meter pixels, which contained greater that 80% of crop area (only corn *and* soybean). Then the lower threshold values for MODIS classification were chosen based on the condition that the number of 250-meter aggregated pixels and the number of MODIS pixels selected were approximately the same. The upper threshold was 0.8 for both States. The lower threshold values were 0.35 for Iowa and 0.40 for Illinois. These values were consistent for developing multiyear crop classifications (2003–2006) using MODIS imagery.

In the next step, filtered NDVI images from DOY 137–209 were selected to distinguish corn and soybean crops. Pixel profiles for these days were approximated by equation (1); second derivatives (equation 3) for DOYs 169 and 177 were calculated. Average values from these 2 days are used to distinguish corn and soybean crops. Fig. 4 shows the average corn and soybean NDVI profiles for Iowa as a result of classification.

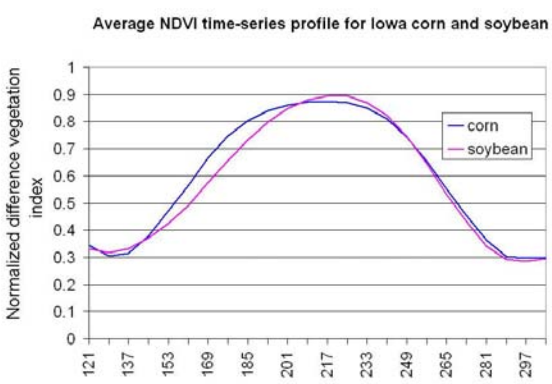


Figure 4. Seasonal dynamics of normalized difference vegetative index for corn and soybean crop

The separation algorithm was tested using the Landsat 2002 year classification for Iowa and Illinois. Landsat 30-meter pixels were aggregated to 250-meter pixels by picking up only those pixels that contained 90% of a specific crop (corn *or* soybean). In Iowa, the number of 250-meter pixels that have 90% of corn or soybean fields was 357,795. The resulting crop classification was compared to the 90% aggregated Landsat pixels. Table 1 shows results of the comparison with an overall producer accuracy of 81.7% and a kappa coefficient of 0.63.

TABLE I. IOWA CLASSIFICATION ACCURACY FOR 2002 CROP SEASON

| | Landsat | | | |
|---|---|---|---|---|
| **MODIS** | ***Corn*** | ***Soybean*** | ***Total*** | ***User's % Accuracy*** |
| ***Corn*** | 160,099 | 28,688 | 188,787 | 84.8 |
| ***Soybean*** | 36,910 | 132,098 | 169,008 | 78.2 |
| ***Total*** | 197,009 | 160,786 | 357,795 | |
| ***Producer's accuracy %*** | 81.3 | 82.2 | | 81.7 |

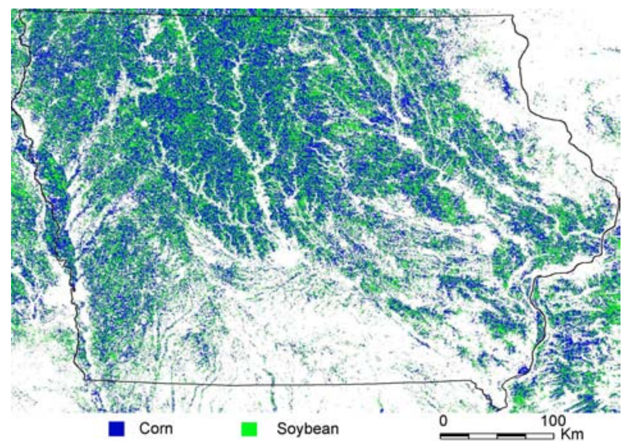


Figure 5. 2002 crop classification map for Iowa

In Illinois the number of 250 m pixels that have 90% corn or soybean fields was 409,108. Comparison of the MODIS classification with the aggregated LANDSAT pixels is shown in table 2. Overall accuracy is 75.1% and a kappa coefficient of 0.50. The results are not as good in Illinois compared to Iowa and may be due to the large north-south variation in corn and soybean crop planting dates and crop development. The 2002 corn and soybean classification maps for Iowa and Illinois are shown in figures 5 and 6.

TABLE II. ILLINOIS CLASSIFICATION ACCURACY FOR 2002 CROP SEASON

| | Landsat | | | |
|---|---|---|---|---|
| **MODIS** | ***Corn*** | ***Soybean*** | ***Total*** | ***User's % Accuracy*** |
| ***Corn*** | 158,654 | 49,037 | 207,691 | 76.4 |
| ***Soybean*** | 52,731 | 148,686 | 201,417 | 73.8 |
| ***Total*** | 211,385 | 197,723 | 409,108 | |
| ***Producer's accuracy %*** | 75.1 | 75.2 | | 75.1 |

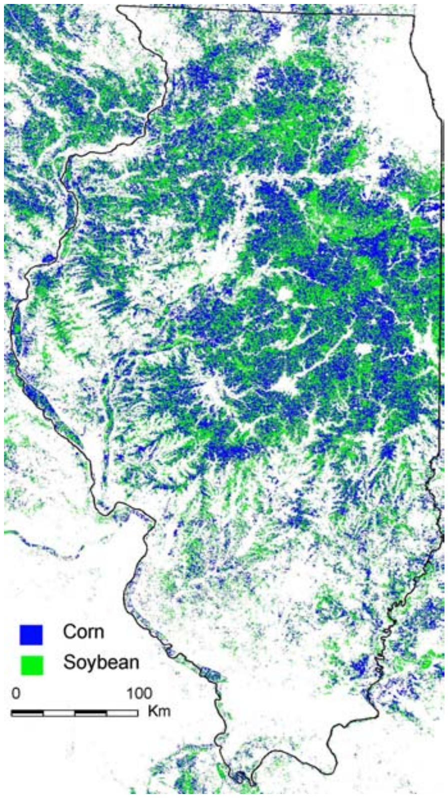


Figure 6. 2002 crop classification map for Illinois

## IV. Conclusion

Operational assessment of crop-specific classification is very important for timely prediction of crop yields. In this study, an algorithm for operational crop classification was developed that uses MODIS 250-meter resolution imagery and applied in the U.S. Corn Belt States of Iowa and Illinois. The algorithm required preprocessing of the MODIS 8-day composite data to remove errors that include discontinuity of data and cloud contamination. The MODIS classification results were compared with results from Landsat classification that was aggregated to the 250-meter resolution. There was a 75–82% user accuracy compared to the Landsat classification that was conducted by the National Agricultural Statistics Service using ground-based field data. The algorithm required no ground-based data, only a percent of corn and soybean planted in the State. This methodology has been successfully applied for operational crop yield prediction for Iowa and Illinois and will be expanded to the rest of the U.S. Corn Belt.

## Acknowledgment

This research was funded by the U.S. Department of Agriculture, Agricultural Research Service, and the National Agricultural Statistics Service.

## References

[1] R.S. DeFries, M.C. Hansen, J.R.G. Townshend, and R.S. Sohlberg, Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers, *International Journal of Remote Sensing*, *19*, 3141–3168, 1998.

[2] M.A. Friedl and C.E. Brodley, Decision tree classification of land cover from remotely sensed data, *Remote Sensing of Environment, 61*, 3, 399–409, 1997.

[3] R. Lawrence, A. Bunn, S. Powell, and M. Zambon, Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis, *Remote Sensing of Environment*, *90*, 331–336, 2004.

[4] R.S. DeFries and J.R.G. Townshend, NDVI-derived land cover classification at a global scale, *International Journal of Remote Sensing*, *15*, 3567–3586, 1994.

[5] P. Jonsson and L. Eklundh, TIMESAT—A program for analyzing time-series of satellite sensor data, *Computer & Geosciences*, *30*, 833– 845, 2004.

[6] L.G. Ferraira, and A.R. Huete, Assessing the seasonal dynamics of the Brazilian Cerrado vegetation through the use of spectral vegetation indices, *International Journal of Remote Sensing*, *2510*, 1837–1860, 2004.

[7] J.F. Knight, R.L. Lunetta, J. Ediriwickrema, and S. Khorram, Regional scale land cover characterization using MODIS NDVI 250 m multi-temporal imagery: a phenology-based approach, *GIScience and Remote Sensing*, *43(1)*, 1–23, 2006.

[8] P. Jonsson and L. Eklundh, TIMESAT—A program for analyzing time-series of satellite sensor data, *Computer & Geosciences*, *30*, 833– 845, 2004.

[9] P.C. Doraiswamy, B. Akhmedov, and A. Stern, Improved techniques for crop classification using MODIS imagery, Presentation at the International Geoscience and Remote Sensing Symposium, July 31–August 4, 2006, Denver, Colorado [CD].

[10] H. Zhang, M.A. Friedl, C.B. Schaaf, A.H. Strahler, J.C.F. Hodges, F. Gao, B.C. Reed, and A. Huete, Monitoring vegetation phenology using MODIS, *Remtoe Sensing of Environment 84*, 471–475, 2003.