Venables, W.N., and B.D. Ripley, 1999. *Modern Applied Statistics with S-PLUS*, Springer-Verlag, New York, 501 p.

Viboud, C., N.B. Ottar, L.S. David, S. Lone, A.M. Mark, and T.G. Bryan, 2006. Synchrony, waves, and spatial hierarchies in the spread of influenza, *Science*, 312(5772):447–451.

Vogt, P., K.H. Riitters, C. Estregull, J. Kozak, T.G. Wade, and J.D. Wickham, 2007. Mapping spatial patterns with morphological image processing, *Landscape Ecology*, 22:171–177.

Wickham, J.D., K. Riitters, T.G. Wade, M. Coan, and C. Homer, 2007. The effect of Appalachian mountaintop mining on interior forest, *Landscape Ecology*, 22:179–187.

Wickham, J.D., K.H. Riitters, R.V. O'Neill, K.B. Jones, and T.G. Wade, 1996. Landscape 'contagion' in raster and vector environments, *International Journal of Geographical Information Systems*, 10(7):891–899.

Wolfram, S., 2002. *A New Kind of Science*, Wolfram Media, Champaign Illinois, 1197 p.

Worboys, M., and M. Duckham, 2004. *GIS: A Computing Perspective*, CRC Press, Boca Raton, Florida, 426 p.

Yang, L., G. Xian, J.M. Klaver, and B. Deal, 2003. Urban land-cover change detection through sub-pixel imperviousness mapping using remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 69(9):1003–1010.

(Received 18 August 2008; accepted 11 December 2008; final version 16 June 2009)

Mapping Cropland and Major Crop Types across the Great Lakes Basin using MODIS-NDVI Data

Yang Shao, Ross S. Lunetta, Jayantha Ediriwickrema, and John Iiames

Abstract

This research evaluated the potential for using the MODIS Normalized Difference Vegetation Index (NDVI) 16-day composite (MOD13Q) 250 m time-series data to develop an annual crop type mapping capability throughout the 480,000 km² Great Lakes Basin (GLB). An ecoregion-stratified approach was developed using a two-step processing approach that included an initial differentiation of cropland versus non-cropland and subsequent identification of individual crop types. Major crop types were mapped for the calendar years of 2002 and 2007. National Agricultural Statistics Service (NASS) census data were used to assess county level accuracies on a unit area basis (2002), and the NASS Crop Data Layer (CDL) was used to generate 231,616 reference data points to support a pixel-wise assessment of the MODIS crop type classification (2007) accuracy across the US portion of the GLB. County level comparisons for 2002 indicated 2.2, −6.8, −6.0, and −5.8 percent of area bias errors for corn, soybeans, wheat, and hay, respectively. Detailed pixel-wise accuracy assessments resulted in an overall crop type classification accuracy of 84 percent (Kappa = 0.73) for 2007. Kappa coefficients ranged from 0.74 to 0.69 for individual ecoregions. The user's accuracies for corn, soybean, wheat, and hay were 87, 82, 81, and 70 percent, respectively. There were spatial variations of classification performances across ecoregions, especially for soybean and hay. Field sizes had a direct impact on the variable classification performances across the GLB.

Introduction

The current movement to increase biofuel crop production could have significant environment consequences for water supply and quality (Carpenter *et al.*, 1998; Osborne and Wiley, 1988), soil fertility and erosion (Groten, 1993; Quarmby *et al.*, 1993), and global carbon balance (Houghton, 1994; Houghton and Hackler, 2000; Lai, 1998; Searchinger *et al.*, 2008). Timely and reliable information about cropland distribution is important to assess the environmental impacts associated with the emergence of cropping practices optimized for biofuel production. Remote sensor data have long been used to characterize cropland at regional and global scales (DeFries *et al.*, 1994; DeFries *et al.*, 1998; Liu *et al.*, 2005; Loveland *et al.*, 2000; Xiao *et al.*, 2003). Past mapping efforts have developed cropland map products at variable spatial resolutions ranging from 1 km AVHRR IGBP DISCOVER dataset (Loveland *et al.*, 2000) to the 30-meter Landsat-derived National Land Cover Datasets (NLCD) products (Homer *et al.*, 2004). The above products provided only a periodic snap-shot of general croplands, and thus provide no documentation about the spatial distributions of biofuel crops (i.e., corn and soybeans).

Recently, the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) generated the cropland data layer (CDL) products. Major crop types were mapped primarily with the AWIFS imagery; however, the CDL products are primarily focused on the agricultural regions in the Midwestern and Mississippi Delta States (**http://www.nass.usda/gov/research/Cropland/SARS1a.htm**). For many other study areas, the annual crop practices and crop rotation information were generally unavailable (Craig, 2001; Wardlow and Egbert, 2008). This has contributed to limiting the soil/water quality assessment efforts that typically require crop map and crop rotation information as inputs (Gassman *et al.*, 2006).

Recent studies have indicated that the Moderate Resolution Imaging Spectroradiometer (MODIS) data has high potential for mapping crops by incorporating both moderate spatial resolution and high temporal resolution attributes. The application of time-series data or phenology-based analysis has been the basis of many cropland mapping applications using MODIS data (Chang *et al.*, 2007; Wardlow *et al.*, 2006; Wardlow *et al.*, 2007; Xiao *et al.*, 2005). Chang *et al.* (2007) examined

Yang Shao is with the National Research Council, U.S. Environmental Protection Agency, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711.

Ross S. Lunetta, and John Iiames are with the U.S. Environmental Protection Agency, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711 (lunetta.ross@epa.gov).

Jayantha Ediriwickrema is with Computer Sciences Corporation, 2803 Slater Road, Morrisville, NC 27560.

Notice: The U.S. Environmental Protection Agency funded and conducted the research described in this paper. Although this work was review by EPA and has been approved for publication, it may not necessarily reflect official Agency policy. Mention of any trade names or commercial products does not constitute endorsement or recommendation for use.

Photogrammetric Engineering & Remote Sensing
Vol. 75, No. 1, January 2010, pp. 73–84.

0099-1112/10/7501–73/$3.00/0

the potential of using MODIS 500 m data for the mapping of corn and soybeans of the United States. Normalized Difference Vegetation Index (NDVI), surface temperature, and surface reflectance were used as inputs to estimate corn and soybean proportions. Using MODIS time-series data, Xiao *et al.* (2005) characterized rice field distributions in southern China. The unique NDVI profiles associated with rice transplanting, growing, and fallow periods were found to be particularly useful in their mapping task. Lobell and Asner (2004) found high inter-class variability and low intra-class separability for MODIS 16-day 250 m NDVI for cropland mapping resulting from sub-pixel heterogeneity. Wardlow *et al.* (2007) analyzed time-series data for individual crop types and concluded that MODIS 250 m NDVI data had sufficient spatial and temporal resolution for major crop type identification in Kansas.

Many researchers have suggested that operational cropland mapping over large study areas represents a significant challenge (Lobell and Asner, 2004; Wardlow *et al.*, 2008; Xiao *et al.*, 2005). To date, most MODIS-NDVI analyses have only focused on single-year mapping effort. A significant impediment has been the lack of quality training data or reference imagery that is not limited in geographic area or temporal resolution. Direct visual interpretation of MODIS-NDVI temporal profiles represent an alternative approach to collect or augment training data.

The analytical approach and selection of a classification algorithm are also important considerations to achieve an optimal result (Duda, 2001). Currently, decision tree is one of the most commonly used in crop mapping research (Chang *et al.*, 2007; Wardlow *et al.*, 2008). Neural network classifiers (NNCS), although intensively used in other remote sensing classifications, have received much less attention for crop mapping (Benediktsson *et al.*, 1990; Goel *et al*, 2003; Huang *et al.*, 2002; Liu *et al.*, 2005). Analytical approaches such as image stratification (Homer *et al.*, 1997; Wardlow *et al.*, 2008) and incorporating ancillary data or prior probabilities (Foody 1995; McIver and Friedl, 2002; Strahler *et al.*, 1980) were also found to be useful in complex classification tasks. Few studies, however, have incorporated agricultural census data or other crop statistics to improve crop mapping results. The validation of large area mapping products also presents a difficult challenge. Past large area mapping efforts did not include accuracy assessment; while others have only provided vague error statistics (Foody, 2002; Justice *et al.*, 2000). The nature of classification errors are still poorly understood especially for sub-regional differences and the interactions between crop field sizes. Accordingly, MODIS-NDVI sub-regional scale classification performances are a research issue of importance to the general scientific community.

Research Objectives

The objective of this research was to evaluate the potential of using the MODIS-NDVI 16-day composite 250 m product (MOD13Q) time-series data to develop an annual cropland mapping capability for the entire 480,000 km^2 Great Lakes Basin (GLB). The cropland mapping design incorporated a two-stage classification protocol to first (a) differentiate cropland versus non-cropland, and then (b) identify individual crop types. The classification of cropland versus non-cropland provided a baseline dataset to support both the more detailed individual crop type identification and to provide a cropland mask for non-agricultural land-cover (LC) classification and change detection efforts. Major GLB crop types considered in this study were corn, hay, soybeans, and wheat. We implemented and assessed crop mapping results for years 2002 and 2007.

The methods developed in this study are potentially important to researchers attempting to characterize the annual change of crop distributions and to monitor crop rotation patterns across large geographic regions. For calendar year 2002, we were interested in the utility of the NASS county level agricultural census data (NASS, 2004) to provide priors to improve the individual crop mapping results and to support non-site specific (unit area) classification accuracy assessments. For calendar year 2007, the availability of CDLs for reference data development supported a pixel-wise (per-pixel) assessment of classification performances. We were also interested in examining the spatial variations (ecosystem) of classification performances across the international GLB study area and investigate the impacts of crop field sizes on the classification performance.

Study Area

The GLB is one of the most heavily industrialized regions in North America, but also a region that supports a variety of unique ecosystem services which include numerous intensively managed land-use activities. Agricultural production in the GLB represents 7 percent of US and 25 percent of Canadian agricultural production (USEPA, 2008). Through population growth and redistribution, urban expansion, and loss of natural areas to development, numerous human induced alterations have occurred resulting in substantial land degradation, deterioration of air and water quality, degradation of watershed habitats, and trajectories of LC change that have substantially altered biodiversity and ecological services (Crosbie *et al.*, 1999; Wolter *et al.*, 2006). Although the US and Canada share stewardship of the GLB, no single or common mapping effort to monitor annual crop distributions was currently available or ongoing.

Methods

Numerous data sets were first assembled including the MODIS 16-Day composite NDVI data (MOD13Q1) from 2000 to 2007 acquired from the USGS EROS Data Center. The 16-day composite MODIS-NDVI data was re-projected from a sinusoidal into an Alber's Equal Area Conic projection. The MODIS-NDVI data were then preprocessed using the method described by Lunetta *et al.* (2006) to provide a high quality uninterrupted data stream to support multi-temporal (phenology) analysis. Non-high quality NDVI data values (i.e., clouds, shadows, etc.) identified by the NASA quality control (QC) flags were removed. A discrete Fourier transformation was then used to decompose the NDVI time-series into low-frequency components and high-frequency components of noise. The NDVI values removed in the filtering process were then transformed into the frequency domain using a discrete Fourier transformation and the signal and noise spectrum separated. Data points were estimated from the frequency domain signal spectrum using a nonlinear deconvolution approach (Roberts *et al.*, 1987). This technique generated a "clean" NDVI temporal profile used to support phenology-based image classifications (Lunetta *et al.*, 2006).

All available Landsat TM and ETM images from 2000 to 2002 were acquired for the GLB from the EROS Data Center's Global Land Cover Facility. Two seamless image mosaics were created from the Landsat imagery including a panchromatic image (15 m) and a false color IR composite (CIR) image using bands 4, 3, and 2 (30 m). These mosaics served as the reference data for identifying training and test sites for the analysis of MODIS-NDVI data. The mosaics were particularly usefully for cropland versus non-cropland classification. The crop data layer (CDL 2007) was obtained from the U.S. Department of Agriculture (USDA) National Agriculture Statistics Service (NASS). The CDL 2007 was derived mainly from AWIFS imagery and provided partial GLB coverage (i.e., Illinois, Indiana, Michigan, Minnesota, Ohio and Wisconsin). The CDL 2007 served as the reference data for the validation of 2007 MODIS-NDVI classification. The overall accuracy of the CDL 2007 was determined by NASS to be >90 percent and the accuracies for the three major crop types (e.g., corn, soybeans, and wheat) >92 percent (NASS, 2008). Additionally, the NASS county level agricultural census data (2002) were used to support the 2002 MODIS-NDVI classification and validation. The NASS 2002 agricultural census represented the most recent agricultural statistics dataset corresponding to the US portion of the GLB. Validation efforts were limited to the US portion of the GLB, because no reference data sets were available for Canada.

Cropland versus Non-Cropland

We applied a stratification of the study region to perform the cropland versus non-cropland classification. The GLB was first stratified into 12 ecoregions (Plate 1). A total of 10 ecoregions were located in the US portion of the GLB and the remaining two in Canada (Omernik, 1987). Training samples were selected independently for each ecoregion corresponding to a minimum of 1,000 cropland and 1,000 non-cropland pixels for each ecoregion. The Landsat seamless mosaic images were used as the primary reference data for the selection of these training data. To account for the coarser resolution MODIS data and minimize the effects of registration and edge effects, training pixels were selected in the center of large homogeneous areas.

We applied a NNC approach for the cropland mapping. The software used for neural network training and classification was the UNIX-based SNNS software package (SNNS, 1993). A multi-layer perceptron (MLP) NNC was designed for this study that consisted of three connected layers including the input, hidden, and output layers. A total of 13 input nodes were used at the input layer, corresponding to 13 NDVI dates (Julian day 81 to 273) from the 2002 MODIS-NDVI MOD13Q data product. The hidden layer used 10 nodes and only one output node was used at the output layer. The network was trained with a back-propagation algorithm and a sigmoid activation function. A three-fold stratified cross-validation was used to improve the overall generalization ability (Bishop, 2006; Duda, 2001). The network was trained to produce output values (codes) corresponding to cropland (1.0) and non-cropland pixels (0.0). An output value near 1.0 suggested a high probability of being cropland. We applied thresholding (>0.5) of the network output to label each pixel. The MODIS-NDVI data were scaled to values between 0.0 to 1.0 prior to their input into the MLP-NNC.

For each ecoregion, the resultant cropland map was overlaid onto the Landsat panchromatic mosaic and the MODIS-NDVI images. An overall visual interpretation was first performed to provide a qualitative assessment of the correspondence between the derived cropland and the reference datasets. For ecoregions determined to have a poor level of correspondence, additional training data points were added to improve the classification performance. A minimal level of post-classification manual editing was also conducted to remove obvious classification errors to produce a final cropland mask.

Crop Type Classification (2002)

The cropland mask was used as the baseline dataset for the subsequent identification of individual crop types. All the non-cropland pixels were eliminated from individual crop type identification processing. The training sets for the individual crop types were developed primarily based on the visual interpretation of NDVI temporal profiles of pixels within the cropland mask. Four main crop types (i.e., corn, soybeans, wheat and hay) represented the majority of cropland (>80 percent) across the US portion of the GLB

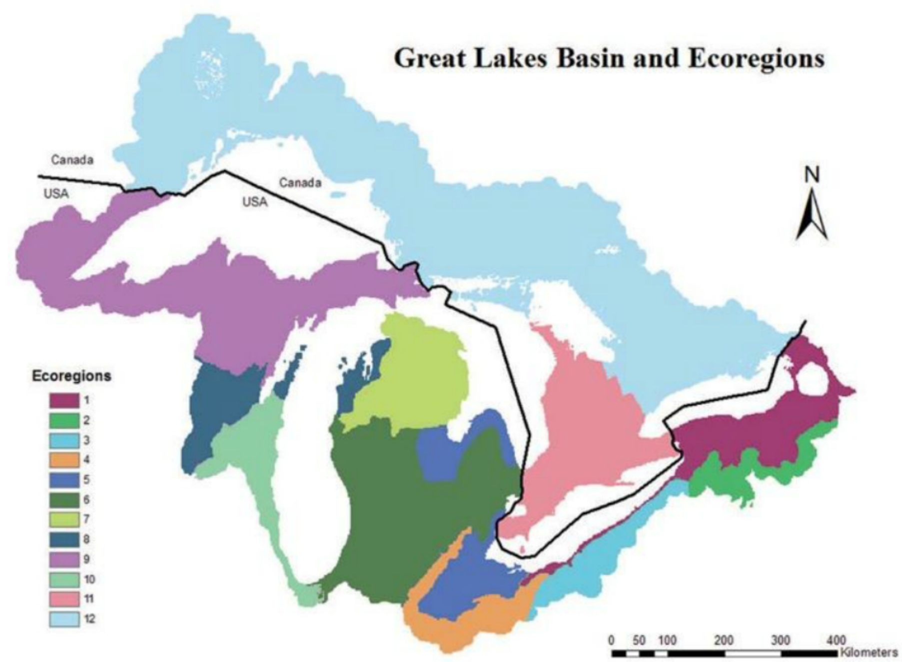

Plate 1. Ecological regions within the United States portion of the Great Lakes Basin: (n = 12). Ecoregion boundaries were used to stratify the MODIS-NDVI classifications (Omernik, 1987).

(NASS, 2004). Importantly, these crop types often have unique NDVI temporal signals that can readily be visual interpretation (Wardlow *et al.*, 2008). The selection of training data for individual crop identification was also ecoregion-based. Initially, a minimum of 250 training pixels for each crop type were selected per each ecoregion. Individual crop type classifications were performed using NNC. As indicated above, a three-layer NNC was designed, 13 and 10 nodes were used at input layer and hidden layer, respectively. Four nodes were used at the output layer, corresponding to the individual crop types of corn, soybeans, wheat, and hay. The target values were set as the 1-of-M target coding system (i.e., 1,0,0,0 = corn; 0,1,0,0 = soybeans; 0,0,1,0 = wheat; 0,0,0,1 = hay).

We used a subset of the 2002 agricultural census statistics to derive class prior probabilities and improve the classification results. Specifically, 19 percent of counties within each ecoregion were randomly selected, totaling 29 counties across the entire GLB. *A priori* class probability, $P(c_i)$, was calculated for each of the four crop types using the county level agricultural census statistics; which represented the percentage of crop area belonging to class c_i. For all input cropland pixels within the selected counties, the neural network outputs were averaged at each of the four output nodes to derive $Ave(Y_i)$. These values should be close to prior class probability $P(c_i)$, if neural network classifier approximates the probabilities of class membership for each class (Bishop, 1995; Richard and Lippmann, 1991). We compared $Ave(Y_i)$ to prior class probabilities using a relative entropy distance measure suggested by Richard and Lippmann (1991):

$$D = \sum_{i=1}^{M} P(c_i) \log \frac{P(c_i)}{Ave(Y_i)}. \quad (1)$$

A large entropy distance typically suggests inaccurate or poor classification performance. The addition of training data points and modification of neural network training protocols were repeated until a minimum entropy distance measure was obtained. The trained neural networks were then used to classify all the cropland pixels in the GLB. We used thresholding of network outputs to label the pixels. Rejection criteria were applied to pixels with low output values (<0.5) at all output nodes.

The accuracy assessment of 2002 MODIS-NDVI cropland classification was first conducted at the county aggregated level. The areas of total cropland versus non-cropland were aggregated for each county completely located inside the GLB boundary (n = 150). Included were counties from Wisconsin (20), Indiana (7), Michigan (83), New York (16), and Ohio (24). Next, cross-plots were also generated for specific crop types (corn, soybeans, wheat, and hay) to evaluate the differences between MODIS-NDVI-derived results and the 2002 USDA census statistics. Out of the 150 counties, 29 were used to support the individual crop type training of NDVI profiles, and 121 were used for the accuracy assessment. Two quantitative measures, the root-mean-square error (RMSE) and systematic error (SE), were calculated to describe average differences. RMSE and SE were calculated using the following equations:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(E_i - A_i)^2}{n}} \quad (2)$$

$$SE = \frac{\sum_{i=1}^{n}(E_i - A_i)}{n} \quad (3)$$

where E_i is the estimated crop area from MODIS-NDVI data, A_i is the crop area from the NASS census statistics, and n is the total number of counties used. The comparison was conducted only for the US portion of GLB (no companion agricultural statistics data were available for Canada).

Crop Type Classification (2007)

The training data for 2007 MODIS-NDVI classification were collected using visual interpretation of MODIS-NDVI temporal profiles. For each ecoregion, a minimum of 250 training pixels were selected for each of the individual crop types. Again, a three-layer NNC was used to classify MODIS-NDVI pixels. The CDL 2007 was primarily used for the accuracy assessments of MODIS-NDVI classification. A large reference dataset was developed from the CDL 2007 to support the accuracy assessment. First, all MODIS-NDVI pixels were identified that corresponded to a 350 × 350 m homogeneous crop patches in CDL 2007. These labeled MODIS-NDVI pixels were further stratified by ecoregions to support accuracy assessments at both GLB and individual ecoregion scales. Error matrices and kappa coefficients were generated for the accuracy assessments (Congalton and Green, 1999). It should be noted that the CDL 2007 covered only a portion (ecoregions 4 to 10) of the GLB study area (Plate 1).

The CDL 2007 data was also used to assess impacts attributable to crop field sizes on MODIS-NDVI classification performance. Thematic layers corresponding to individual crops were first extracted from the CDL 2007 to represent each of the four major crop types. Next, a majority filter with a 3 × 3 window was used to remove inclusions (salt and pepper). For each of the four individual crop types, we then identified all homogenous patches using an eight-cell contiguity rule (McGarigal and Marks, 1995). The size of each crop patch was calculated to provide an average crop patch sizes for individual ecoregions. The mean crop patch sizes were further examined in the context of MODIS-NDVI spatial resolution and the regional differences of crop mapping performances were assessed.

Results

Cropland versus Non-Cropland

The results for cropland and non-cropland classification were compared to the NASS census statistics (Figure 1). The mean NASS census cropland area for GLB counties was 410.4 km^2 compared to 420.6 km^2 for the MODIS-NDVI classification in 2002. In general, the MODIS-NDVI classification over-estimated cropland area (SE = 10.2 km^2) by approximately 2.5 percent across the US portion of the GLB. This may be attributed to the over-sampling of cropland pixels in the training data, because classification results from non-parametric classifiers such as neural networks are sensitive to the class frequency distributions of training data (Foody *et al.*, 1995; McIver and Friedl, 2002). The cross-plot showed relatively larger scattering for counties with less than 200 km^2 cropland acreages. The RMSE value was 65.5 km^2. For 84 of 150 counties, the estimated cropland acreages were within the 15 percent error bounds of the NASS census data.

SE and RMSE values were calculated and compared for the 10 US GLB ecoregions (Table 1). The mean of cropland acreage in each ecoregion was also reported for the NASS census data and the MODIS-NDVI estimates, respectively. For NASS census, the means of cropland acreages ranged from 61.0 km^2 (ecoregion 9) to 809.9 km^2 (ecoregion 5). The MODIS-NDVI classification slightly overestimated cropland for almost all ecoregions except ecoregion 4 (SE = −18.6 km^2). RMSE values ranged from 37.3 to 98.8 km^2 across ecoregions. Ecoregion 8 had the highest RMSE value (98.8 km^2). The high

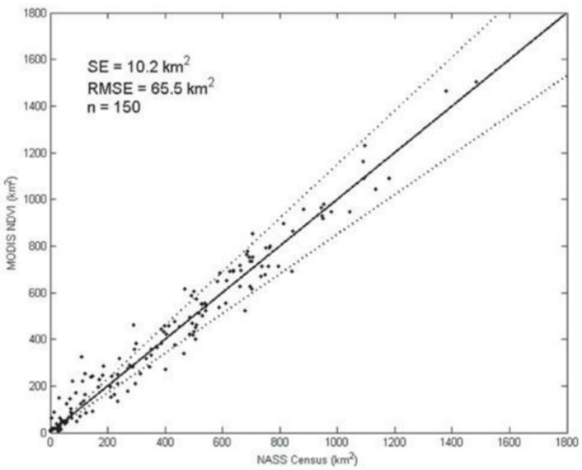

Figure 1. Comparisons of total cropland estimates using end-members corresponding to individual ecoregions (n = 12) for MODIS-NDVI classifications versus NASS county level census data (2002).

RMSE was attributed to one county that had very large difference that inflated the MODIS estimated 2002 cropland acreage value. Ecoregions 4 and 5 had the intermediate levels of RMSE, but they were relatively small compared to the means of cropland acreages. Initial visual interpretation also indicated that there was reasonable agreement between MODIS-NDVI derived cropland map and Landsat seamless mosaic. Visual interpretation of the MODIS-NDVI classification result also suggested that ecoregion 9 was the most problematic region. The SE (10.3 km^2) and RMSE (37.3 km^2) values were relatively high compared to its mean cropland acreage; some forest and wetland pixels were falsely labeled as cropland pixels. One reason is that ecoregion 9 is located further north and includes Michigan's Upper Peninsula. Confusion of croplands and forests often occurred at high latitudes for phenology-based NDVI classification (Friedl *et al.* 2000; Loveland *et al.*, 1999). Degraded MODIS-NDVI quality might be another explanation because NDVI data quality tends to decrease at high latitudes.

Some individual counties initially had large differences (i.e., >150 km^2) of cropland estimations compared to the NASS census statistics. For example, in both Montcalm County (Michigan) and Pulnam County (Ohio), the MODIS-NDVI classification greatly underestimated the total cropland. We determined that there were some croplands with unique NDVI temporal profiles, and the training signals selected for the ecoregion did not adequately representative for these counties. Therefore, we added these missing training signals for neural network training and reprocessed the counties. Also, the MODIS-NDVI classification largely overestimated cropland for several counties with substantial suburban residential areas. Low density residential areas often contained mixed vegetation and impervious cover types. Depending on the level of spatial mixing and the vegetation phenology attributes, pixels in low density residential areas often mimicked the phonological cycle of cropland areas, resulting in classification errors. Minimal post-classification manual editing was conducted to remove the obvious classification errors. For the majority of counties, a visual inspection of the cropland map with the Landsat seamless mosaic did not reveal any obvious errors associated with the MODIS-NDVI classification (i.e., other cover types being classified as cropland). The observed errors were largely attributable to the 250 m resolution limitation of MODIS-NDVI data. Because mixed pixels were labeled as a single cover type, the total cropland estimations may differ substantially compared to the actual cropland acreages (Lobell and Asner, 2004).

TABLE 1. MODIS-NDVI CLASSIFICATION RESULTS FOR CROPLAND CORRESPONDING TO INDIVIDUAL ECOREGIONS. CROPLAND ESTIMATIONS DERIVED FROM MODIS-NDVI CLASSIFICATION (2002) WERE COMPARED WITH THE NASS CENSUS DATA (2002). MEAN VALUES WERE CALCULATED BY DIVIDING THE TOTAL AREA BY THE NUMBER OF COUNTIES PER ECOREGION. ECOREGIONS 1, 2, AND 3 WERE COMBINED TO PROVIDE A SUFFICIENT NUMBER OF COUNTIES TO SUPPORT A COMPARATIVE ANALYSIS. UNITS ARE KM^2 FOR ALL STATISTICS

Ecoregion ID	Mean_NASS	Mean_MODIS	SE	RMSE
Ecoregion 1,2,&3	329.9	333.4	3.5	68.3
Ecoregion 4	781.6	763.0	−18.6	73.6
Ecoregion 5	809.9	839.7	29.7	72.3
Ecoregion 6	488.1	500.4	12.3	61.5
Ecoregion 7	90.8	97.6	6.8	54.5
Ecoregion 8	336.3	368.9	32.6	98.8
Ecoregion 9	61.0	71.3	10.3	37.3
Ecoregion 10	510.1	512.5	2.4	71.5
GLB Overall	410.4	420.6	10.2	65.5

Mean_NASS (NASS census data)
Mean_MODIS (MODIS-NDVI classification)
SE (systematic error)
RMSE (root-mean-square-error)

Crop Type Classifications (2002)

Individual crop types typically have unique NDVI temporal profiles. This allowed the analyst to collect training pixels through visual interpretation of the MODIS-NDVI time-series data. Figure 2 shows the 2002 NDVI temporal profiles of four major crop types for selected training samples from ecoregion 6, which corresponded to areas in southern Michigan. Winter wheat had higher NDVI values in spring (Julian day 129) compared to the other three major crop types, and then

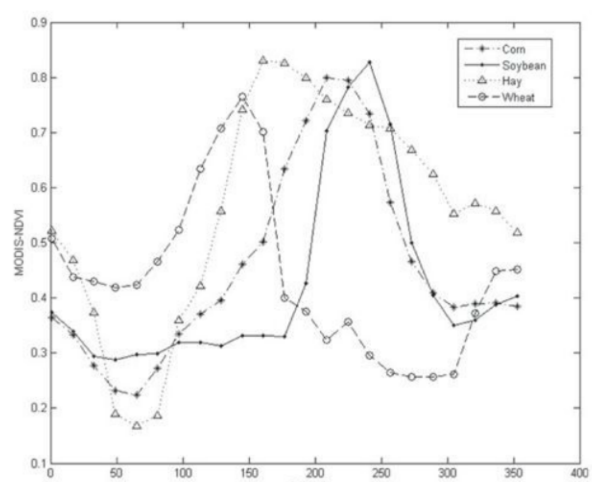

Figure 2. MODIS-NDVI temporal profiles (end-members) representing the major Michigan crop types. Profiles represent average values for corn, soybeans, hay, and wheat.

values quickly drop to the lowest among four crop types around Julian day 209. Michigan corn began growing in early-May and peaked in late-July (Julian day 209). The planting of soybeans was about two weeks later than corn, resulting in large NDVI differences in the range of Julian day 177 to 209. A majority of Michigan hay was alfalfa and similar to winter wheat; NDVI values peaked in early-June (Julian day 161). Depending on the timing of cutting, there might be local peaks in the temporal profiles (Wardlow *et al.*, 2007).

An initial crop classification was conducted by using 250 training samples for each class. The resulting crop type map was aggregated to county level and compared to the NASS 2002 census statistics. For corn, soybeans, wheat, and hay, the SE values were 10.1, −18.3, 9.8, and 15.1 km^2 and RMSE 65.4, 58.5, 34.5, 39.6 km^2, respectively. The large degree of variability between the MODIS-NDVI estimated crop acreage and the NASS 2002 census statistics for all four crop types prompted additional NDVI signature refinement. For most ecoregions, additional training pixels were added iteratively to improve classification performance by comparing the MODIS-NDVI-derived crop acreages and the estimated priors for a selected subset of counties (i.e., 29 of 150).

Final crop area estimations from 2002 MODIS-NDVI classifications were then compared to the 2002 county level agricultural census statistics (Figure 3a through 3d). The RMSE values (n = 121) were 41.6, 45.6, 19.5, and 22.8 km^2 for corn, soybeans, wheat, and hay, respectively. These values were much smaller than the initial RMSE values derived without incorporating prior information. By incorporating priors, the MODIS-NDVI classification also obtained smaller systematic errors for corn (SE = 3.0 km^2), soybean (SE = −9.5 km^2), wheat (SE = −1.8 km^2), and hay (SE = −3.2 km^2). These values corresponded to 2.2, −6.8, −6.0, and −5.8 percent of bias compared to the means of crop acreages from NASS 2002 census data. These results suggested good generalization ability, because only 19 percent of counties were utilized for prior estimation. Similar studies have been conducted by McIver and Friedl (2002) to improve their agricultural land mapping effort, although there were differences in the selection/use of classification algorithm and the sources of prior information. Without in situ training data or other reference data, the agricultural statistics are probably the only available dataset that can be used as quality control for individual crop mapping.

From the Figure 3 scatter plots, it appears that there are points with large differences between the MODIS-NDVI classification and the NASS 2002 census statistics. For example, the MODIS-NDVI classification greatly underestimated wheat acreages for Sanilac and Huron Counties (Michigan). This may be attributable to error propagations from the initial cropland and non-cropland classification (Some wheat fields may have been misclassified as non-croplands in the initial cropland classification). However, the large overestimation of wheat acreages in Paulding and Wood counties (Ohio) is difficult to explain. Visual interpretation of NDVI temporal profiles did not reveal major commission errors from other crop types.

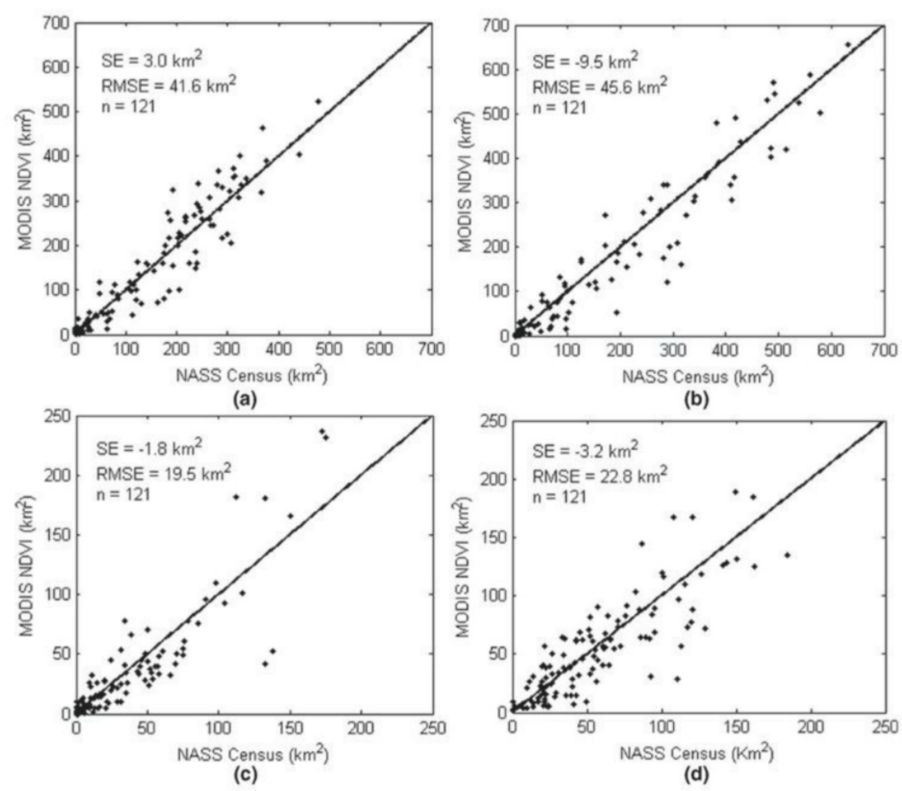

Figure 3. Comparisons of individual crop type estimates from the MODIS-NDVI 2002 classifications versus the NASS 2002 census data for (a) corn, (b) soybeans, (c) wheat, and (d) hay.

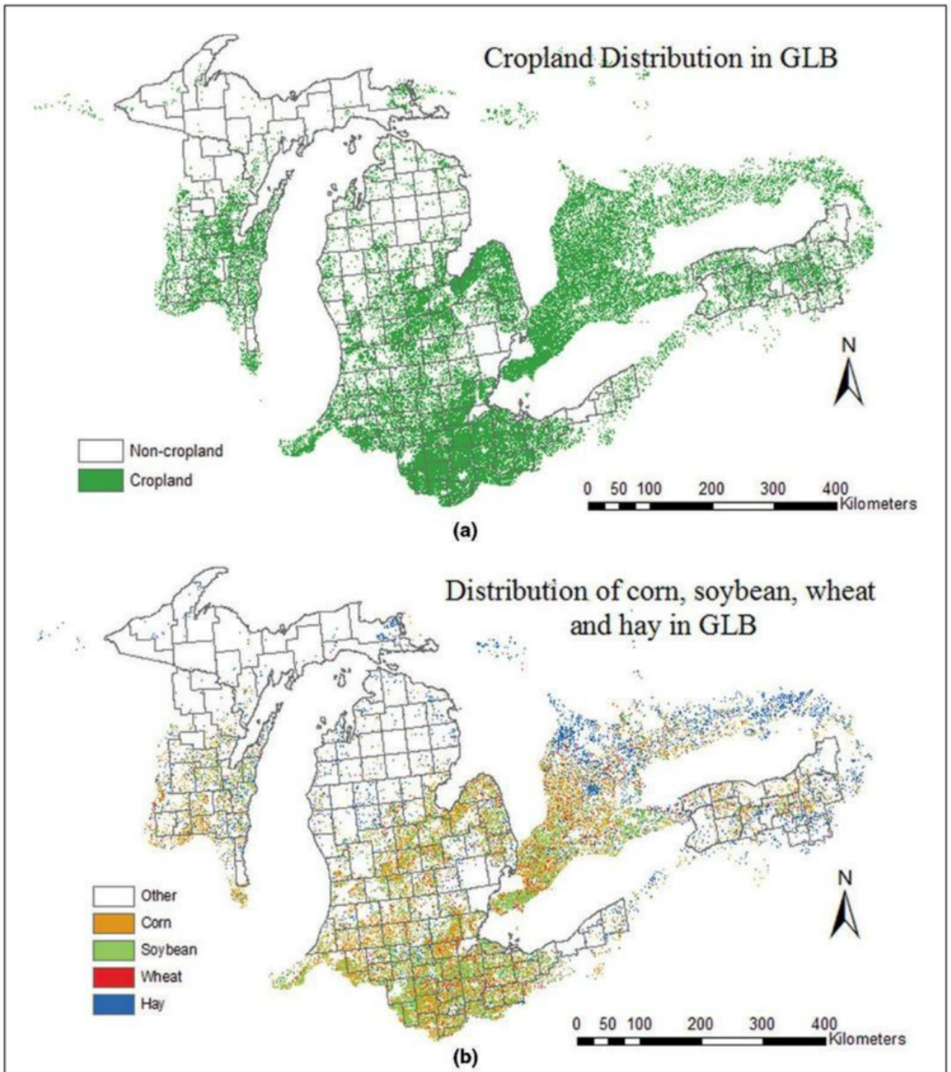

Plate 2. MODIS-NDVI cropland classification results for 2002: (a) cropland distribution, and (b) the distribution of the four major crop types (corn, soybeans, wheat, and hay) across the entire Great Lakes Basin.

Plate 2a and 2b shows the general cropland extent and the spatial distributions of the four major crop types in the study region. Corn and soybean were dominant crop types in ecoregions 4, 5, and 6. The two crop types contribute over 70 percent of the total crop acreage for most counties in these ecoregions. In ecoregions 7 and 9 (Michigan's Upper Peninsula), hay was the dominant crop type (>60 percent) for most counties. A majority of wheat fields were located in ecoregions 4 and 5. For ecoregions 8 and 10, corn and hay were the dominant crop types. Table 2 shows the comparisons of SE and RMSE values derived for eight individual ecoregions (combinations). The mean of cropland acreage in each ecoregion was also reported for both the NASS census statistics and the MODIS-NDVI estimates, respectively. Mean values were calculated to allow a meaningful comparison between ecoregions by dividing the total crop area by the number of counties. There were large variations of crop distributions for all four crop types across ecoregions. For example, the means of soybean acreages ranged from 1.8 km^2 for ecoregion 9 to 402.4 km^2 for ecoregion 4. For all four crop types, the estimated means of crop acreages were close to the mean values derived from 2002 agricultural census. Across ecoregions, SE values were only slightly higher or lower than 0 for all four individual crop types, suggesting unbiased estimation from MODIS-NDVI classification. The RMSE values had large variations across ecoregions, but most were proportional to the means of crop acreages.

Crop Type Classifications (2007)

A pixel-wise accuracy assessment of the 2007 MODIS-NDVI classification was performed using CDL 2007 as reference. For the entire GLB study area, the overall accuracy was 84 percent (n = 231,616) with a Kappa coefficient of 0.73 (Table 3). The accuracies of corn and soybeans were 87 and 82 percent, respectively. Both soybeans and corn are summer crops with similar NDVI temporal profiles. The

TABLE 2. MODIS-NDVI CLASSIFICATION RESULTS FOR CROP TYPES CORRESPONDING TO INDIVIDUAL ECOREGIONS. CROP TYPE ESTIMATIONS DERIVED FROM MODIS-NDVI CLASSIFICATION (2002) WERE COMPARED WITH THE NASS CENSUS DATA (2002). MEAN VALUES WERE CALCULATED BY DIVIDING THE TOTAL AREA BY COUNTIES PER ECOREGION. UNITS ARE KM^2 FOR ALL STATISTICS.

Ecoregion ID	Mean_NASS	Mean_MODIS	SE	RMSE
Corn				
Ecoregion 1,2,&3	102.0	112.2	10.2	38.2
Ecoregion 4	263.8	265.3	1.5	50.9
Ecoregion 5	252.7	275.2	22.5	54.0
Ecoregion 6	180.3	175.1	−5.2	50.6
Ecoregion 7	16.4	18.0	1.5	13.5
Ecoregion 8	127.0	141.4	14.3	23.1
Ecoregion 9	10.0	11.7	1.7	5.4
Ecoregion 10	197.3	186.4	−10.9	51.6
Soybean				
Ecoregion 1,2,&3	47.9	34.8	−13.1	33.4
Ecoregion 4	402.4	388.3	−14.1	79.0
Ecoregion 5	389.4	372.3	−17.1	51.5
Ecoregion 6	172.2	162.6	−9.6	53.7
Ecoregion 7	6.6	3.4	−3.3	8.8
Ecoregion 8	42.8	43.7	0.8	21.9
Ecoregion 9	1.8	0.7	−1.2	5.4
Ecoregion 10	86.5	71.7	−14.9	59.8
Wheat				
Ecoregion 1,2,&3	24.1	24.6	0.5	14.0
Ecoregion 4	80.4	76.2	−4.2	25.8
Ecoregion 5	84.4	78.3	−6.1	43.0
Ecoregion 6	27.1	25.5	−1.6	12.3
Ecoregion 7	3.6	4.2	0.6	5.1
Ecoregion 8	11.3	7.8	−3.5	8.8
Ecoregion 9	0.6	0.3	−0.3	1.5
Ecoregion 10	31.4	30.7	−0.7	10.4
Hay				
Ecoregion 1,2,&3	86.1	79.8	−6.3	30.7
Ecoregion 4	25.3	26.7	1.4	18.6
Ecoregion 5	33.9	27.8	−6.1	16.3
Ecoregion 6	61.7	54.4	−7.3	24.5
Ecoregion 7	42.1	46.8	4.7	19.6
Ecoregion 8	75.4	70.3	−5.0	27.4
Ecoregion 9	34.8	30.8	−4.0	16.5
Ecoregion 10	87.7	101.0	13.3	24.7

Mean_NASS (NASS census data)
Mean_MODIS (MODIS-NDVI classification)
SE (systematic error)
RMSE (root-mean-square-error)

moderate levels of commission/omission errors indicated the level of confusions between these two classes. This is consistent with results from previous research of corn and soybean mapping (Chang *et al.*, 2007). Wheat had a classification accuracy of 81 percent with commission errors mostly attributed to both corn and soybeans. The accuracy of hay classification was the lowest at 70 percent and the large commission error was largely associated with corn. The low accuracy of hay identification was possibly related to variable harvest times. Farmers typically cut hay several times a year. This caused large variations of NDVI temporal profiles and the negatively impacted classification performance.

The accuracy assessments were also conducted for individual ecoregions (Table 4). There was no reference data available for ecoregions 1 and 2. Ecoregions 3 and 9 were also dropped due to insufficient sample size for statistical analysis. Across ecoregions, the overall accuracies ranged from 77 percent (ecoregion 7) to 86 percent (ecoregion 8). Kappa coefficients were >0.70 for all ecoregions except No.7. The classification of corn performed best as accuracies were above 80 percent for all 12 ecoregions. The classification performances of wheat were also consistent spatially. Accuracies approached 80 percent for all regions except ecoregion 7 (70 percent). For soybean and hay, however, accuracies were highly varied. Ecoregions 4, 5, and 6 had good soybean classification performances with above 75 percent accuracy, while ecoregions 7, 8, and 10 had high commission errors of 49, 44, and 33 percent, respectively. Large numbers of corn pixels were falsely classified as soybean. However, hay classification accuracies associated with ecoregions 7, 8, and 10 were much higher than the numbers at ecoregions 4, 5, and 6.

Using CDL 2007 data, we further examined the sizes of crop patches and their impacts on the classification performances. Figure 4 shows the mean crop patch sizes of four crop types for selected ecoregions. In general, large crop patch size suggested less sub-pixel mixing and a more "pure" crop phenology. Larger patch sizes also tend to mitigate MODIS image registration problems with the CDL 2007 reference data. Higher classification accuracies were expected for crop types with large homogeneous patches. Across the US portion of the GLB, corn had relatively large mean patch sizes (>1.5 MODIS pixel) and consistently high classification accuracies across regions. For soybean, ecoregions 4, 5, and 6 had relatively large mean patch sizes (>2.0 MODIS pixels), while ecoregions 7, 8, and 10 had mean patch sizes of less than 1.5 MODIS pixels. This explained the differences of classification accuracies between ecoregions 4, 5, 6 and ecoregions 7, 8, and 10. The mean crop patch sizes of hay were less than 1.0 MODIS pixel in ecoregions 4, 5, and 6. Large classification errors were expected as most hay pixels were not "pure" pixels. The impacts of crop patch sizes on wheat classification were less interpretable. The accuracies of wheat classification were consistently high across ecoregions while the mean crop patch sizes were relatively small (<1.5 MODIS pixels) in ecoregions 6, 7, 8, and 10. It should be noted that there are many other factors, such as the shapes of crop patches and inaccuracies

TABLE 3. NORMALIZED ACCURACY ASSESSMENT STATISTICS FOR THE GLB USING THE NASS CDL (2007) AS REFERENCE DATA

Reference							
	Corn %	Soybean %	Wheat %	Hay %	Total %	% Correct	% Commission
Corn	**45.48**	5.99	0.34	0.24	52.05	87	13
Soybean	6.22	**29.59**	0.28	0.05	36.15	82	18
Wheat	0.64	0.44	**5.02**	0.11	6.21	81	19
Hay	0.99	0.31	0.39	**3.89**	5.59	70	30
Total %	53.33	36.33	6.04	4.29	100.00	**84**	**(n = 231,616)**
% Correct	85	81	83	91			
% Omission	15	19	17	9			**Kappa = 0.73**

TABLE 4. ERROR MATRICES USING NASS 2007 DATA AS REFERENCE. NOTE THAT ECOREGIONS 1, 2, 3, 9, 11, AND 12 WERE NOT INCLUDED. NASS 2007 DATA WAS NOT AVAILABLE (1 AND 2) OR INSUFFICIENT (3 AND 9) AND NO DATA (11 AND 12).

Reference							
	Corn	Soybean	Wheat	Hay	Total	% Correct	% Commission
Ecoregion 4							
Corn	18403	3971	81	4	22459	82	18
Soybean	2857	20056	99	11	23023	87	13
Wheat	245	323	2170	10	2748	79	21
Hay	228	144	144	181	697	26	74
Total	21733	24494	2494	206		83	(n = 48,927)
% Correct	85	82	87	88			
% Omission	15	18	13	12			Kappa = 0.70
Ecoregion 5							
Corn	27260	4237	395	48	31940	85	15
Soybean	4724	27079	427	29	32259	84	16
Wheat	602	462	4533	26	5623	81	19
Hay	320	274	249	414	1257	33	67
Total	32906	32052	5604	517		83	(n = 71,079)
% Correct	83	84	81	80			
% Omission	17	16	19	20			Kappa = 0.71
Ecoregion 6							
Corn	42816	3833	122	63	46834	91	9
Soybean	6086	20146	113	38	26383	76	24
Wheat	396	188	2951	43	3578	82	18
Hay	658	208	139	1435	2440	59	41
Total	49956	24375	3325	1579		85	(n = 79,235)
% Correct	86	83	89	91			
% Omission	14	17	11	9			Kappa = 0.71
Ecoregion 7							
Corn	771	72	20	53	916	84	16
Soybean	201	260	10	41	512	51	49
Wheat	40	14	258	58	370	70	30
Hay	69	11	36	750	866	87	13
Total	1081	357	324	902		77	(n = 2,664)
% Correct	71	73	80	83			
% Omission	29	27	20	17			Kappa = 0.67
Ecoregion 8							
Corn	5710	491	48	204	6453	88	12
Soybean	184	246	2	6	438	56	44
Wheat	70	8	502	59	639	79	21
Hay	403	23	91	2926	3443	85	15
Total	6367	768	643	3195		86	(n = 10,973)
% Correct	90	32	78	92			
% Omission	10	68	22	8			Kappa = 0.74
Ecoregion 10							
Corn	10374	1276	132	175	11957	87	13
Soybean	359	747	9	2	1117	67	33
Wheat	130	17	1219	51	1417	86	14
Hay	623	61	250	3313	4247	78	22
Total	11486	2101	1610	3541		84	(n = 18,738)
% Correct	90	36	76	94			
% Omission	10	64	24	6			Kappa = 0.70

associated with the CDL 2007 reference data were not considered. Accordingly, the CDL-based assessment provided a conservative estimate of accuracy. Further insights could be derived by thoroughly examining their interactions and impacts on the crop mapping performances.

Discussion

This study examined the potential of MODIS-NDVI data for the classification of cropland and the identification of four individual crop types across the GLB. The classification of cropland versus non-cropland was developed first utilizing a Landsat TM/ETM seamless mosaic image as a locational information source to derive phenology training data for the MODIS-NDVI data classification. The estimated county level cropland acreages were consistent with the NASS county level census data for a majority of counties in the US portion of the GLB. The main confusion was between cropland and forest, as well as cropland and low density residential cover. The changing illumination angle may affect NDVI temporal profiles (Goward *et al.*, 1991), causing misclassification of forest and cropland in high

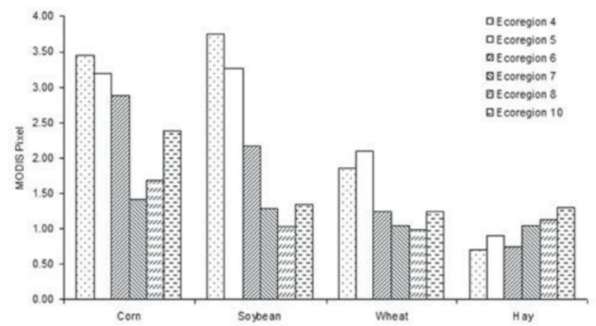

Figure 4. Mean patch size for corn, soybeans, wheat, and hay by ecoregion.

latitude regions. Confusion between cropland and urban land-cover has been reported by many urban remote sensing classification researches using single date imagery analysis (i.e., Paola and Schowengerdt, 1995; Wu and Murray, 2003). We expected that the phenology-based MODIS-NDVI classification would improve the performance due to the multi-temporal approach, however, the challenge remains for low density residential areas where MODIS-NDVI pixels have mixed cover types of vegetation and impervious cover. Further research is needed to understand the composition and the levels of spatial mixture, as well as their impacts on the MODIS-NDVI temporal profiles. One of the main challenges for individual crop type classification was the lack of available training data. This is especially the case if crop map products for multiple years or crop rotation information are needed. The research objective of this study was to assess the classification performance using MODIS-NDVI image-based training data (end-member) selection. Training data points were selected primarily through the visual interpretation of MODIS-NDVI temporal profiles. This approach could be employed in other studies, regions or years (independent of reference crop data). For the 2002 MODIS-NDVI classification, selected agricultural statistics were also incorporated as priors to improve the neural network training and classification. This can be particularly useful for crop mapping in the absence of available ground truth or training data, because agricultural statistics are generally available for most agricultural regions. Although no pixel-wise accuracy assessments were conducted for 2002 MODIS-NDVI classification, we expected an improved classification performance by incorporating priors, as suggested by many other researchers (McIver and Friedl, 2002; Strahler *et al.*, 1980).

The pixel-wise accuracy assessments for the 2007 MODIS-NDVI classifications were compared across a sub-set of ecoregions. The dominate sources of confusions were between corn and soybean, resulting in moderate levels of commission and omission errors. These two crop types exhibited similar temporal NDVI profiles. To increase future performance, other MODIS products such as surface reflectance and surface temperature, could be included as input features to enhance separability (e.g., Chang *et al.* 2007). This research also demonstrated that the spatial variations of classification errors were closely related to crop field sizes; especially, for soybeans and hay. Ecoregions 4, 5, and 6 had good classification performance for soybeans because of large soybean patch sizes. However, they had small crop patch sizes for hay, resulted in difficulty of hay classification. The scale factor in remote sensing (Woodcock and Strahler, 1987) and the impacts of image misregistration (Townshend *et al*, 1992) are well suited for understanding the spatial variation of classification performances.

The phenology-based classification was conducted at ecoregion scale across the GLB. This was attributed to the highly variable nature of cropland phenology across large geographic regions (Lobell and Asner, 2004; Wardlow *et al.*, 2008). The main disadvantage of image stratification before classification is level-of-effort. Instead of using a single universal classifier, we developed 12 individual classifiers for the entire study region. Independent training, validation and testing were also developed for each classifier. This suggests a trade-off of classification performance and cost. Future study on this topic is needed to determine the feasibility of reducing the total number of stratification elements and still maintain the overall classification performance. Also, further research is needed to examine the feasibility of using a "temporal generalization" of MODIS-NDVI classification over multiple years to support future crop type mapping efforts.

Conclusions

The 250 m MODIS-NDVI data was used for developing cropland mask and identification of four major crop types in the GLB. For cropland classification, the accuracies of crop acreage estimates at the county level were determined to be representative to those published by the USDA NASS for the same calendar year (SE = 10.2 km^2). Outliers included some counties with poor NDVI data quality, those with substantial plantings of crops other than corn, soybeans, wheat, and hay, and northern counties that were dominated by hay. For 2002 individual crop identification, part of agricultural census statistics were incorporated in the neural network classification, thus improved the crop mapping results. Pixel-wise accuracy assessments were conducted for 2007 crop type mapping. At GLB scale, the accuracies for corn, soybeans, and wheat are all at the acceptable level. Hay has the lowest accuracy because of highly variable temporal profiles. For soybeans and hay, there were large regional differences of classification performances which can be explained by varying crop patch sizes across ecoregions. Overall results suggested that a phenology-based MODIS-NDVI classification approach has high potential for regional scale monitoring of annual agricultural land-use activities across intensively managed croplands located in the Midwest region of the United States and Canada.

Acknowledgments

The authors would like to thank David Johnson and Rick Mueller for their assistance in providing the USDA NASS data in support of this research. This research was supported by the EPA's Global Earth Observation System of Systems (GEOSS) under the Advanced Monitoring Initiative (AMI), Project No. 07–35. The U.S. Environmental Protection Agency funded and conducted the research described in this paper. Although this work was reviewed by EPA and has been approved for publication, it may not necessarily reflect official Agency policy. Mention of any trade names or commercial products does not constitute endorsement or recommendation for use.

References

Benediktsson, J.A., P.H. Swain, and O.K. Ersoy, 1990. Neural network approaches versus statistical methods on classification of multisource remote sensing data, *IEEE Transactions on Geoscience Remote Sensing*, 28:40–551.

Bishop, C.M., 1995. *Neural Networks for Pattern Recognition*, Oxford University Press, Inc., New York, 504 p.

Bishop, C.M., 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, New Jersey, 738 p.

Carpenter, S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley, and V.H. Smith, 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen, *Ecological Applications*, 8:59–568.

Chang, J., M.C. Hansen, K., Pittman, M. Carroll, and C. DiMiceli, 2007. Corn and soybean mapping in the United States using MODIS time-series data sets, *Agronomy Journal*, 99:1654–1664.

Congalton, R.G., and K. Green, 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, Boca Raton, Florida, 160 p.

Craig, M.E., 2001. The NASS cropland data layer program, *Proceedings of the Third Annual International Conference on Geospatial Information in Agriculture*, Denver, Colorado, 05–07 November.

Crosbie, B., and P. Chow-Fraser, 1999. Percentage land use in the watershed determines the water and sediment quality of 22 marshes in the Great Lakes basin, *Canadian Journal of Fisheries Aquatic Sciences*, 56:1781–1791.

DeFries, R.S., and J.G.R. Townshend, 1994. NDVI derived land cover classifications at a global scale, *International Journal of Remote Sensing*, 5:3567–3586.

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

Duda, R.O., P.E. Hart, and D.G. Stork, 2001. *Pattern Classification*, John Wiley & Sons, Inc., New York, 654 p.

Foody, G.M., M.B. McCulloch, and W.B. Yates, 1995. The effect of training set size and composition on artificial neural network classification, *International Journal of Remote Sensing*, 16(9):1707–1723.

Foody, G.M., 2002. Status of land cover classification accuracy assessment, *Remote Sensing of Environment*, 80:185–201.

Friedl, M.A., D.K. McIver, J.C.F. Hodges, X.Y. Zhang, D. Muchoney, A.H. Strahler, C.E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf, 2002. Global land cover mapping from MODIS: Algorithms and early results, *Remote Sensing of Environment*, 83:287–302.

Gassman, P.W., S. Secchi, and M. Jha, 2006. Upper Mississippi River Basin modeling system Part 1: SWAT input data requirements and issues,*Coastal Hydrology and Processes* (V.P. Singh and Y.J.Xu, editors), pp. 103–116.

Goel, P.K., S.O. Prasher, R.M. Patel, J.A. Landry, R.B. Bonnell, and A. Viau, 2003. A. classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn, *Computers and Electronics in Agriculture*, 39(2):67–93.

Goward, S.N., B. Markham, D.G. Dye, W. Dulaney, and J. Yang, 1991. Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer, *Remote Sensing of Environment*, 35:257–277.

Groten, S.M.E., 1993. NDVI-crop monitoring and early yield assessment of Brukina Faso, *International Journal of Remote Sensing*, 14:1495–1515.

Homer, C., G.R.D. Ramsey, T.C. Edwards, and A. Falconer, 1997. Land cover-type modeling using a multi-scene Thematic Mapper mosaic, *Photogrammetric Engineering & Remote Sensing*, 63(1):59–67.

Homer, C., C. Huang, Y. Limin, B. Wylie and M. Coan, 2004. Development of a 2001 National Land Cover Database for the United States, *Photogrammetric Engineering & Remote Sensing*, 70(7):829–840.

Houghton, R.A., 1994. The worldwide extent of land-use change, *BioScience*, 44:305–313.

Houghton, R.A., and J.L. Hackler, 2000. Changes in terrestrial carbon storage in the United States, 1: The roles of agriculture and forestry, *Global Ecology and Biogeography*, 9(2):125–144.

Huang, C., L.S. Davis, and J.R.G. Townsend, 2002. An assessment of support vector machines for land cover classification, *International Journal of Remote Sensing*, 23:725–749.

Justice, C., A. Belward, J. Morisette, P. Lewis, J. Privette, and F. Baret, 2000. Developments in the 'validation' of satellite sensor products for the study of the land surface, *International Journal of Remote Sensing*, 21:3383–3390.

Knight, J.K., R.S. Lunetta, J. Ediriwickrema, and S. Khorram, 2006. 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.

Lal, R., L.M. Kimble, R.R. Follett, and C.V. Cole, 1998. *The Potential of U.S. Cropland to Sequester C and Mitigate the Greenhouse Effect*, Ann Arbor Press, Chelsea, Michigan.

Liu, J., M. Liu, H. Tian, D. Zhuang, Z. Zhang, W. Zhang, X. Tang, and X. Deng, 2005. Spatial and temporal patterns of China's cropland during 1990–2000: An analysis based on Landsat TM data, *Remote Sensing of Environment*, 98:442–456.

Liu, W., 2005. Comparison of non-linear mixture models: Sub-pixel classification, *Remote Sensing of Environment*, 94:145–154.

Lobell, D.B., and G.P. Asner, 2004. Cropland distributions from temporal unmixing of MODIS data, *Remote Sensing of Environment*, 93:412–422.

Loveland, T.R., B.C. Reed, J.F. Brown, D.O. Ohlen, Z. Zhu, L.Yang, and J.W. Merchant, 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, *International Journal of Remote Sensing*, 21(6–7):1303–1365.

Lunetta, R.S., J.F. Knight, J. Ediriwickrema, J.G. Lyon, L.D. Worthy, 2006. Land-cover change detection using multi-temporal MODIS-NDVI data, *Remote Sensing of Environment*, 105:142–154.

McGarigal, K., and B.J. Marks. 1995. *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure*, General Technical Report PNW-GTR-351, Portland, Oregon: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

McIver, D.K., and M.A. Friedl, 2002. Using prior probabilities in decision-tree classification of remotely sensed data, *Remote Sensing of Environment*, 81:253–261.

Muchoney, D., J. Borak, H. Chi, M. Friedl, S. Gopal, N. Hodges, N., Morrow, and A. Strahler, 1999. Applications of the MODIS global supervised classification model to vegetation and land cover mapping of Central America, *International Journal of Remote Sensing*, 21:1115–1138.

NASS (National Agriculture Statistical Service), 2004. *2002 Census of Agriculture: Volume 1 Geographic Area Series*, U.S. Government Printing Office, Washington, D.C., URL: *http://www.nass.usda.gov/census*, 561 p. and Appendices (last date accessed: 04 August 2009).

NASS (National Agriculture Statistical Service), 2008. URL: *http://www.nass.usda.gov/research/Cropland/metadata/metadata_mi07.htm* (last date accessed: 04 August 2009).

Omernik, J.M., 1987. Ecoregions of the conterminous United States, *Annals of the Association of American Geographers*, 77(1):118–125.

Osborne, L.L., and M.J. Wiley, 1988. Empirical relationships between land use/cover and stream water quality in an agricultural watershed, *Journal of Environmental Management*, 26:9–27.

Paola, J., and R.A. Schowengerdt, 1995. A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification, *IEEE Transactions on Geoscience Remote Sensing*, 33(4):981–996.

Quarmby, N.A., M. Mimes, T.L. Hindle, and N. Silleos, I993. The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction, *International Journal of Remote Sensing*, 14:199–210.

Richard, M.D., and R.P. Lippmann, 1991. Neural network classifiers estimate Bayesian a posteriori probabilities, *Neural Computation*, 3:461–483.

Roberts, H., J. Lehar, and J.W. Dreher, 1987. Time series analysis with CLEAN, I. - Derivation of a spectrum, *Astronomical Journal*, 93(4):968–989.

Searchinger, T., R. Heimlich, R.A., Houghton, F. Dong, A. Elobeid, J. Fabiosa, S. Tokgoz, D. Hays, and T.H. Yu, 2008. Use of U.S. Croplands for biofuels increases greenhouse gases through emissions from land-use change, *Science*, 319:1238–1240.

Strahler, A.H., 1980. The use of prior probabilities in maximum likelihood classification of remotely sensed data, *Remote Sensing of Environment*, 10:135–163.

Townshend, J.R.G., C.O. Justice, C. Gurney, and J. McManus, 1992. The impact of misregistration on change detection, *IEEE Transactions on Geoscience Remote Sensing*, 30(5):1054–1060.

USEPA (U.S. Environmental Protection Agency), 2008. Explore our Multimedia Page, URL: *http://www.epa.gov* (last date accessed: 04 August 2009).

Xiao, X., J. Liu, D. Zhuang, S. Frolking, S. Boles, and B. Xu, 2003. Uncertainties in estimates of cropland area in China: A comparison between an AVHRR-derived dataset and a Landsat TM-derived dataset, *Global and Planetary Change*, 37:297–306.

Xiao, X., S. Boles, J. Liu, D. Zhuang, S. Frolking, and C. Li, 2005. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images, *Remote Sensing of Environment*, 95(4):480–492.

Wolter, P.T., C.A. Johnston, and G.J. Niemi, 2006. Land use land cover change in the U.S. Great Lakes basin 1992 to 2001, Journal of Great Lakes Research, 32:607–628.

Wardlow, B.D., S.L. Egbert, and J.H. Kastens, 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains, *Remote Sensing of Environment*, 108(3):290–310.

Wardlow, B.D., S.L. Egbert, 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains, *Remote Sensing of Environment*, 112(3):1096–1116.

Wardlow, B.D., J.H. Kastens, and S.L. Egbert, 2006. Using USDA crop progress data for the evaluation of greenup onset date calculated from MODIS 250-meter data, *Photogrammetric Engineering & Remote Sensing*, 72(11):1225–1234.

Woodcock, C.E., and A.H. Strahler, 1987. The factor of scale in remote sensing, *Remote Sensing of Environment*, 21:311–332.

Wu, C., and A.T. Murray, 2003. Estimating impervious surface distribution by spectral mixture analysis, *Remote Sensing of Environment*, 84:493–505.

Zell, A., N. Mache, R. Hubner, G. Mamier, M. Vogt, K. Hermann, M. Schmalzl, T. Sommer, A. Hatzigeorgiou, S. Doring, and D. Posselt, 1993. *SNNS: Stuttgart Neural-Network Simulator, Institute Parallel Distributed High-Performance Systems*, University Stuttgart, Germany, Technical Report 3/93, 338 p.

Zhan, X., R. Sohlberg, J.R.G. Townshend, C. DiMiceli, M. Carroll, J.C. Eastman, M. Hansen, and R.S. DeFries, 2002. Detection of land cover changes using MODIS 250 m data, *Remote Sensing of Environment*, 83:337–351.

(Received 07 August 2008; accepted 11 December 2008: revised 27 January 2009)