Geocarto International
2011, 1–18, i*First* article

Taylor & Francis
Taylor & Francis Group

Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program

Claire Boryan*, Zhengwei Yang, Rick Mueller and Mike Craig

Department of Agriculture, National Agricultural Statistics Service, 3251 Old Lee Highway, Room 305, Fairfax VA 22030, USA

(*Received 2 November 2010; final version received 8 February 2011*)

The National Agricultural Statistics Service (NASS) of the US Department of Agriculture (USDA) produces the Cropland Data Layer (CDL) product, which is a raster-formatted, geo-referenced, crop-specific, land cover map. CDL program inputs include medium resolution satellite imagery, USDA collected ground truth and other ancillary data, such as the National Land Cover Data set. A decision tree-supervised classification method is used to generate the freely available state-level crop cover classifications and provide crop acreage estimates based upon the CDL and NASS June Agricultural Survey ground truth to the NASS Agricultural Statistics Board. This paper provides an overview of the NASS CDL program. It describes various input data, processing procedures, classification and validation, accuracy assessment, CDL product specifications, dissemination venues and the crop acreage estimation methodology. In general, total crop mapping accuracies for the 2009 CDLs ranged from 85% to 95% for the major crop categories.

Keywords: cropland classification; agriculture; Advanced Wide Field Sensor; crop estimates

Downloaded By: [George Mason University] At: 13:33 14 April 2011

1. Introduction

The mission of the US Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) is to provide timely, accurate and useful statistics in service to US agriculture. In 2009, the NASS Cropland Data Layer (CDL) program played an important role toward fulfilling this mission by providing operational in-season acreage estimates to the NASS Agricultural Statistics Board (ASB) and Field Offices (FOs) for 15 crops in 27 states. The 2009 CDL program covered many different crops, such as corn, soybeans, wheat, rice and cotton, etc. It provided updated acreage estimates throughout the growing season as increased quantities of farmer reported and satellite data became available. Revised CDLs, for several key states, were generated and estimates provided to the ASB and FOs up to six times during the growing season to provide input in setting acreage estimate updates.

The CDL product is a comprehensive, raster-formatted, geo-referenced, crop-specific land cover classification with a spatial resolution of 56 m that utilizes ortho-rectified imagery to accurately and geospatially identify field crop types. On 4 January 2010, 48 state-level CDL land cover products, for crop year 2009, were

*Corresponding author. Email: claire_boryan@nass.usda.gov

ISSN 1010-6049 print/ISSN 1752-0762 online
This work was authored as part of the Contributors' official duties as Employees of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law.
DOI: 10.1080/10106049.2011.562309
http://www.informaworld.com

publicly disseminated as Geographic Information System (GIS) data layers. Twenty-seven state-level CDL products were completed in season and 21 were completed in the post season. These GIS products are valuable resources for government agencies, private sector organizations, scientists, educators, and students who use land cover information.

CDL products have been used in a variety of research applications including assessing the utility of 500 m Moderate Resolution Imaging Spectroradiometer (MODIS) Time-Series Data for mapping corn and soybeans in the US (Chang *et al.* 2007), validating plant functional type maps developed from MODIS data using multisource evidential reasoning (Sun *et al.* 2008), examining the relationship between agricultural chemical exposure and cancer (Maxwell *et al.* 2010) to flood mapping assessment with satellite images (Shan *et al.* 2010). The CDL was also used to evaluate the use of high spatial resolution aerial imagery to monitor tree cover in agricultural landscapes in North and South Dakota (Liknes *et al.* 2010) and to assess automated determination of management units for precision soil conservation (Gelder *et al.* 2008). Additional reported uses of the CDL products include agribusiness, change detection, yield, crop intensity and rotation, education, ethanol, epidemiology, as well as assessments of water use, watershed, environmental risk, disaster response and forest fire potential.

This paper provides an overview of the current NASS CDL program, including method and inputs used in CDL production. Additionally, the description of CDL applications is provided to help users more wisely interpret and take advantage of the freely available crop-specific land cover classifications for alternative applications. The major inputs to the recent CDL program are detailed including satellite and ancillary data, sources of ground truth, software, classification and estimation procedures, accuracy assessment, results, and metadata. Figure 1 illustrates the 2009 state-level CDL image products. The legend identifies aggregated agricultural and non-agricultural land cover categories by decreasing acreage.

Downloaded By: [George Mason University] At: 13:33 14 April 2011

2. Background

The image processing and acreage estimation software first used to create the CDL was known as Peditor. This ‘in-house’ software, based on Pascal and FORTRAN, was originally written in the 1970s and was updated and maintained by NASS through 2006. It included digitizing, labelling, clustering, data pre-processing, Maximum Likelihood classifier, and acreage estimation components. Advantages of Peditor included the ability to produce statewide CDL image products and accuracy assessments, link multiple programmes, and most importantly estimate crop acreage with a simple linear regression method. The quality of the CDL products was high with classification accuracies ranging in the low to mid-90% for major crops. At the time, no commercial software could conduct all of the necessary operations performed by Peditor (Ozga and Craig 1995). Additionally, in the early 1990s, the Remote Sensing Project software was developed using Microsoft Visual FoxPro to manage the ground truth data collection, digitization and field acreage correction efforts.

From 1997 to 2005, the NASS CDL program used ground truth collected during the June Agricultural Survey (JAS). Every June, approximately 11,000 one-square mile segments are surveyed as a part of the JAS. The JAS segments are made up of approximately 41,000 individual farms that are enumerated to identify the planting

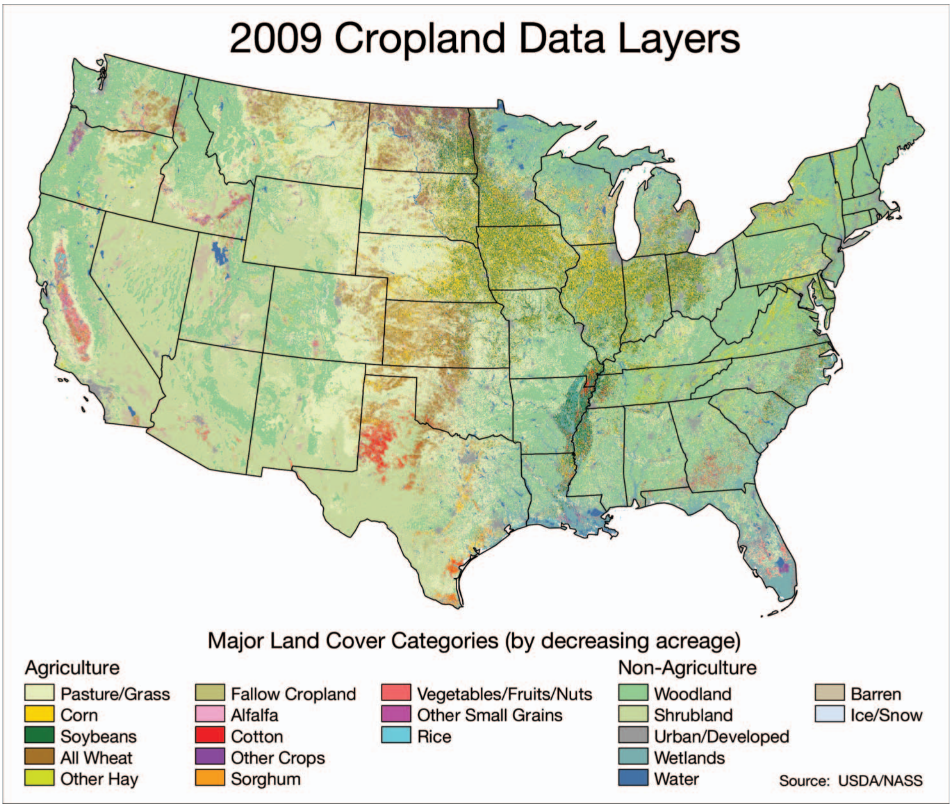

Figure 1. The 2009 cropland data layer products. The legend identifies aggregated agricultural and non-agricultural land cover categories by decreasing acreage.

intentions for all agricultural land within the segments, including planted acreage and acreage intended for harvest. The selection of JAS segments is based on a national area sampling frame (ASF) that is the statistical foundation for providing estimates with complete coverage of US agriculture. The ASF is a stratification of land cover in the US by percent cultivated cropland.

During this period, the JAS crop data were used as ground truth for maximum likelihood-based supervised classification. JAS segments were also utilized to perform a simple linear regression to derive crop-specific acreage estimates (Allen and Hanuschak 1988, Ozga and Craig 1995). One drawback of the JAS segment data was that the segments required manual digitization of all field-level boundaries prior to use in the CDL program, a labour intensive activity. By 2007, the JAS segment data were no longer utilized within the CDL program as ground truth but were still used as an independent data source for the regression estimator.

The NASS used multi-spectral satellite imagery beginning in the 1970s to estimate acreage of large area crops in major producing states. NASS remote sensing programs initially used imagery from the Landsat Multi-Spectral Scanner instruments through the 1987 crop season at which time NASS began evaluating Landsat 5 Thematic Mapper (TM) and SPOT Image data as possible replacements. In 1991, Landsat TM was adopted for use in the program. By April 1999, Landsat TM and

Downloaded By: [George Mason University] At: 13:33 14 April 2011

Landsat Enhanced Thematic Mapper (ETM+) data were used in combination to produce crop acreage estimates and CDL image products for six major crop producing states (Mueller 2000, Craig 2001). On 31 May 2003, the Landsat ETM+ sensor experienced an anomaly in its scan line corrector at which time NASS began to evaluate alternative sources of data including imagery from the Indian Remote Sensing Satellite (IRS) RESOURCESAT-1 launched in October 2003. The IRS RESOURCESAT-1 Advanced Wide Field Sensor (AWiFS) became the sensor of choice for the NASS CDL program after careful, quantitative evaluation and comparison of AWiFS with Landsat data for CDL production (Boryan and Craig 2005, Seffrin 2007, Johnson 2008).

The growth of the CDL program to include more states from 1997 to 2006 was primarily through partnerships and cooperative agreements with federal and state governments and universities. It was determined, however, that producing CDLs within NASS headquarters was the most efficient means to expand the program.

Beginning in 2006, the CDL program underwent a major restructuring and modernization effort. The original software and data inputs were replaced with a commercial suite of software including Rulequest Research's See5 decision tree software, ERDAS Imagine remote sensing software, Environmental Systems Research Institute's (ESRI) ArcGIS, Statistical Analysis Software (SAS) and new data sources including RESOURCESAT-1 AWiFS data, and 578 Administrative and Common Land Unit (CLU) data from the Farm Service Agency (FSA). Tremendous efficiency gains were achieved due to the modernization allowing for the generation of in-season crop acreage estimates, a goal never achieved using the older operational process, methods and data.

In 2007, the CDL program provided acreage estimates for 13 states and nine crops to the NASS ASB for the October Crop Production Report (PR). For the first time, remote sensing estimates were used in season for setting the NASS official state acreage estimates, a milestone for the program. An additional eight CDL state image products were generated after the growing season for a total of 21 2007 CDL state products. In 2008, research was conducted by Boryan *et al*. (2008) to determine if accurate estimates could be derived earlier in the growing season. A total of 35 2008 CDL state products were generated and, based upon the previous research, acreage estimates were provided to the NASS ASB for the first time to meet June, August, September and October production deadlines.

Downloaded By: [George Mason University] At: 13:33 14 April 2011

3. Cropland data layer program inputs

The major inputs to the current CDL program include AWiFS, Landsat TM and ETM+, MODIS satellite data, the FSA CLU data for agricultural ground truth and the National Land Cover Data set (NLCD) 2001 for non-agricultural ground truth and ancillary data sources including US Geological Survey (USGS) digital elevation, NLCD 2001 tree canopy and NLCD 2001 imperviousness data layers.

3.1. Imagery

The primary source of satellite data used by the CDL program is acquired by the IRS RESOURCESAT-1 sensor launched in 2003. The payload of RESOURCE-SAT-1 includes three sensors: the Linear Imaging Self Scanner (LISS) IV, LISS-III and AWiFS that is the primary sensor for the CDL program. AWiFS specifications include a 56-m spatial resolution at nadir, a large swath width (740 km), four channels including green, red, near-infrared (NIR) and middle-infrared (MIR), a rapid revisit (5-day repeat) capability, 10-bit quantization and a 5-year design life. The AWiFS has a moderate spatial resolution that is appropriate for identifying large homogenous crop fields. The large swath is made possible with identical AWiFS multispectral cameras (A and B) acquiring data with an 8.4 km overlap and is particularly useful as large geographic areas can be acquired in single day passes. The spectral characteristics of AWiFS correspond closely with Landsat TM, which is no coincidence as AWiFS designers matched bands closely to bands two through five of Landsat TM. Table 1 lists the sensor specifications of Landsat TM vs. AWiFS.

Landsat TM bands two through five are particularly useful for vegetation assessments specifically, band 2: 0.52–0.60 μm (green) to the green reflectance of healthy vegetation, band 3: 0.63–0.69 μm (red) for vegetative discrimination, band 4: 0.76–0.90 μm (NIR) to the percentage of vegetative biomass present and band 5: 1.55–1.75 μm (MIR) to the water content of plants (Jensen 2007).

As a member of the USDA's Satellite Image Archive (SIA) administered by the Foreign Agricultural Service, NASS has the opportunity to utilize any and all available AWiFS data collected by the SIA for CDL processing. The AWiFS data are collected by cameras A & B mounted side by side and acquisitions are identified by path/row/quad. Camera A (western side of path) acquires data in quads A and C and camera B (eastern side of path) acquires data in quads B and D. Figure 2 illustrates an AWiFS single date acquisition with quad collections superimposed on the image.

The majority of AWiFS acquisitions purchased by the SIA cover the Midwestern and Great Plains states where most of the corn, soybeans and winter wheat are grown in the US. The data are ortho-rectified and GeoTIFF formatted. They have 10 bit quantization and Lambert Conformal Conic projection. The NASS reprojects the data to Albers Conical Equal Area (Albers), GRS 1980 (spheroid) and NAD83 (datum), and mosaics same day acquisitions.

In 2009, NASS regularly supplemented AWiFS data with Level 1T (terrain corrected) Landsat TM and ETM+ data for CDL production as the entire USGS Landsat Data Archive became available at no charge (USGS 2010). The Landsat data were downloaded from Glovis (http://glovis.usgs.gov). Image data processing steps included converting the data from GeoTIFF to ERDAS Imagine image (.img) format, reprojecting from Universal Transverse Mercator (UTM) to Albers,

Table 1. Landsat Thematic Mapper and Advanced Wide Field Sensor specifications.

	TM	AWiFS
Altitude	705 km	817 km
Equatorial crossing time	9:45 ± 15 min	10:30 ± 5 min
Temporal resolution	16 days	5 days
Spatial resolution	30 × 30 m (reflective), 120 × 120 m (thermal)	56 × 56 m
Radiometric resolution	8 bit (256)	10 bit (1024)
Spectral resolution	6 (B, G, R, NIR, SWIR, MIR) + Thermal IR	4 (G, R, NIR,SWIR)
Swath width	185 km	740 km
Scene size	184 × 170 km	370 × 370 km

Downloaded By: [George Mason University] At: 13:33 14 April 2011

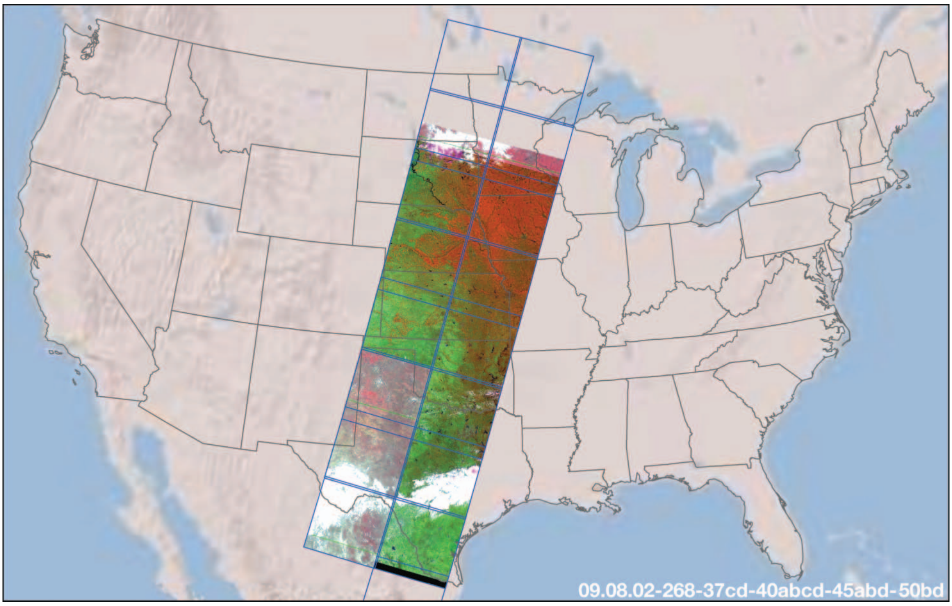

Figure 2. Indian remote sensing satellite resources at 1 – Advanced Wide Field Sensor imagery acquired on 2 August 2009. Acquisition descriptions include path/row/quad information. The brightly coloured quads are those used in CDL processing.

resampling from 30 to 56 m using bilinear interpolation and mosaicing same day acquisitions. The bilinear interpolation method was selected to more closely represent the spectral values of the original neighbouring pixels.

National Aeronautics and Space Administration MODIS 16-day Normalized Difference Vegetation Index (NDVI) composites were also used to supplement the AWiFS and Landsat TM and ETM+ data. The 250 m MODIS data were downloaded from the USGS's Land Processes Distributed Active Archive Center, resampled to 56 m and reprojected to Albers.

To produce the 2009 CDLs of all 48 conterminous states 477 AWiFS scenes, 1357 Landsat TM scenes, 138 Landsat ETM+ scenes, and 26 MODIS 16 day NDVI composite images were utilized. AWiFS and Landsat TM and ETM+ data were selected based on a low percentage of cloud cover and with the goal of matching the dates of available imagery with the phenological cycle of the crops. Crop progress and condition information for major crops in all 48 states was utilized by analysts to determine optimal dates for imagery selection. Crop progress and condition charts are available on the NASS web site at http://www.nass.usda.gov/Charts_and_Maps/Crop_Progress_&_Condition/2009/index.asp.

3.2. Ground truth

The main source of agricultural ground truth for the CDL supervised classification training is the USDA's FSA CLU data. This standardized GIS data layer of the nation's farms and fields was established to support farm commodity, conservation programs and disaster response (Heard 2002, Anderson *et al.* 2005). CLU data are updated every growing season when producers report crop type and crop acreage for

Downloaded By: [George Mason University] At: 13:33 14 April 2011

their fields to FSA county offices. The FSA CLU program is operational in over 2300 FSA county offices. The program includes all states and extensive coverage of 'major crops', which are those for which farmers receive financial subsidies. The CLU system creates digitized polygon boundaries of semi-permanent 'fields' in ESRI shape file format. Attribute information is maintained in a separate database format known as FSA 578 Administrative Data (Heard 2002). Two important advantages of the FSA CLU data for CDL processing are the sheer volume of agricultural data and that the CLU polygons are digitized in the FSA county offices thereby creating a comprehensive agricultural data set that requires no manual digitizing by NASS staff. The FSA CLU data are confidential data sources and are not provided or shared with anyone outside of NASS. Figure 3 illustrates FSA CLU ground truth polygons of a 184 km^2 area in Nebraska. The yellow polygons are corn fields, dark green polygons are soybeans and pale green polygons are pasture/grass.

The preparation of FSA CLU data for use in the CDL production occurs in three phases, the first involves the delineation of CLU polygons using ESRI arc GIS 9.3 software. Certified CLUs are provided by FSA in shape file format at the county level. The original CLUs are merged to create a state-level shape file and buffered inward by 30 or 56 m, depending on the state, so that the centre of the crop fields, rather than the field boundaries or edges are targeted for sampling. The state shape files are reprojected from UTM to Albers. At this point, the attributes attached to the CLU polygons include state, county of administration, county of geography, CLU polygon acreage, farm, tract, CLU number and an NASS unique identifier that

Figure 3. Farm service agency common land unit ground truth polygons.

Downloaded By: [George Mason University] At: 13:33 14 April 2011

is generated from a combination of these attributes. The CLU shape file does not contain crop-specific information. These steps are performed in sequence using python scripts and ArcGIS processing tools. This phase in ground truth preparation requires processing only once per state annually. Once the FSA CLU county polygon data are merged to the state level and buffered, they are ready for linking with the FSA 578 attribute data that includes all crop-specific information including crop type, status and intention codes.

Updating the FSA 578 attribute data provides the opportunity to utilize the most current ground truth available, as farmers continue to report and/or update their cropping intentions throughout the growing season. The CLU fields are sorted by crop type, size and attributes so that when separated into ground truth and validation data sets, they include the optimal range of crops and acreages. CLU polygons that are either planted to more than one crop or have acreage discrepancies of more than 10% between the CLU polygon and the 578 attribute data are excluded from the final ground truth data set. For example, Nebraska FSA 578 attribute data accessed on 15 September 2009 included 484,410 CLU polygon records. After filtering on acreage discrepancies and multiple crop types, 251,016 CLU polygon records remained in the ground truth data set.

Once the FSA CLU polygons are linked with the 578 data, the state-level CLU shape files are prepared for use with the See5 decision tree software. An important requirement of See5 is that all inputs must be in raster format of identical cell size and projection. The shape files are divided into separate training and validation fields using a 70% training and 30% validation breakdown and converted into continuous raster layers. The cell size of all raster layers are set to 56 m, a predetermined cell size for all inputs to match the AWiFS spatial resolution and the extent of FSA CLU raster layers are set to match the extent of all other inputs to the classification.

For all of the advantages of the FSA data, there still exists a shortcoming. Many CLU polygons include more than one crop type per CLU (Craig 2005). In order to use the FSA data, CLUs with mixed crop types, except certain double crops such as winter wheat followed by soybeans, are excluded from the ground truth used in the classification process. Fortunately, this shortcoming is greatly outweighed by the sheer volume of crop data available from the FSA CLU program. The CLU data currently stands as the cornerstone of the CDL program. Being a comprehensive agricultural data set that requires minimal preparation and can be updated multiple times during the growing season greatly outweighs the disadvantage. Using the FSA CLU and 578 attribute data for training has dramatically increased the volume and timeliness of available ground truth and thereby increased the scope, efficiency, and accuracy of the operational CDL program.

The current ground truth data source for acreage estimation is still from the JAS. The 11,000 area segments selected nationwide for the JAS account for approximately 2.5% of total land area in the US JAS segments range in size from one-tenth of one-square mile in urban areas to approximately one-square mile in cultivated areas to as much as 4–8 square miles in open range. This stratification of land facilitates the identification and higher selection rate for segments in intensively cultivated land areas that takes place at a rate of approximately 1:125. Segments in less-cultivated areas are selected at a rate of 1:250 to 1:500. The JAS data are based on a probability survey and considered statistically robust. The 150–400 square miles of ground truth collected on average per state during the JAS provides the basis for building the

Downloaded By: [George Mason University] At: 13:33 14 April 2011

regression estimation model. The farmer reported data collected in the JAS are only used internally by NASS and held strictly confidential.

3.3. Ancillary data

Several raster-based data layers from USGS were used as ancillary data sources in the production of the 2009 CDL products. These include the National Elevation Data set (NED), the NLCD 2001 tree canopy and the NLCD 2001 imperviousness products. The NED is 30 m in spatial resolution. The tree canopy and imperviousness layers are by-products of the 2001 NLCD, a national product completed in January of 2007 (Homer *et al*. 2004, 2007). These data sets were merged to create a US national level product, reprojected to Albers and resampled to 56 m to match the native AWiFS pixel resolution. These ancillary products facilitated the separation of agricultural from non-agricultural land cover categories. Figure 4 illustrates the NLCD 2001 data of an area in Nebraska, US. Representations include grey – urban, green – grassland, dark blue – water and light blue – wetlands. Approximate image area is 765 km^2.

The NLCD 2001 was the source of non-agricultural ground truth for CDL processing. Features such as water, urban, barren, forest, shrub/scrub, grassland herbaceous and wetlands were sampled from the NLCD 2001. Since NASS and the FSA, do not collect non-agricultural ground truth, the NLCD 2001 was deemed to be the best available source. Although the NLCD 2001 is a dated product, NASS has found that by using current imagery, the See5 classifier has correctly identified areas of urban expansion, agricultural land conversion and forest clearing. The NASS has not made an attempt to quantify these changes.

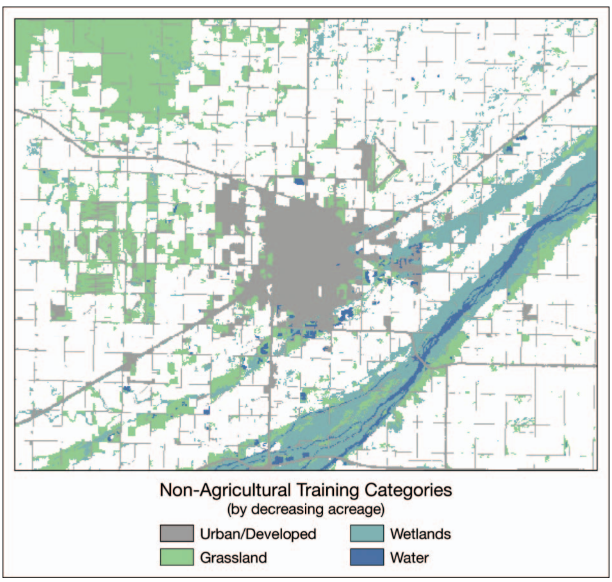

Figure 4. The National Land Cover Data set 2001 of Nebraska. Categories sampled from the National Land Cover Data set 2001 include non-agricultural categories such as urban, water, wetlands and forest.

Downloaded By: [George Mason University] At: 13:33 14 April 2011

4. Classification

Supervised classification of the cropland cover type with raw imagery and ancillary data is performed using the FSA CLU and NLCD 2001 ground truth sample points as training for the See5 decision tree classifier. Training samples (pixels) are used by the classifier to derive the state-level decision trees. State-level samples are collected from the FSA CLU data to create agricultural training data and from the NLCD 2001 to create non-agricultural training data. The NLCD sampling tool kit provided by USGS is an ERDAS Imagine plug-in component that interfaces ERDAS Imagine with See5. The NLCD sampling tool kit was customized by NASS to increase the number of bands of data (83–1000) that could be used as inputs to the classification process.

Pre-processed AWiFS, Landsat and ancillary data are loaded into the sampling tool as 'independent variables'. In the pre-processing phase, images are selected based on optimal dates for separation of crop types and with maximum geographic coverage. FSA and NLCD samples are collected separately. When deriving ground truth sample points, the FSA CLU data layer (or USGS NLCD 2001 data) is loaded as the 'dependent layer'. A number, per cent or all points within the dependent layer is sampled. A random stratified sampling scheme based upon crop or non-agriculture categories is utilized. Names and data files are outputs of this process. The names files identify the number of training samples selected, values ignored, sampling method, output form, the dependent layer including the directory path and all independent layers listed as individual bands.

In classification, See5's boosting algorithm is set to 10 trials and global pruning at 25% based on positive results in the literature (Quinlan 1996). Analysis is performed at the pixel level. Positive attributes of See5 include allowing for an abundance of satellite imagery to be used in the classification process; the powerful See5 boosting algorithm that reviews the results multiple times to refine or 'prune' the decision tree; and See5's tolerance of image noise, such as clouds, haze or even scan gaps in the Landsat ETM+ imagery. The raw state-level CDL image products and the corresponding confidence layers are produced without any form of smoothing or filtering of results, the only exception being the citrus category in the state of Florida. A description of the CDL confidence layer is included in Section 6.3 of this paper.

Downloaded By: [George Mason University] At: 13:33 14 April 2011

5. Accuracy

The accuracies of the CDL agricultural crop categories are derived by comparing the CDLs with independent validation data extracted from the FSA CLU ground truth data. During the ground truth preparation phase, 30% of the available FSA data (at the polygon level) are set aside for the purpose of validating the output product at the pixel level. In CDL production, the Kappa coefficients were used for measuring the difference between the actual agreement in the accuracy matrix and the agreement that would occur by chance (Congalton and Green 1999). The number of 'correct pixels' in the accuracy table represents the total number of independent validation pixels correctly identified and quantifies the abundance of crops within a state. The producer and user accuracies are generally 85% to 95% correct for major crop categories. Accuracy statistics are included in the metadata provided with all CDL image products. Accuracies for the non-agricultural categories are not provided. Table 2 contains an example of the accuracy statistics generated for the Nebraska 2009 CDL.

Downloaded By: [George Mason University] At: 13:33 14 April 2011

Table 2. The accuracy table for the Nebraska 2009 Cropland Data Layer.

USDA, National Agricultural Statistics Service, 2009 Nebraska Cropland Data Layer								
Statewide agricultural accuracy report								
Crop-specific covers only		*Correct	Accuracy	Error	κ			
Overall Accuracy		1669764	91.65%	8.35%	0.8618			
Cover Type	Attribute code	*Correct pixels	Producer's accuracy (%)	Omission error (%)	κ	User's accuracy (%)	Commission error (%)	Cond'l κ
Corn	1	1019448	97.12	2.88	0.9335	98.64	1.36	0.9679
Sorghum	4	7675	59.87	40.13	0.5966	81.81	18.19	0.81668
Soybeans	5	447419	96.43	3.57	0.9523	97.44	2.56	0.9657
Sunflowers	6	1237	40.58	59.42	0.4054	84.55	15.45	0.8453
Popcorn	13	608	19.31	80.69	0.1928	85.39	14.61	0.8537
Barley	21	18	15.93	84.07	0.1593	78.26	21.74	0.7826
Durum wheat	22	0	n/a	n/a	n/a	0.00	100.00	0.0000
Spring wheat	23	76	11.33	88.67	0.1132	83.52	16.48	0.8351
Winter wheat	24	92688	90.72	9.28	0.9021	96.79	3.21	0.9660
Other grains	25	103	13.05	86.95	0.1305	69.13	30.87	0.6911
WW/soybeans	26	29	13.00	87.00	0.1300	28.71	71.29	0.2870
Rye	27	128	8.82	91.18	0.0881	55.17	44.83	0.5514
Oats	28	997	25.94	74.06	0.2588	64.20	35.80	0.6412
Millet	29	1574	37.19	62.81	0.3709	53.48	46.52	0.5337
Alfalfa	36	30093	72.82	27.18	0.7231	89.09	10.91	0.8884
Other hays	37	10778	15.85	84.15	0.1526	86.18	13.82	0.8565
Beets	41	1360	86.02	13.98	0.8601	91.09	8.91	0.9108

(*continued*)

Downloaded By: [George Mason University] At: 13:33 14 April 2011

Table 2. (*Continued*).

USDA, National Agricultural Statistics Service, 2009 Nebraska Cropland Data Layer								
Statewide agricultural accuracy report								
Crop-specific covers only		*Correct	Accuracy	Error	κ			
Overall Accuracy		1669764	91.65%	8.35%	0.8618			
Cover Type	Attribute code	*Correct pixels	Producer's accuracy (%)	Omission error (%)	κ	User's accuracy (%)	Commission error (%)	Cond'l κ
Potatoes	43	1271	83.02	16.98	0.8301	97.62	2.38	0.9762
Other crops	44	0	0.00	100.00	0.0000	n/a	n/a	n/a
Misc. vegetables	47	0	n/a	n/a	n/a	0.00	100.00	0.0000
Watermelon	48	0	n/a	n/a	n/a	0.00	100.00	0.0000
Peas	53	0	0.00	100.00	0.0000	n/a	n/a	n/a
Clover/wildflowers	58	0	0.00	100.00	0.0000	n/a	n/a	n/a
Idle/fallow	61	54262	85.96	14.04	0.8550	94.56	5.44	0.9437
Grapes	69	0	0.00	100.00	0.0000	n/a	n/a	n/a
Christmas trees	70	0	0.00	100.00	0.0000	n/a	n/a	n/a

Note: *Correct pixels represents the total number of independent validation pixels correctly identified in the error matrix. The shaded areas highlights the Nebraska 2009 Cropland Data Layer overall accuracy and specific accuracies for the three major (most acreage) crops grown in Nebraska in 2009.

Downloaded By: [George Mason University] At: 13:33 14 April 2011

6. Cropland data layer products

6.1. Crop acreage estimates

One of the major purposes for producing the CDL is to derive the supplementary crop acreage estimates for various crops. Intuitively, crop acreage can be derived from counting pixels of a specific crop type. Pixel counting estimates, however, consistently underestimate the actual acreage number as compared with NASS official estimates. Therefore, NASS builds a linear regression model from the CDL pixel data and segment summary data collected as part of the NASS JAS as follows:

$$Y = a + bX \tag{1}$$

where Y is the estimated acres and X is the independent variable representing CDL classified acres.

The coefficients a and b are estimated from JAS reported acres and CDL classified acres using a least square estimation method. This method computes the best-fitting regression line for the observed data (CDL pixels) by minimizing the sum of the squares of the vertical deviations from each data point (JAS segment) to the line. The regression is performed at the segment level for all strata on the JAS segments and classified pixel data. The reported acres of JAS segments and the pixel summaries of the geographically corresponding fields on the CDL represent dependent and independent variables, respectively. This CDL-JAS regression estimation is preferred as it is able to improve upon the JAS estimate based on the correlation between the JAS reported acres and the CDL pixel count in each stratum. The remote sensing based acreage estimate from the CDL-JAS regression model leads to an independent acreage estimate with a lower error rate (coefficient of variation) than direct expansion alone or direct pixel counting.

In the modelling process, segments identified as outliers that do not fit the linear regression relationship are reviewed and removed from consideration if in error. The correlation coefficient R^2 is used to measure the goodness of fitting of the regression line, i.e. the correlation between the CDL classified pixels and the JAS segment summary data. Figure 5 illustrates a linear regression performed on corn. Pixels classified to corn in the CDL (X axis) are regressed against JAS segment data (Y axis).

The regression scatter plot depicts corn planted in stratum 11 ($>80\%$ cultivated) in Nebraska 2009. The X axis reflects acres classified in the CDL product. The Y axis reflects reported acres in the JAS survey data. The small symbols represent JAS segments. The black, red and green symbols represent segments considered in the regression formula and were used to generate the acreage estimates. The blue dots are outliers as identified in the legend.

Presently, the CDL program provides supplementary acreage estimates to the NASS ASB and FOs to meet the June (winter wheat), August (corn and soybeans), September (winter wheat, corn, soybeans, cotton, rice, peanuts and other small grains), October (corn, soybeans and all other major field crops) and December (county acreage estimates for major crops) production deadlines. To meet this requirement, updated CDL products are generated multiple times during the season to provide acreage estimates with the highest accuracy at each point in the growing season. In 2009, state-level crop acreage estimates were provided to meet NASS production deadlines for 15 states for the June PR, 14 states for the August PR,

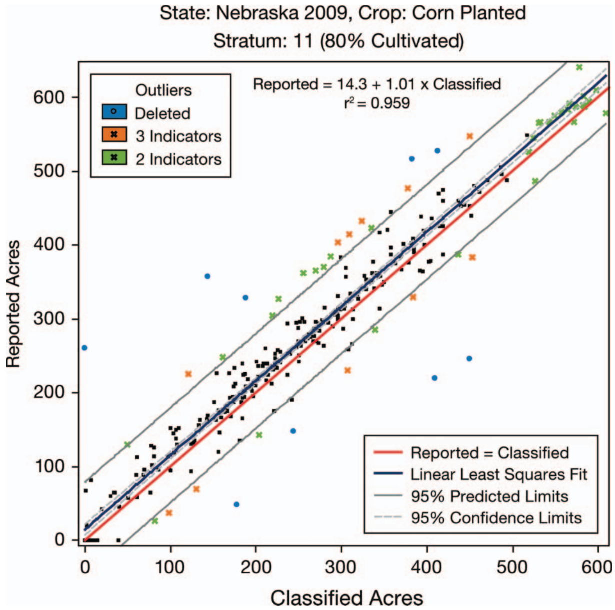

Figure 5. A linear regression performed on corn. Pixels classified to corn in the Cropland Data Layer (*X* axis) are regressed against June Agricultural Survey segment data (*Y* axis).

15 states for the September PR, 15 states for the Small Grain Summary and 27 states and a total of 15 crops for the final October PR.

6.2. The 2009 cropland data layer image products

In 2009, 27 CDL image products were created during the crop season to provide state-level acreage estimates to the NASS ASB and state FOs. Using funds provided by the US Environmental Protection Agency Landscape Ecology Branch, 21 additional CDL state image products were generated in the off season for a total of 48 statewide 2009 CDL products. The final CDL products are generated at the end of the crop season for the October crop report. CDL products created for earlier reporting deadlines are not released to the public. The CDL products have a spatial resolution of 30 m for CDLs produced prior to 2006 and 56 m for CDLs produced from 2007 to 2009. The CDL products on the Geospatial Data Gateway are provided in GeoTIFF format, UTM and NAD83 or World Geodetic System 1984 (WGS84) map projection. The 2009 CDLs are aggregated to standardized categories emphasizing agricultural land cover. The 2009 CDL image products, as well as all historic CDLS, can be downloaded free of charge from the National Resources Conservation Service Geospatial Data Gateway at http://datagateway.nrcs.usda.gov. Table 3 summarizes the historic record of statewide CDLs that are available for free download.

The NASS, in cooperation with George Mason University/Centre for Spatial Information Science and Systems, recently released a new interactive visualization portal called CropScape coincident with the release of the 2010 CDL products. CropScape serves all CDL data as a web service-based interactive map visualization,

Downloaded By: [George Mason University] At: 13:33 14 April 2011

Table 3. The historic record of state-level cropland data layers (1997–2009) available to the public.

State	Year 1997–2006	2007	2008	2009
Alabama			X	X
Arizona			X	X
Arkansas	1997–2006	X	X	X
California		X		X
Colorado			X	X
Connecticut	2002			X
Delaware	2002		X	X
Florida	2004			X
Georgia			X	X
Idaho	2005	X		X
Illinois	1999–2006	X	X	X
Indiana	2000–2006	X	X	X
Iowa	2000–2006	X	X	X
Kansas	2006	X	X	X
Kentucky			X	X
Louisiana	2004–2006	X	X	X
Maine				X
Maryland	2002		X	X
Massachusetts				X
Michigan		X	X	X
Minnesota	2006	X	X	X
Mississippi	1999–2006	X	X	X
Missouri	2001–2006	X	X	X
Montana		X		X
Nebraska	2001–2006	X	X	X
Nevada			X	X
New Hampshire				X
New Jersey	2002		X	X
New Mexico			X	X
New York	2002		X	X
North Carolina	2002		X	X
North Dakota	1997–2006	X	X	X
Ohio	2006	X	X	X
Oklahoma	2006	X	X	X
Oregon		X		X
Pennsylvania	2002		X	X
Rhode Island	2002			X
South Carolina			X	X
South Dakota	2006	X	X	X
Tennessee			X	X
Texas			X	X
Utah			X	X
Vermont				X
Virginia	2002		X	X
Washington	2006	X		X
West Virginia	2002		X	X
Wisconsin	2003–2006	X	X	X
Wyoming			X	X

Downloaded By: [George Mason University] At: 13:33 14 April 2011

dissemination and querying system. The CropScape web service provides open geospatial access and navigation, online mapping, statistical analysis, change detection, data retrieval and distribution. The CropScape web portal is available at http://nassgeodata.gmu.edu/CropScape.

6.3. Metadata

Each CDL product has a metadata file associated with it. The metadata includes the following information: identification, data quality, spatial data organization, spatial reference, entity and attribute distribution and reference. The associated metadata for each CDL is included with the Geospatial Data Gateway download and at http://www.nass.usda.gov/research/Cropland/metadata/meta.htm.

6.4. Classification confidence layer

Supplemental accuracy assessment data, in the form of associated confidence layers, which are not available through the CropScape web portal or the Geospatial Data Gateway are available by contacting the authors or HQ_RDD_GIB@nass.usda.gov. The confidence value is not a measure of accuracy for a given pixel in the classification but rather a measure of how well the decision to identify a pixel within a specific category fit within the decision tree rule set. Liu *et al.* (2004) provided additional information on the use of confidence layers in land cover classification.

7. Conclusion

This overview of the NASS CDL program included a brief history followed by a description of the major inputs to the CDL program including the use of AWiFS; Landsat TM and ETM+; MODIS satellite data; the FSA CLU and NLCD 2001 for ground truth and ancillary data sources. Additionally, descriptions of the software utilized including ArcGIS 9.3, See5, ERDAS Imagine, NLCD tool kit, and SAS; classification and estimation procedures, accuracy assessment, results and metadata were provided.

Recently, the CDL program covered all NASS speculative program crops providing updated acreage estimates throughout the growing season using the most up to date farmer reported and satellite data available. Additionally, for the first time in 2009, the freely available CDL products were created for all 48 conterminous states in the US. Having achieved this level of coverage, it is the goal of the CDL program to continue to provide yearly updates, at the state level, to meet the growing needs of our agricultural stakeholders.

The CDL program will continue to evaluate its ability to expand the quantity, scope and quality of crop acreage estimates provided to the NASS ASB and FOs to further the NASS mission of providing the most timely, accurate and useful agricultural statistics possible. Research will continue in an attempt to improve the CDL image products and acreage estimates. Techniques for enhancing the quality of available ground truth, improving the accuracy of small area but high value crops, improvements to spatial resolution and cropping intensity and rotational analysis are being investigated.

Downloaded By: [George Mason University] At: 13:33 14 April 2011

Acknowledgements

The authors thank the current NASS team working on the CDL program and the many analysts who worked on the program over the past 40 years. They extend a special thanks to Karla Koudelka (NASS) and Lee Ebinger (NASS) for their help in preparing the tables and graphics and to Dr. Barry Haack (George Mason University) for his valuable suggestions during the writing of this paper.

References

Allen, J.D. and Hanuschak, G.A., 1988. *The remote sensing applications program of the National Agricultural Statistics Service: 1980–1987*. U.S. Department of Agriculture, NASS Staff Report No. SRB-88-08.

Anderson, T., *et al.*, 2005. *USDA service center agencies geospatial data management team data management for common land unit data*. Available from: http://www.itc.nrcs.usda.gov/scdm/docs/DMP-CLU-DataManagementPlan.pdf [Accessed February 10 2009].

Boryan, C.G. and Craig, M.E., 2005. Multiresolution landsat TM and AWiFS sensor assessment for crop area estimation in Nebraska. *Proceedings from Pecora 16*, 22–27 October 2005, Sioux Falls, South Dakota.

Boryan, C.G., Craig, M.E., and Lindsey, M., 2008. Deriving essential dates of AWiFS and MODIS for the identification of corn and soybean fields in the U.S. heartland. *In*: *Proceedings from Pecora 17*, November 2008, Denver, Colorado.

Chang, J.C., *et al.*, 2007. Corn and soybean mapping in the United States using MODIS time-series data sets. *Agronomy Journal*, 99, 1654–1664.

Congalton, R.G. and Green, K., 1999. *Assessing the accuracy of remotely sensed data: principles and practices*. Boca Raton: Lewis Publishers.

Craig, M., 2005. *Using FSA administrative data in the NASS cropland data layer*. Fairfax, VA: NASS/RDD/GIB/SARS. Draft as of 9/7/2005; write-up of FSA data used for Nebraska 2002-2004 research; circulated administratively only in NASS.

Craig, M.E., 2001. The NASS cropland data layer program. *Presented at the Third International Conference on geospatial information in agriculture and forestry*, November 2001, Denver, Colorado.

Gelder, B., Cruse, R.M., and Kaleita, A.L., 2008. Automated determination of management units for precision conservation. *Journal of Soil and Water Conservation*, 63 (5), 273–279.

Heard, J., 2002. *USDA establishes a common land unit*. ESRI ArcUser Online. Available from: http://www.esri.com/news/arcuser/0402/usda.html [Accessed February 10 2010].

Homer, C., *et al.*, 2004. Development of a 2001, national land cover database for the United States. *Photogrammetric Engineering & Remote Sensing*, 70 (7), 829–840.

Homer, C., *et al.*, 2007. Completion of the 2001 national land cover database for the conterminous United States. *Photogrammetric Engineering & Remote Sensing*, 73 (4), 337–341.

Jensen, J.R., 2007. *Remote sensing of the environment: an earth resource perspective*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall.

Johnson, D.M., 2008. A comparison of coincident landsat-5 TM and resourcesat-1 AWiFS imagery for classifying croplands. *Photogrammetric Engineering & Remote Sensing*, 74 (11), 1413–1423.

Liknes, G., Perry, C., and Meneguzzo, D., 2010. Assessing tree cover in agricultural landscapes using high-resolution aerial imagery. *The Journal of Terrestrial Observation*, 2 (1), 38–55.

Liu, W., Gopal, S., and Woodcock, C.E., 2004. Uncertainty and confidence in land cover classification using a hybrid classifier approach. *Photogrammetric Engineering & Remote Sensing*, 70 (8), 963–971.

Maxwell, S.K., Meliker, J., and Goovaerts, P., 2010. Use of land surface remotely sensed satellite and airborne data for environmental exposure assessment in cancer research. *Journal of Exposure Science and Environmental Epidemiology*, 20, 176–185.

Mueller, R., 2000. Categorized mosaicked imagery from the National Agricultural Statistics Service Crop Acreage Estimation Program. *In*: *Proceedings of the ASPRS 2000 Conference*, ASPRS [Available on the CD], May 2000, Bethesda, MD.

Downloaded By: [George Mason University] At: 13:33 14 April 2011

Ozga, M. and Craig, M.E., 1995. PEDITOR – statistical image analysis for agriculture. *In*: *Presentation at the Washington Statistical Society (WSS) Seminar, USDA/NASS*, April 1995, Washington, DC.

Quinlan, J.R., 1996. Bagging, boosting, and C4.5. *In*: *Proceedings AAAI-96 fourteenth National Conference on Artificial Intelligence*, Portland, OR.

Seffrin, R., 2007. Evaluating the accuracy of 2005 multitemporal TM and AWiFS imagery for cropland classification of Nebraska. *In*: *Proceedings of the ASPRS 2007 Annual Conference*, 7–11 May 2007, Tampa, Florida.

Shan, J., *et al.*, 2010. Flood mapping with satellite images and its web service. *Photogrammetric Engineering & Remote Sensing*, 76 (2), 102–104.

Sun, W., *et al.*, 2008. Mapping plant functional types from MODIS data using multisource evidential reasoning. *Remote Sensing of Environment*, 112 (3), 1010–1024.

USGS, 2010. Available from: http://landsat.usgs.gov/products_data_at_no_charge.php [Accessed May 10 2010].

Downloaded By: [George Mason University] At: 13:33 14 April 2011