THE NASS CROPLAND DATA LAYER PROGRAM [*]

Michael E. Craig
National Agricultural Statistics Service, USDA
Fairfax, Virginia, USA

ABSTRACT

The USDA National Agricultural Statistics Service (NASS) entered a new era of data dissemination this spring with the release of products from its Cropland Data Layer Program. The CD-ROM products are public domain, and are available to the general public. The Cropland Data Layer is a crop specific digital data layer, suitable for direct use in commercial geographic information systems (GIS) applications. Currently, the Program encompasses seven states: Arkansas, Illinois, Indiana, Iowa, Mississippi, New Mexico, and North Dakota. Small area pilot demonstrations are underway in Nebraska, Missouri, and Florida. The output products combine categorized imagery from the Landsat V Thematic Mapper (TM) and the Landsat VII Enhanced Thematic Mapper (ETM+) sensors. A digital geo-referenced output mosaic for the entire state is created each year for five of the Program states. Only the eastern halves of Arkansas and New Mexico are included in their mosaics at this time. This Program represents a cooperative venture between three USDA Agencies (headquarters units of NASS, the Foreign Agriculture Service, and the Farm Service Agency) plus in-state agreements between NASS State Statistical Offices and their respective state government and/or university partners. The image files are in Erdas Imagine .GIS format; with an .IMG version also available as a WinZip file. Each CD-ROM contains the rasterized Cropland Data Layer of a particular state, categorization accuracy statistics by Analysis District, ancillary vector Shapefile data layers (Agricultural Statistics Districts, NASS area frame land use stratum, county boundaries, and major roads), a screen show of the Program's methodology, ArcExplorer (from ESRI) freeware software, and an ArcExplorer project that bundles up the raster and vector data sets together for display.

1.0 INTRODUCTION TO NASS, CROP ACREAGE, AND DATA LAYERS

One specific task of the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture is the production and distribution of official domestic crop acreage statistics. The NASS Agricultural Statistics Board (ASB) uses indications from all available sources, including farmer reported data from surveys, Farm Service Agency administrative data, agri-business reported data and the Census of Agriculture data (every five years) to generate the best numbers for each specific crop's acreage given all of the inputs. Only the official number is published; publication dates are standard and are known well in advance. For spring planted crops, the first crop acreage estimates are published at the

[*] Presented at the Third International Conference on Geospatial Information in Agriculture and Forestry, Denver, Colorado, 5-7 November 2001. A web page for information on this Program is: http://www.nass.usda.gov/research/Cropland/SARS1a.html .

end of March in the *Prospective Plantings Report*. Actual plantings are released in the *Acreage Report* at the end of June. Yield forecasts are published about July 10 for oats, barley, and spring wheat; and August 10 for corn, soybeans, sorghum, and cotton, along with forecasts of acres for harvest. The end-of-season summary of *Crop Production*, containing the final state estimates of crop acreage, is published in early January of the following year; county estimates are published starting in February.

The Cropland Data Layer Program is an offshoot of the Agency's Remote Sensing for Agricultural Estimation Program, which is an ongoing effort to reduce the statistical variance of acreage indications from farmer reported surveys. Satellite data categorized by crop is used as an auxiliary variable in a regression approach to refine state and county crop acreage estimates. Ground data for the categorization is based on a sample of land areas visited every year in June and enumerated with respect to crop acreage among other variables. Although more precise than just using the farmer reported ground survey, the indications generated using remote sensing inputs cannot meet the NASS/ASB time requirements of the crop estimates system, described above, that American agriculture has come to expect.

Early in the growing season, there would be considerable misclassification among different crops if acreage were estimated by remote sensing. Until crops reach nearly their full canopies, the satellite-based algorithms would be attempting to predict acreage based on soil types. Some studies have shown that the optimum time to accurately separate corn from soybeans in the upper Midwest is about mid-August; for rice and cotton the optimum date is in mid to late July. Combining early season imagery with mid-to-late season imagery increases the classification accuracy also. However, NASS has published three major crop acreage indications (prospective plantings, actual plantings, and acreage for harvest) before satellite imagery could be accurately used. Since county crop acreage estimates are not released until early in the next calendar year (following the crop season), remote sensing based acreage indications are being used in Program states for county estimates and can be used for other small area statistics (such as for watersheds).

One by-product of the remote sensing estimation process is satellite imagery categorized into crop specific classes. Originally these categorized images were used only in displays or for demonstration purposes, but were not considered a marketable product by themselves. However, the rapid improvement of image processing and geographic information systems software capabilities, and the release of EarthSat's GeoCover products, led to the development of mosaics containing many scenes stitched together to cover large areas, up to entire states in some cases. This new product was seen as very useful to many existing Agency customers and has generated new customers in related fields. For example, one firm used this product to help place the construction of a wheat straw processing plant, another plans to use it to monitor crops by watershed, another to locate migratory bird landing areas, and for state water quality evaluation.

2.0 SPECIFICATIONS OF THE CROPLAND DATA LAYER PRODUCT

The Cropland Data Layer for a state is a large area (up to the entire state) precision registered mosaic merging categorizations of multiple separate satellite images into one digital file. Since 1999, only Landsat V Thematic Mapper or Landsat VII Enhanced Thematic Mapper imagery was used, giving the data layer a 30-meter resolution. SPOT Multi-Spectral Scanner and Indian Resource Satellite LISS III imagery could easily be used to fill in for areas cloud covered in Landsat overpasses for the estimation process; however, these images would have to be re-projected into 30 meters to fit the mosiac process now in place. The data layer is produced annually once a State enters the Program. Currently, full Program states include North Dakota, Illinois, Indiana, Iowa, and Mississippi. The eastern halves of

Arkansas and New Mexico are also included in the Program. Indiana and Iowa both entered the Program in 2000, so 1999 data layers are not available for them. See Figure 1 for a view of the CDL's for both Arkansas and Mississippi. Additionally, pilot areas are being processed for the boot heel area of Missouri and south eastern Nebraska in 2001.

The current data layer for each Program state is distributed on a CD-ROM, with the previous year data layer included when available. The major content of each CD-ROM is the raster Cropland Data Layer in ERDAS Imagine ♣GIS♠ format. When space allows, compressed (using the WinZip utility) ERDAS ♣IMG♠ format images of the data layer are included. Ancillary vector spatial data layers are available each CD-ROM, they include the Agricultural Statistics Districts (ASD) and County boundaries, major roads, and the Agency's Area Frame Land Use Stratification for the chosen state. See Figure 2 for a close up of the Arkansas and Mississippi Area Sampling Frames. All of the vector data layers are provided in ESRI's Shapefile format. Also included are categorization accuracy statistics by Analysis District (i.e. by date combinations of the input satellite imagery), a screen show of the Program methodology, ESRI♣s ArcExplorer freeware software, and an ArcExplorer project that bundles up the raster and vector datasets together for easy display in ArcExplorer. Users are under NO copyright restrictions with either the NASS categorized imagery or the freeware ESRI ArcExplorer. For more information on ESRI and ArcExplorer: see http://www.esri.com . The NASS Cropland Data Layer imagery is considered to be in the public domain and FREE to redistribute.

Using the NASS in-house image processing system, PEDITOR, the categorized images were co-registered, using automated block correlation techniques, to the ortho-rectified GeoCover Stock Mosaic images produced by Earth Satellite Company, Incorporated (EarthSat); the GeoCover website is: http://www.geocover.com/ . The EarthSat images were chosen as they currently provide the best available large area ortho-rectified images as a basis to register large volume Landsat images. The GeoCover Stock Mosaics are rated to have within a 50 meter root mean squared error overall. The Cropland Data Layer mosaics are projected into the following UTM zones: New Mexico (13), North Dakota (14), Arkansas (15), Iowa (15), Mississippi (16), Illinois (16), and Indiana (16). The spheroid and datum are based on the WGS84 coordinate system.

Each CD-ROM contains classification accuracy metadata by Analysis District (unique imagery date combinations) in .html format. Figure 3 shows an example of one Analysis District's metadata for Arkansas 2000. The metadata contains the Analysis District imagery date coverage, sensors used, percent correct and kappa coefficients, regression analysis by Analysis District, the sampling frame scheme, and the original number of cover type signatures, all of which are provided in browser .html format. The metadata should be studied to determine the strong and weak points/areas of any given classification mosaic. Classification accuracy can be limited by many factors, including: different planting dates of the same crop, differences between healthy plants and plants of the same crop but under serious stress, the density of crop planting and the presence or absence of weeds and/or water. On top of all the other factors, the atmosphere (clear versus haze or clouds) through which the crop response is being measured is not the same from one day to the next, or from one part of the image to the next.

3.0 A SUMMARY OF METHODOLOGY AND BACKGROUND

As mentioned earlier, the Cropland Data Layer Program is derived from the Agency's Remote Sensing Acreage Estimation Program. Expansion of the acreage estimation effort was positively influenced by several factors in late 1996 and early 1997. Using remote sensing based estimators at the county level became more important after a US farm policy change in the 1996 farm bill that led to a

reduction in small area (county) administrative data on crop acreage. Figures 3 and 4 show two years of CDL data for a selected county in the Delta Region of Arkansas. The major factor however, was an agreement reached between NASS and the Production Estimates and Crop Assessment Division (PECAD) of the USDA's Foreign Agriculture Service (FAS) to provide NASS with Landsat TM data from their current year's purchases. This agreement greatly reduced the cost per scene of Landsat data to NASS, and also brought new licensing regulations which allowed program expansion based on hardware and analyst resources instead of on availability of sufficient satellite imagery. The Cropland Data Layer was originally designed as a vehicle to attract external (to NASS) users interested in sharing resources for the Acreage Estimation Program, but it quickly became a viable product by itself. Partners for sharing were found with state government groups, other federal agencies, and universities.

This paper did not attempt to cover in detail several topics associated with the Cropland Data Layer Program. In general, current statistics and information on crop estimation procedures can be found on the Agency♣s website at http://www.usda.gov/nass . The NASS crop acreage estimation system is covered by Vogel (1999) and Allen (2001). The references listed below should aid in covering these topics. Data Layer methodology concerns are addressed in more detail in Mueller (2000), Craig (2001), and Allen (2001). The resource sharing partnerships and cooperative agreements necessary to maintain this Program are listed and discussed in Craig (2001) and Hanuschak (2001). The economics of today♣s Program versus those attempted in the past are analyzed in Hanuschak (2001). Finally, the Cropland Data Layer Program has its own webpage containing links to Product Examples, Frequently Asked Questions, Order Forms, Disclaimer, and Program Methodology. The webpage specifically for the CDL is http://www.nass.usda.gov/research/Cropland/SARS1a.html .

4.0 REFERENCES

R.A. Allen, G.A. Hanuschak, M.E. Craig. ■Remote Sensing for Crop Acreage Estimates.• White paper available from the USDA National Agricultural Statistics Service, Washington, D.C., April 2001.

M. E. Craig, ■A Resource Sharing Approach to Crop Identification and Estimation.• In *Proceedings of the ASPRS 2001 Conference*, St. Louis, MO, pp. xxxx, April 2001.

G.A. Hanuschak, R.C. Hale, M.E. Craig, R.W. Mueller, and G. Hart. ■The New Economics of Remote Sensing for Agricultural Statistics in the United States. In *Proceedings of the CAESAR Conference*, Volume2, Part 2, Paper#XXII, Italian Statistics Agency (ISTAT), Rome, Italy, June 2001.

R.W. Mueller, ■Categorized Mosaicked Imagery from the National Agricultural Statistics Service Crop Acreage Estimation Program.• In *Proceedings of the ASPRS 2000 Conference*, Washington, DC, pp. Xxxx, May 2000.

F.A. Vogel, and G. Bange. ■Understanding USDA Crop Forecasts.• NASS/WAOB Miscellaneous Publication No. 1554, U.S. Department of Agriculture, Washington DC-USA, 1999.

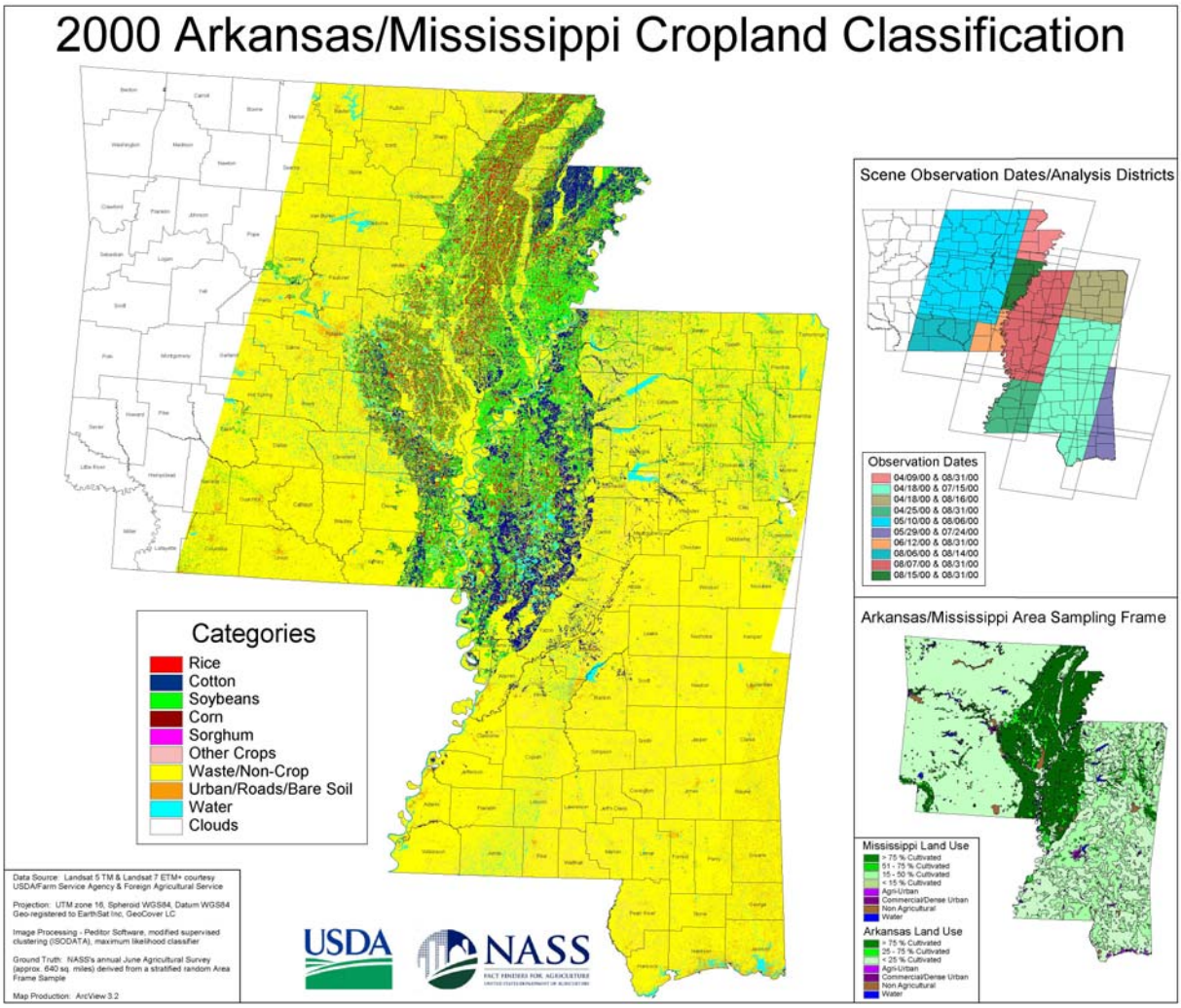

Figure 1. Cropland Data Layers for Arkansas and Mississippi in 2000.

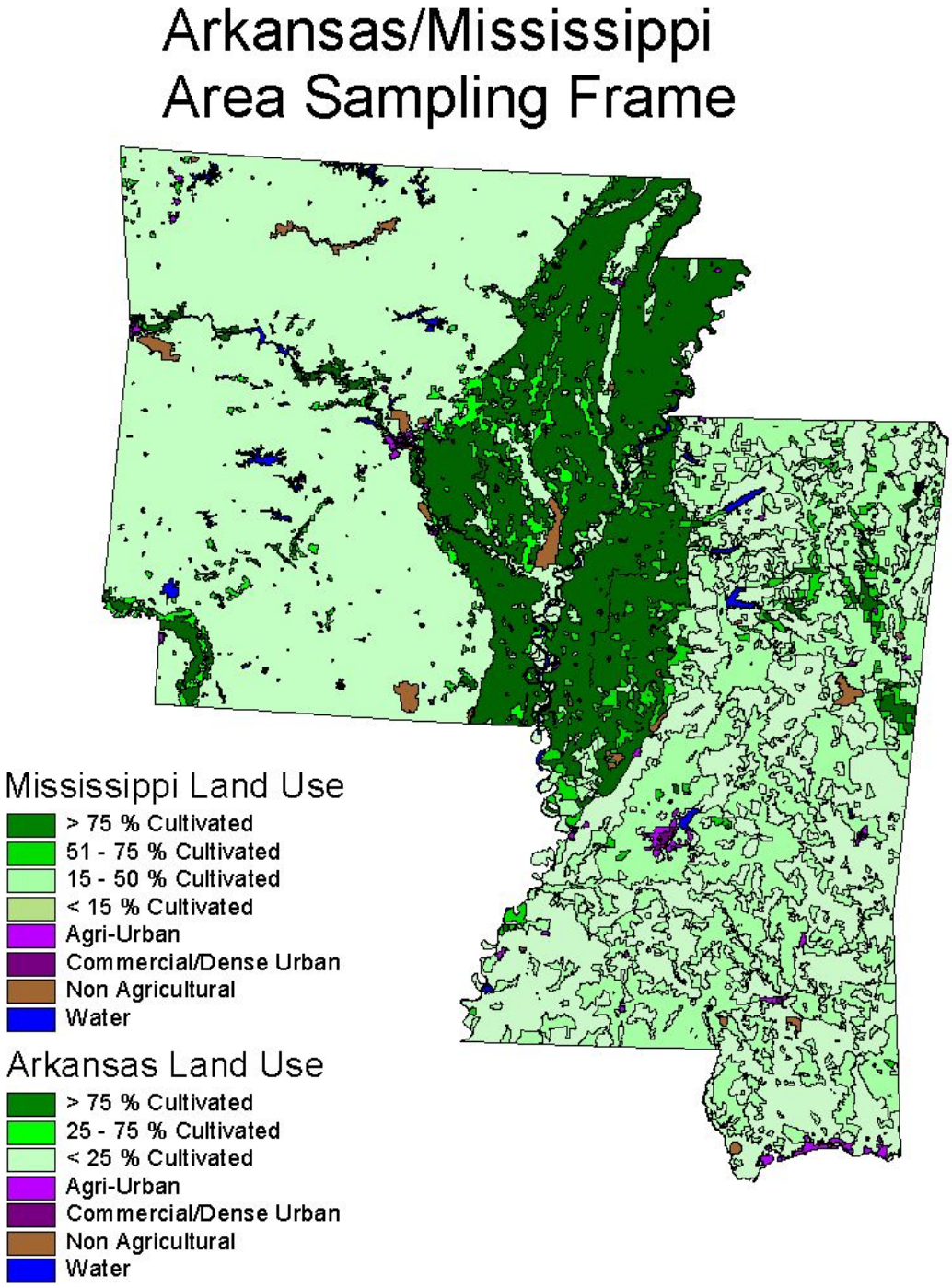

Figure 2. Area Frame Stratification for Arkansas and Mississippi

ANALYSIS DISTRICT AD04 - DATES = 5/10/00 & 8/06/00
LANDSAT 5 TM & 7 ETM+ - PATH: 24, ROWS: 35,36; 14 CHANNELS
164 CROP / COVER TYPE SIGNATURES

Mosaic Category	Crop / Cover *	Original # Categories	Original # Pixels	Percent Correct*	Commission Error (%)	Kappa Coefficient
1	CORN	8	3899	95.31	5.61	95.17
2	COTTON	13	8188	95.77	17.88	95.46
69	FESCUE SEED	5	388	94.07	57.95	94.03
81	FILLER	1	0	0.00	0.00	0.00
61	IDLE CROP	31	3270	95.90	53.75	95.69
62	NON-AGRIC	17	19440	76.23	15.80	72.79
62	OTHER	1	0	0.00	0.00	0.00
25	OTHER HAY	9	755	94.17	24.76	94.13
62	PERM PAST	9	1771	82.27	42.95	81.94
3	RICE	10	39756	95.16	5.44	93.20
4	SORGHUM	10	901	96.00	41.79	95.96
5	SOYBEANS	13	52843	82.93	3.49	74.67
62	UNKNOWN	0	221	0.00	0.00	0.00
82	URBAN	5	0	0.00	0.00	0.00
83	WATER	7	0	0.00	0.00	0.00
67	WATERMELONS	0	14	0.00	0.00	0.00
24	WIN WHEAT(STUBBLE)	21	2641	96.44	35.91	96.34
63	WOODS	4	5110	54.05	47.88	52.23
_	OVERALL *	164	139197	86.12	_	81.93

*** NOTE: Signatures for covers such as CLOUDS, URBAN and WATER were determined from pixels outside of the original ground truth sample; these cover types will have '0.00'**

Figure 3. Classification Accuracy Statistics Example.

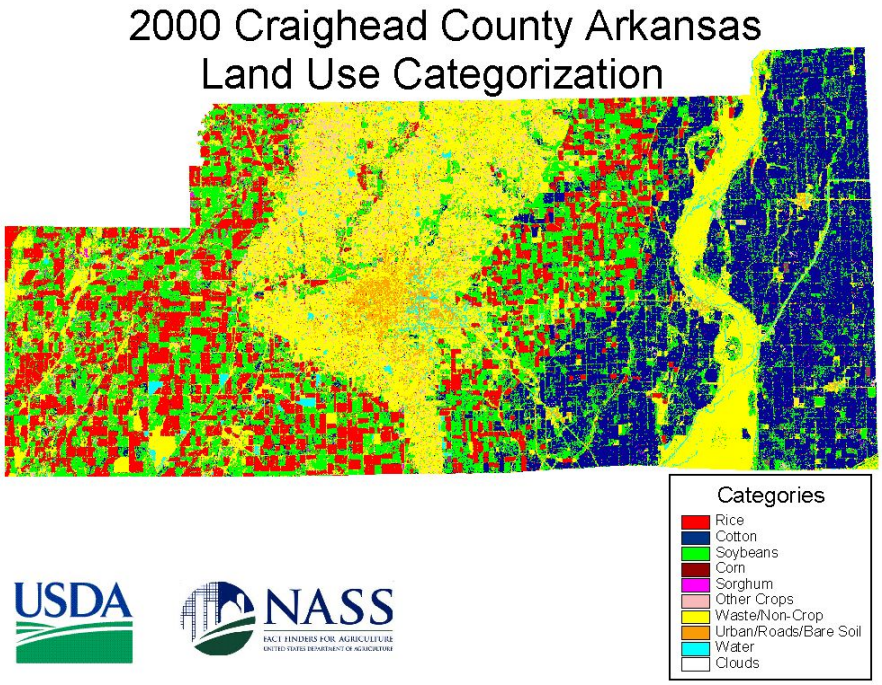

Figure 4. A Closer Look at the 2000 CDL in Arkansas

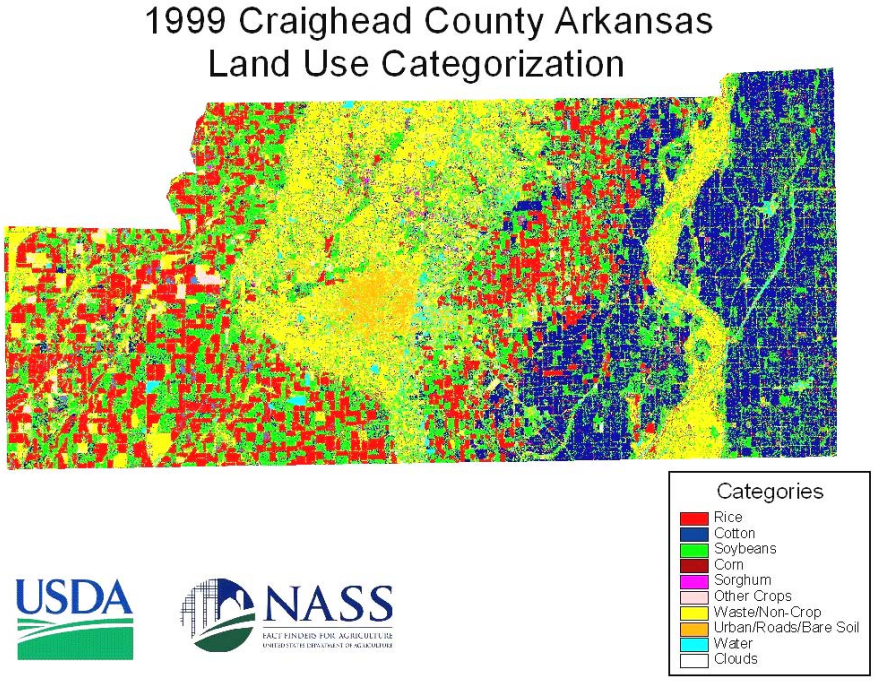

Figure 5. A Closer Look at the 1999 CDL for Arkansas.