

Crop Specific Covariate Data based on the NASS Cropland Data Layer for Area Frame Stratification

Claire G. Boryan, Geographer
USDA\NASS\Research and Development Division
Fairfax, VA



“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”



Background

- NASS provides timely, accurate, and useful statistics in service to U.S. agriculture

The screenshot shows the NASS Data and Statistics website. The main navigation bar includes Home, About NASS, Newsroom, Publications, Data and Statistics, Census, Surveys, and News. The 'Data and Statistics' section is highlighted. Below the navigation, there are search options and a 'Quick Stats' section. The 'Quick Stats' section includes a link to 'Query our Quick Stats Data Base' and a description of the data available. The 'Census of Agriculture' section provides information on how to query census data. The 'Interactive Data' section offers tools for interacting with census data, including 'Interactive Census Maps for 2002 Census Highlights' and 'Table Lens Application for 1997 Census Data'. The footer indicates the last modified date as 11/21/05.

2001 Wildlife Damage Survey

7.7 Percent of Crop Value Lost to Deer and Geese

lost \$17.2 million of corn, soybeans and wheat to deer or geese during 2001 and farmers losing 7.7 percent of the crop value for deer and geese. Soybean economic loss, totaling \$9.1 million, 11 percent. Corn losses were \$6.6 million, 5.6 percent. Deer damage resulted in losses of \$13.6 million, 6.1 percent, while geese, 1.6 percent.

stated 6.0 million bushels. Corn losses were 3.2 million bushels, soybean loss and wheat accounted for 0.6 million bushels. Production losses to deer were 4.1.3 million bushels.

losses to deer were most severe in Central and Western Maryland, while geese damage was most severe in the Eastern Shore region. The Eastern Shore received the highest number of deer damage reports. Corn losses were 3.2 million bushels, soybean loss and wheat accounted for 0.6 million bushels. Production losses to deer were 4.1.3 million bushels.

USDA NEWS RELEASE
NATIONAL AGRICULTURAL STATISTICS SERVICE
United States Department of Agriculture • Washington, DC 20250
Ag Statistics Hotline: (800) 853-8763

USDA FORECASTS RECORD

Washington, Aug. 10, 2007 - U.S. farmers' history in 2007, according to the Crop Profitability of Agriculture's National Agricultural Statistics Service, is 13.1 billion bushels, 19.6 percent above the 11.6 billion bushels of 2006. Based on conditions as of August 1, 2007, the 2007 crop production is expected to be 13.1 billion bushels, 19.6 percent above the 11.6 billion bushels of 2006.

WISCONSIN AGRICULTURAL STATISTICS SERVICE
P.O. Box 8034 Madison, WI 53708-8034
In cooperation with WI Department of Agriculture, Trade and Consumer Protection

2002 Dairy Producer Opinion Survey
November 2002

Wisconsin Milk Production to Recover

Milk production is expected to increase in Wisconsin during the next five years according to a survey conducted by the Wisconsin Agricultural Statistics Service. This statewide survey of producers asked for their plans with the assumption that milk prices for the next five years will be at the same level as the past five years. The survey was conducted during May and June.

Based on the survey, 60 percent of producers expect to keep the same herd size, 20 percent plan to increase herd size, and 20 percent intend to discontinue milking by 2007. Actual results will depend on future milk prices, input prices, financing availability, crop yields, and other factors.

The number of herds projected for 2007 shows that the diversity of small to large herds will continue. The most prevalent herd size will remain at 50 to 99 cows.

The screenshot shows the NASS 2002 Census of Agriculture - 500 Interactive Mapping website. The main navigation bar includes Home, About NASS, Newsroom, Publications, Data and Statistics, Census, Surveys, and News. The 'Data and Statistics' section is highlighted. Below the navigation, there are search options and a 'Quick Stats' section. The 'Quick Stats' section includes a link to 'Query our Quick Stats Data Base' and a description of the data available. The 'Census of Agriculture' section provides information on how to query census data. The 'Interactive Data' section offers tools for interacting with census data, including 'Interactive Census Maps for 2002 Census Highlights' and 'Table Lens Application for 1997 Census Data'. The footer indicates the last modified date as 11/21/05.

All Milk Price, Wisconsin Annual Average, 1985 - 2002

The chart shows a line graph of the annual average milk price in Wisconsin from 1985 to 2002. The price fluctuates significantly, with a notable peak around 1995 and a low around 2002.

Wisconsin Dairy Herds by Herd Size

Milk cow herd size	May 2002 herds	May 2007 herds (projected) (%)	Change 2007/2002
1 - 29	2,800	1,440	-48
30 - 49	4,700	2,440	-27
50 - 99	7,400	5,600	-24
100 - 199	1,800	2,080	+9
200 - 499	700	900	+29
500+	200	440	+120
Total	17,500	19,900	+20

17% the May 2007 projection is based on farmers' opinions May-June 2002, with the assumption that milk prices for the next five years will be at the same level as the past five years.

Percent of Herds by Size Group 2007 Projection

The pie chart shows the projected distribution of dairy herds by size group in 2007. The largest group is 100-199 cows, followed by 50-99 cows, 200-499 cows, 1-29 cows, 30-49 cows, and 500+ cows.

Applications of the Cropland Data Layer within NASS

- **Crop Acreage Estimation (1997-2012)**
- Area Sampling Frame Stratification (2010-2013)
- Area Sampling Frame Substratification (2013)



“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”



What are Area Sampling Frames?

- NASS Area Sampling Frames been used as the **primary tool** to conduct agricultural surveys since 1954.
- The NASS Area Sampling Frames are the basis for the annual **June Area Survey** in which approximately 11,000 segments are enumerated in early June to collect crop acreage and other agricultural information.
- NASS ASFs are based on a **stratification of land cover** in the U.S. defined by percent cultivated cropland.



“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”



What is the stratification?

- The word “**stratify**” comes from the Latin words meaning “to make layers”; we divide the population into subpopulations, called strata.
- In statistics, the goal is to make the strata as **homogeneous** as possible.
- Stratified sampling generally gives more precise (**lower variance**) estimates for population means and totals than simple random sampling alone.



“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”



How Is Stratification Performed at NASS?

- It has been conducted by Area Frame staff since 1954 using **visual interpretation** of aerial photography, and later moderate resolution Landsat TM data.
- The NASS Cropland Data Layer products have been used in recent years to aid in the visual interpretation process.
- In the past two years **Cropland Data Layer (CDL) automated stratification** has begun to be implemented.

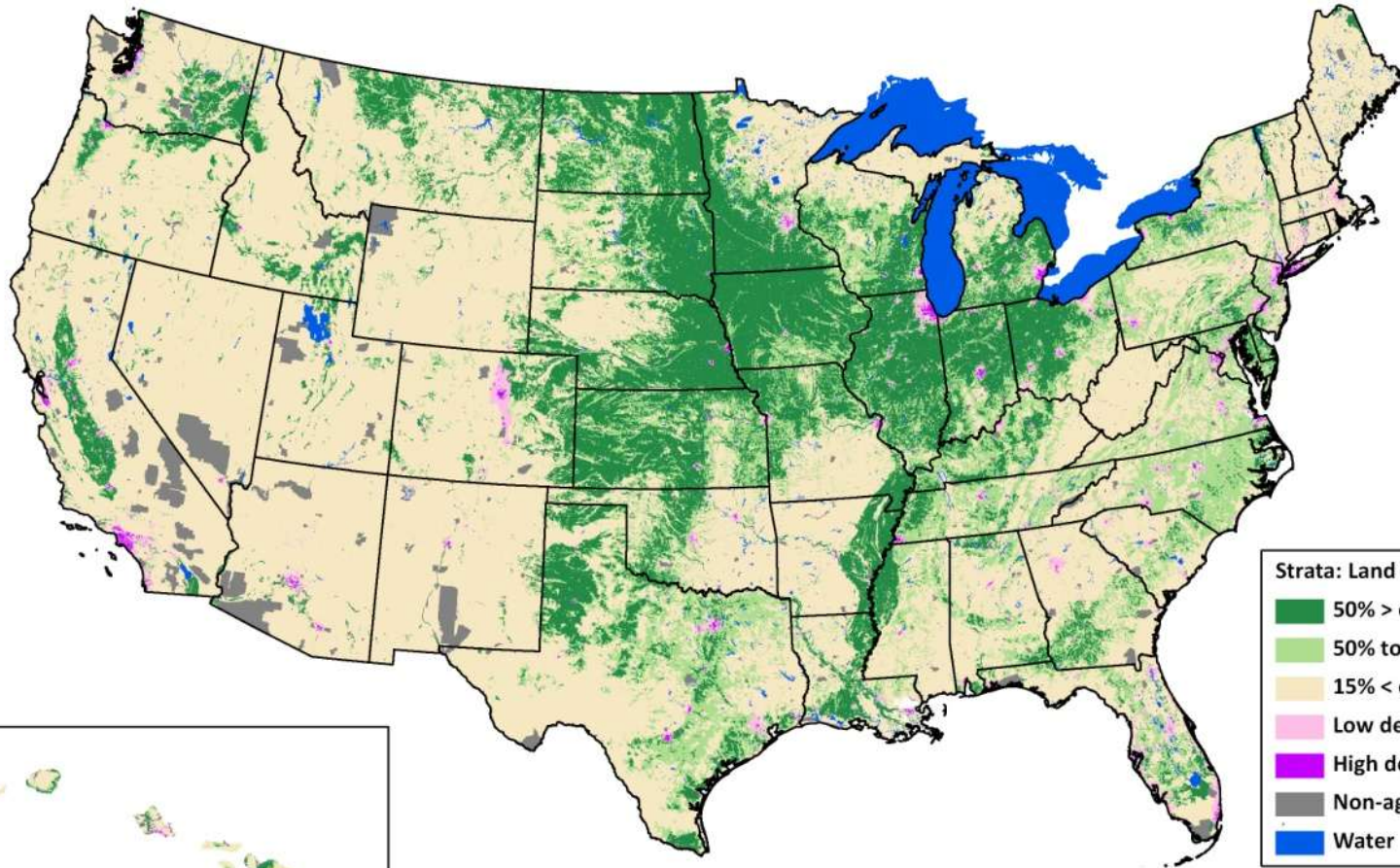


“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”



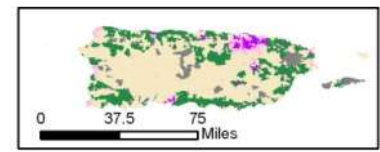
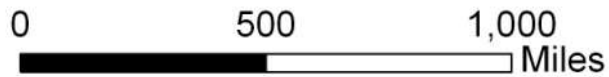
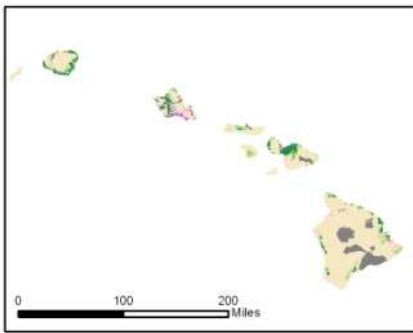


National Agricultural Statistics Service Land Use Area Frame



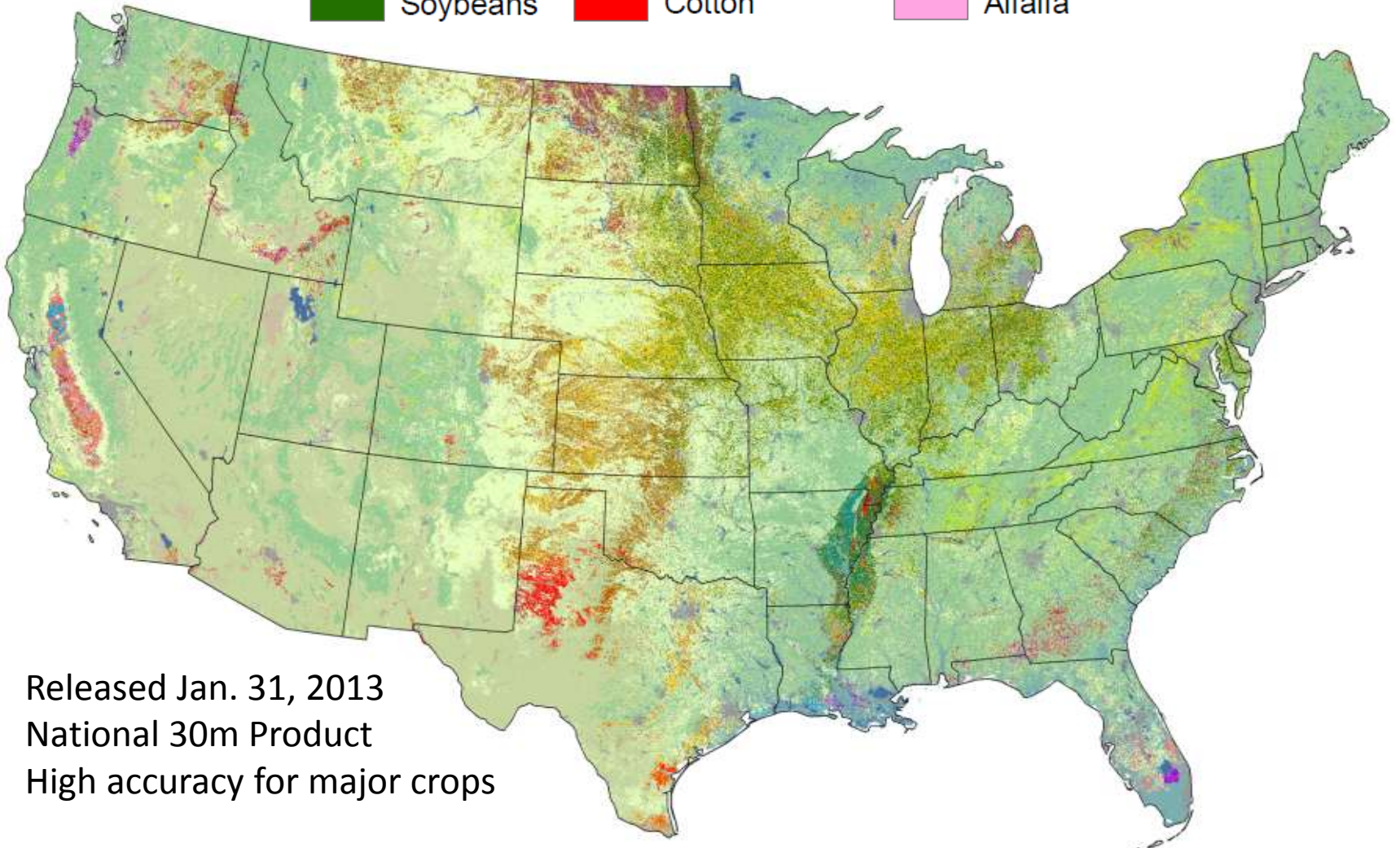
Strata: Land Use

- 50% > cultivation
- 50% to 15% cultivation
- 15% < cultivation
- Low density urban
- High density urban
- Non-agriculture
- Water



What is the Cropland Data Layer (CDL)?

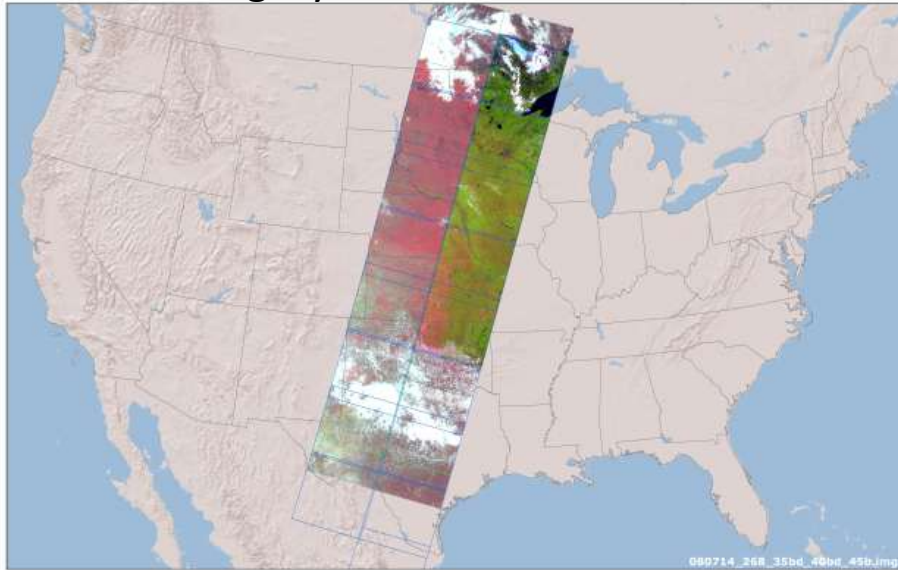
The Cropland Data Layer product is a raster-formatted, crop specific, land cover data set used to generate in-season acreage estimates for 19 crops in 41 states.



Released Jan. 31, 2013
National 30m Product
High accuracy for major crops

2012 Cropland Data Layer Inputs

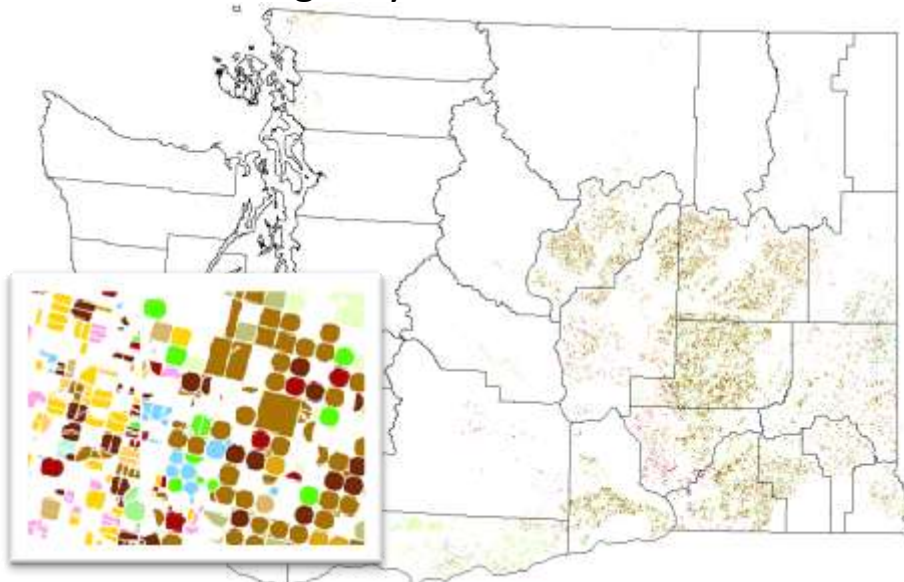
Satellite Imagery – Landsat



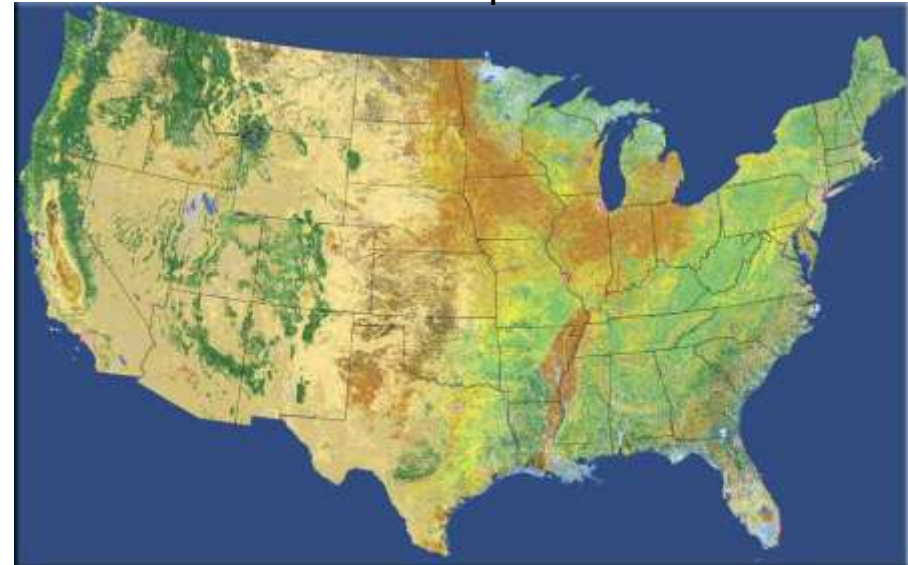
Satellite Imagery – Deimos & UK2



Farm Service Agency: Common Land Unit



2006 NLCD & Derivative products









2011 Continental United States Land Cover Categories (by decreasing acreage)

National Cropland Data Layer Legend

Agriculture

 Pasture/Grass	 Triticale	 Dbl Crop Lettuce/Cotton
 Corn	 Citrus	 Mustard
 Soybeans	 Safflower	 Plums
 Winter Wheat	 Pistachios	 Dbl Crop Barley/Sorghum
 Fallow/Idle Cropland	 Blueberries	 Broccoli
 Other Hay/Non Alfalfa	 Christmas Trees	 Radishes
 Alfalfa	 Dbl Crop Barley/Soybeans	 Garlic
 Cotton	 Tomatoes	 Speltz
 Spring Wheat	 Onions	 Vetch
 Sorghum	 Flaxseed	 Apricots
 Dbl Crop WinWht/Soybeans	 Dbl Crop Oats/Corn	 Caneberries
 Rice	 Pop or Orn Corn	 Greens
 Barley	 Herbs	 Nectarines
 Oranges	 Misc Veggies & Fruits	 Cucumbers
 Oats	 Olives	 Other Small Grains
 Sunflower	 Other Tree Crops	 Turnips
 Dry Beans	 Dbl Crop Corn/Soybeans	 Dbl Crop Lettuce/Cantaloupe
 Peanuts	 Sweet Potatoes	 Camelina
 Durum Wheat	 Peaches	 Cauliflower
 Sugarbeets	 Cranberries	 Rape Seed
 Potatoes	 Tobacco	 Honeydew Melons
 Canola	 Cantaloupes	 Celery
 Sugarcane	 Prunes	 Dbl Crop Durum Wht/Sorghum
 Almonds	 Dbl Crop Barley/Corn	 Eggplants
 Sod/Grass Seed	 Dbl Crop Soybeans/Cotton	 Gourds
 Grapes	 Pears	 Dbl Crop Lettuce/Barley
 Apples	 Lettuce	
 Rye	 Dbl Crop Lettuce/Durum Wht	
 Peas	 Watermelons	
 Millet	 Switchgrass	
 Walnuts	 Asparagus	
 Lentils	 Carrots	
 Pecans	 Strawberries	
 Dbl Crop WinWht/Cotton	 Pumpkins	
 Dbl Crop WinWht/Sorghum	 Squash	
 Sweet Corn	 Cabbage	
 Aquaculture	 Peppers	
 Clover/Wildflowers	 Dbl Crop Soybeans/Oats	
 Other Crops	 Hops	
 Dbl Crop WinWht/Com	 Mint	
 Cherries	 Pomegranates	

Non-Agriculture

 Forest
 Shrubland
 Developed
 Wetlands
 Water
 Barren
 Perennial Ice/Snow

CropScape

USDA United States Department of Agriculture
National Agricultural Statistics Service

CropScape - Cropland Data Layer

Layers Legend

- Background Layers
- Cropland Data Layers
 - 2012
 - 2011
 - 2010
 - 2009
 - 2008
 - 2007
 - 2006
 - 2005
 - 2004
 - 2003
 - 2002
 - 2001
 - 2000
 - 1999
 - 1998
 - 1997
- Crop Mask Layer
- Boundary Layers
- Water Layers
- Road Layers

2860306.28345, 3792995.44022

<http://nassgeodata.gmu.edu/CropScape>

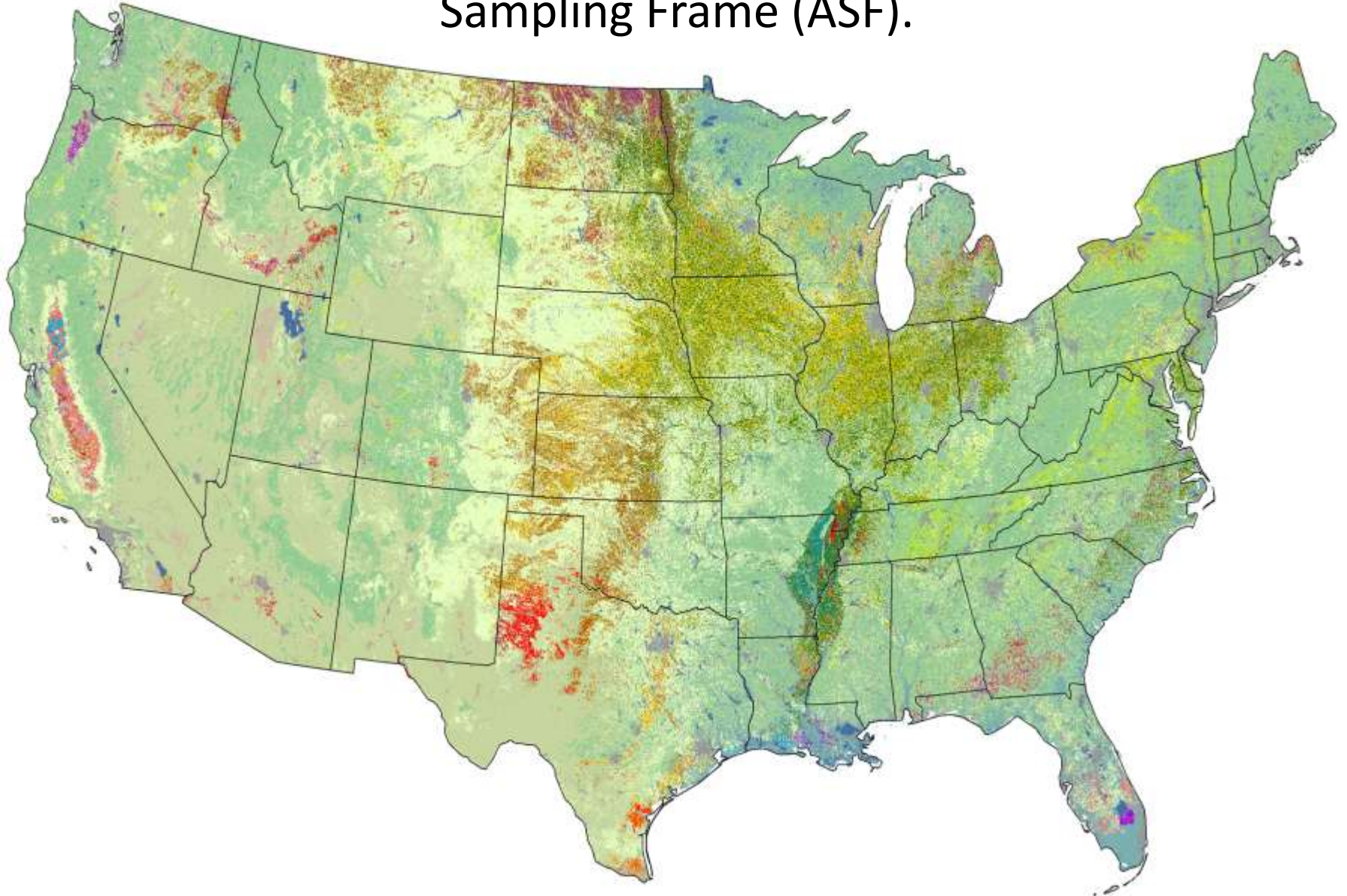
USDA Home | NASS Home | Research and Development Division | About CDL | Citation | Contact
Copyright © General Science and Technology Solutions Inc., 2009 - 2014

f t in

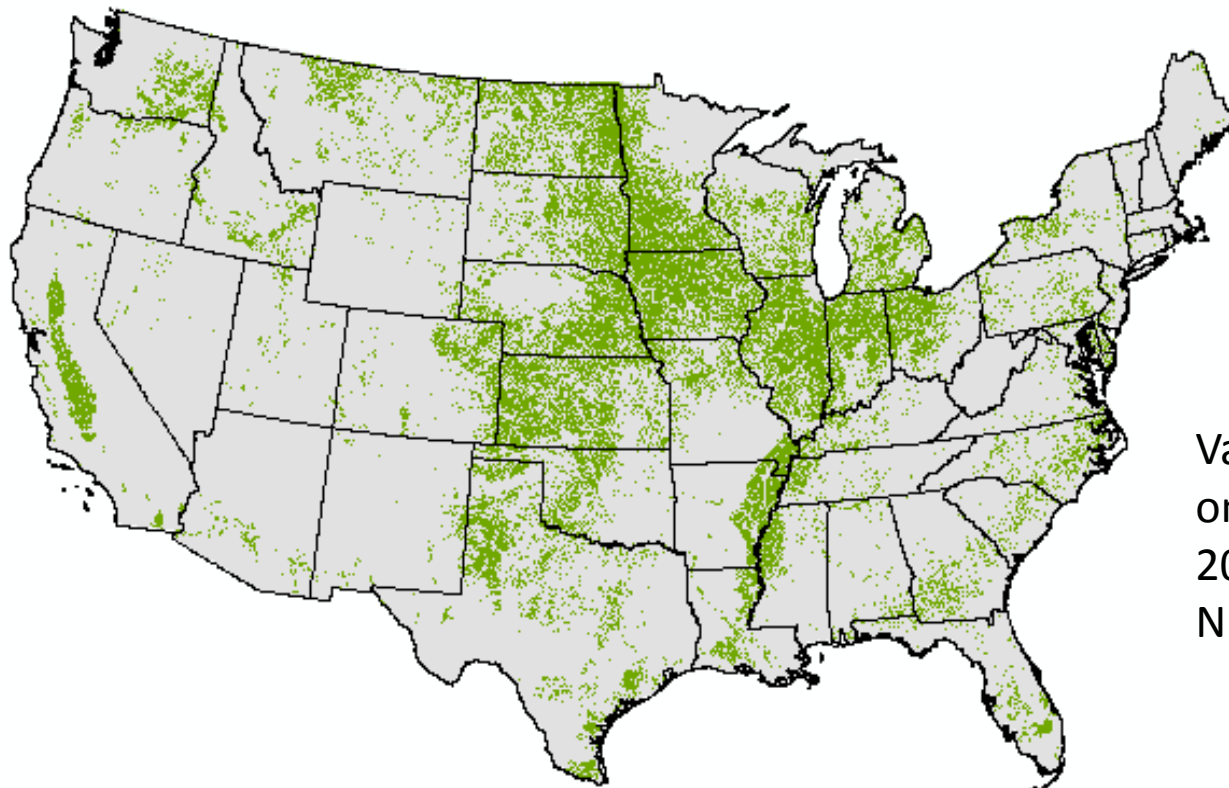
Interactive CDL visualization/exploration/dissemination portal



Research is being conducted to develop new methods and data sets to utilize the CDL in the construction of the NASS Area Sampling Frame (ASF).



The NASS 2012 Cultivated Layer created from 2008-2012 CDL data



Validation based on independent, 2012, FSA CLU and NLCD, 2006 data

2012

Class Name	Code	Producer Accuracy	Omission Error	Kappa	User Accuracy	Commission Error	Conditional Kappa
Non-Cultivated	1	96.6%	3.4%	0.940	97.2%	2.8%	0.951
Cultivated	2	97.9%	2.1%	0.951	97.4%	2.6%	0.940

Metadata and Download - <http://www.nass.usda.gov/research/Cropland/Release/>

Applications of the Cropland Data Layer within NASS

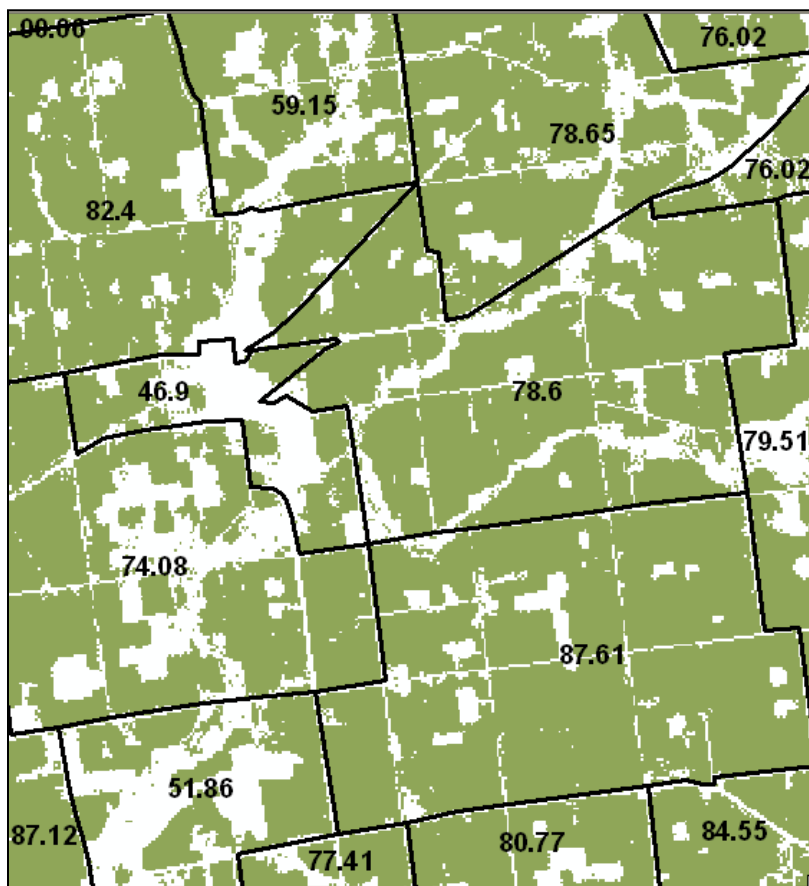
- Crop Acreage Estimation (1997-2012)
- **Area Sampling Frame Stratification (2010-2013)**
- Area Sampling Frame Substratification (2013)



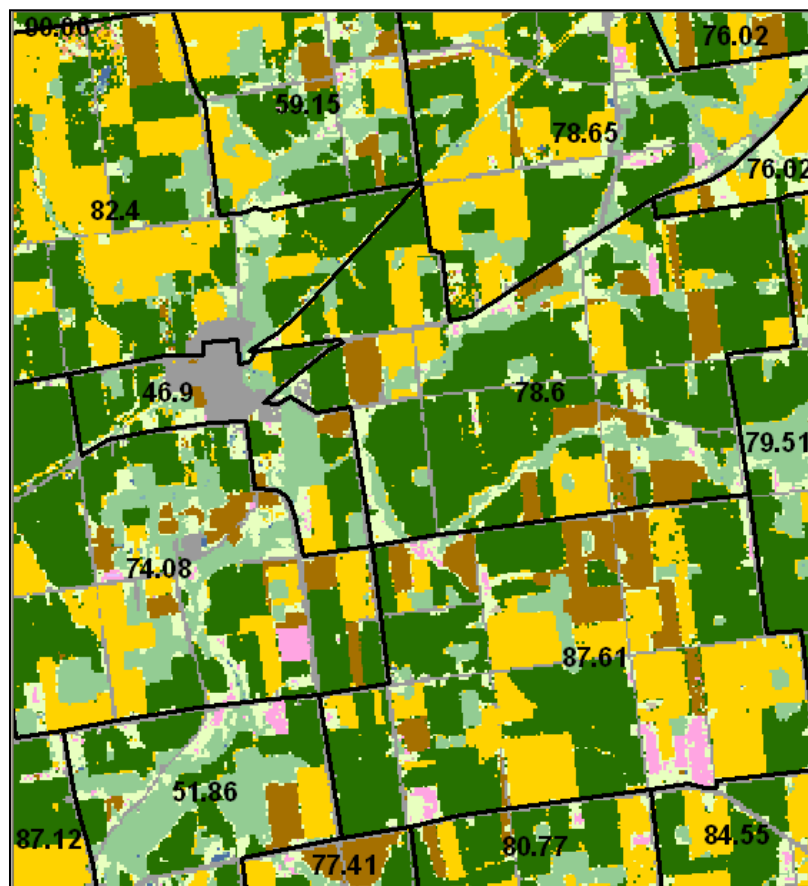
“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”



Automated Stratification of the NASS Area Sampling Frame based on the CDL

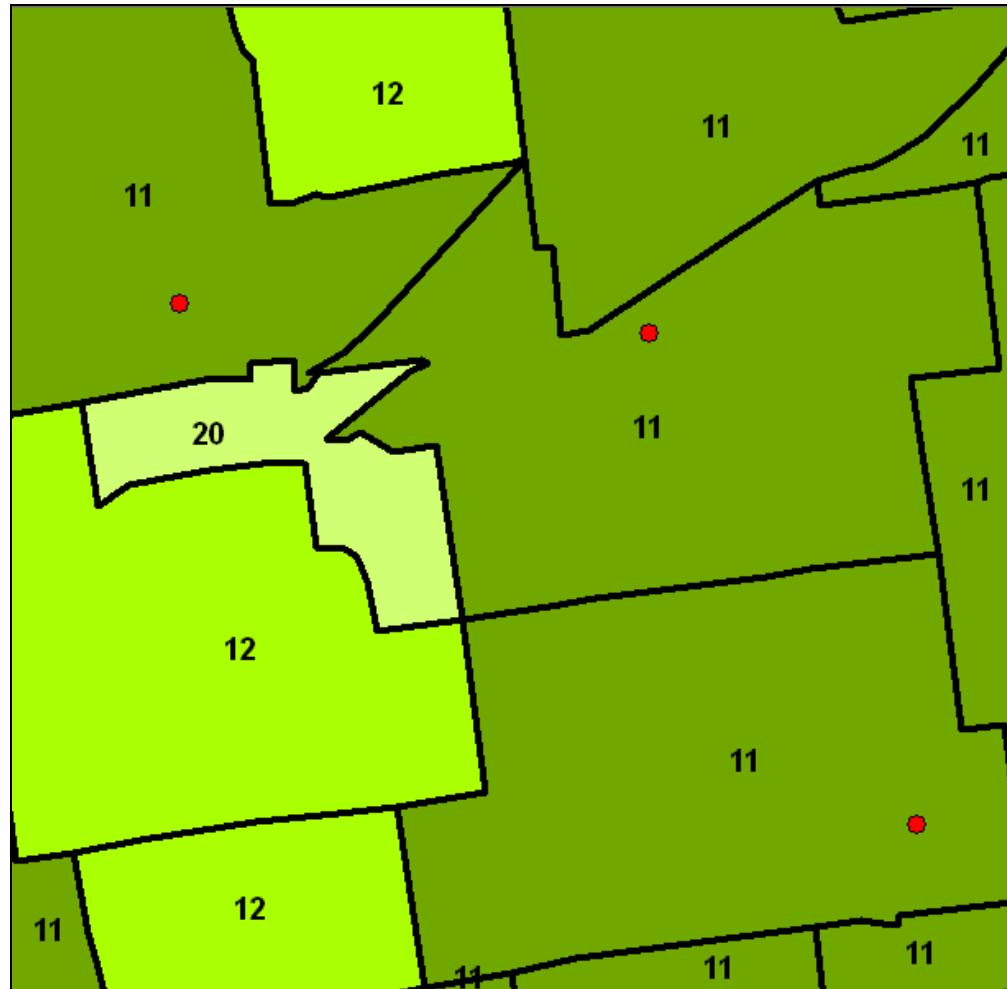


Primary Sampling Units with CDL percent cultivation



Primary Sampling Units with CDL percent cultivation, overlaying a 2010 CDL image product

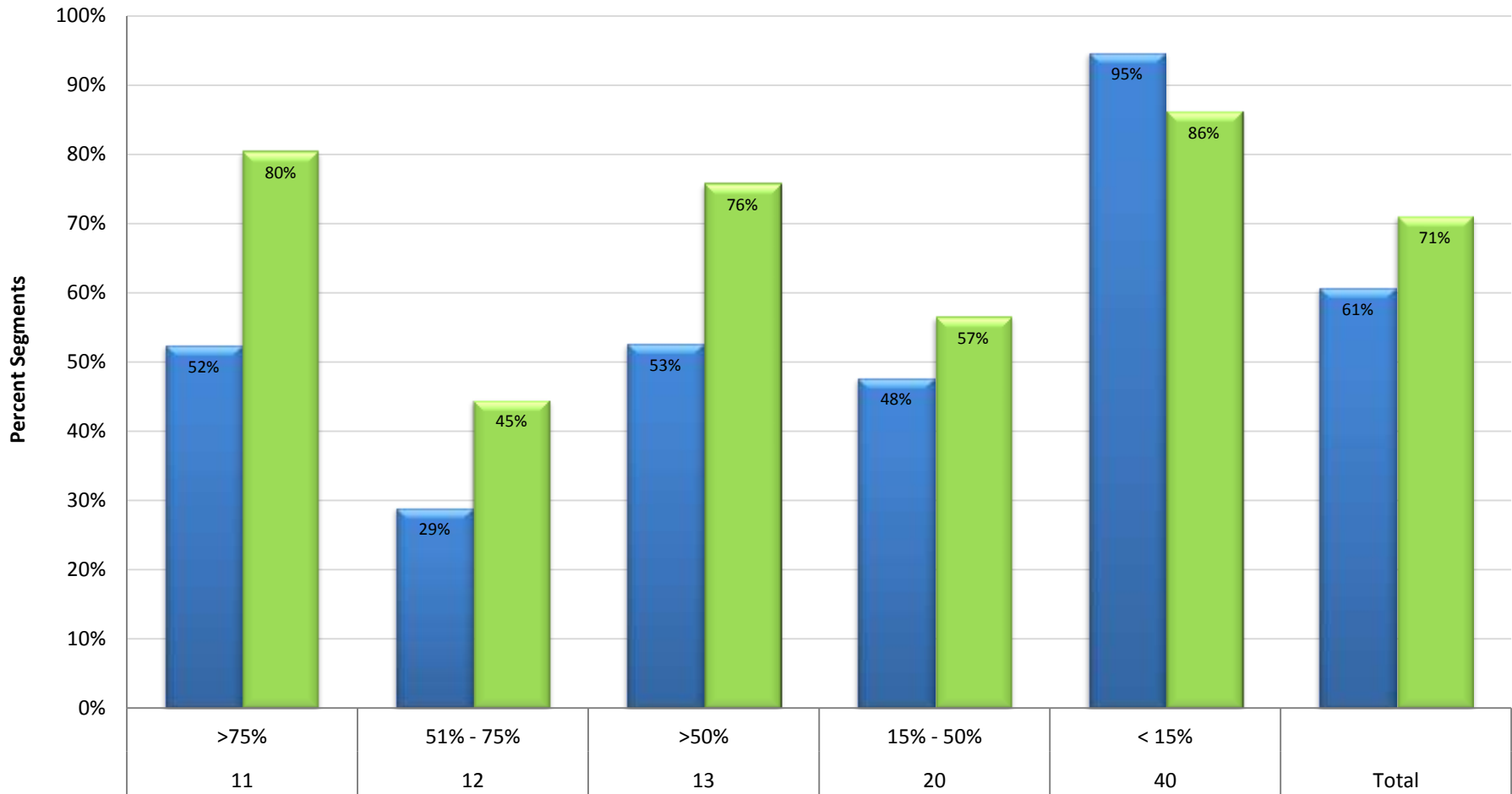
CDL based stratification of a NASS Area Sampling Frame (ASF)



Area Frame (Traditional) vs CDL Stratification

Strata Summary 2010

■ AREA FRAME Stratification ■ CDL Stratification



Stratum & Percent Cultivation

Accuracy based on 2010 June Area Survey in situ data

New state Area Sampling frames created using the CDL automated stratification method

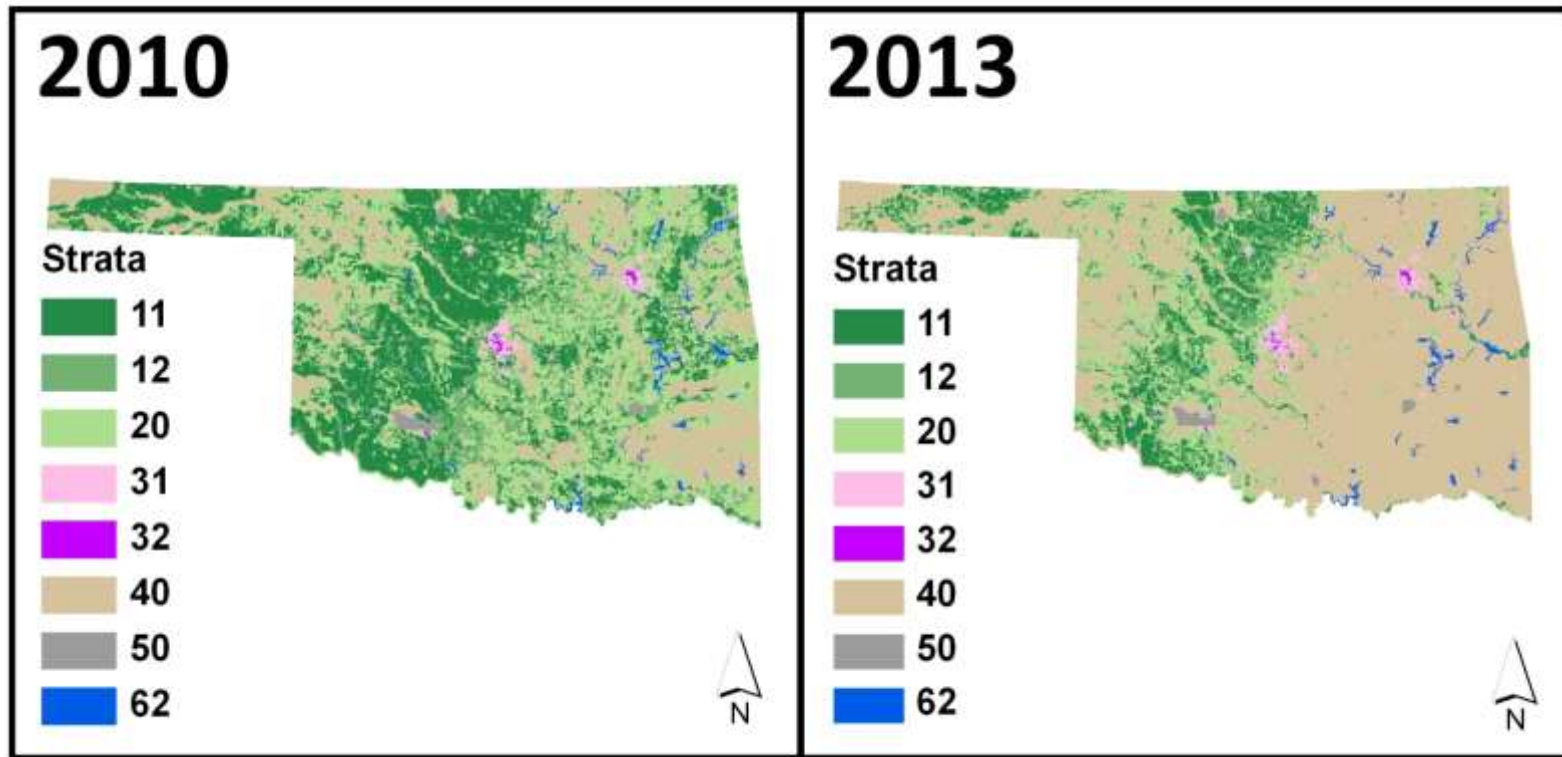
- *Arizona - 2014*
- *Georgia - 2014*
- *New Mexico - 2014*
- *Oklahoma - 2013*
- *South Dakota - 2014*
- *North Carolina - 2014*



“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”



CDL Automated Stratification in NASS Operations



The Oklahoma Area Sampling Frames (2010 and 2013). Stratum 11 (>75% cultivated) was overestimated in the 2010 ASF which was created using the traditional method and updated to more accurately reflect conditions in the 2013 ASF using the CDL automated stratification method. (Graphic courtesy of Kevin Hunt - AF Section -NASS)

Oklahoma 2013 Area Frame:

- New state frames are being built at reduced cost with improved objectivity, efficiency and accuracy.
- The 2010 Oklahoma frame (traditional method)
 - 4552 employee hours
- The 2013 Oklahoma frame (CDL automated method)
 - 1980 employee hours
- Predictive accuracy of the 2013 Oklahoma frame – 78% of segments meeting stratum definition vs. 34% for the 2010 Oklahoma frame (based on 2012 JAS data)



“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”



Applications of the Cropland Data Layer within NASS

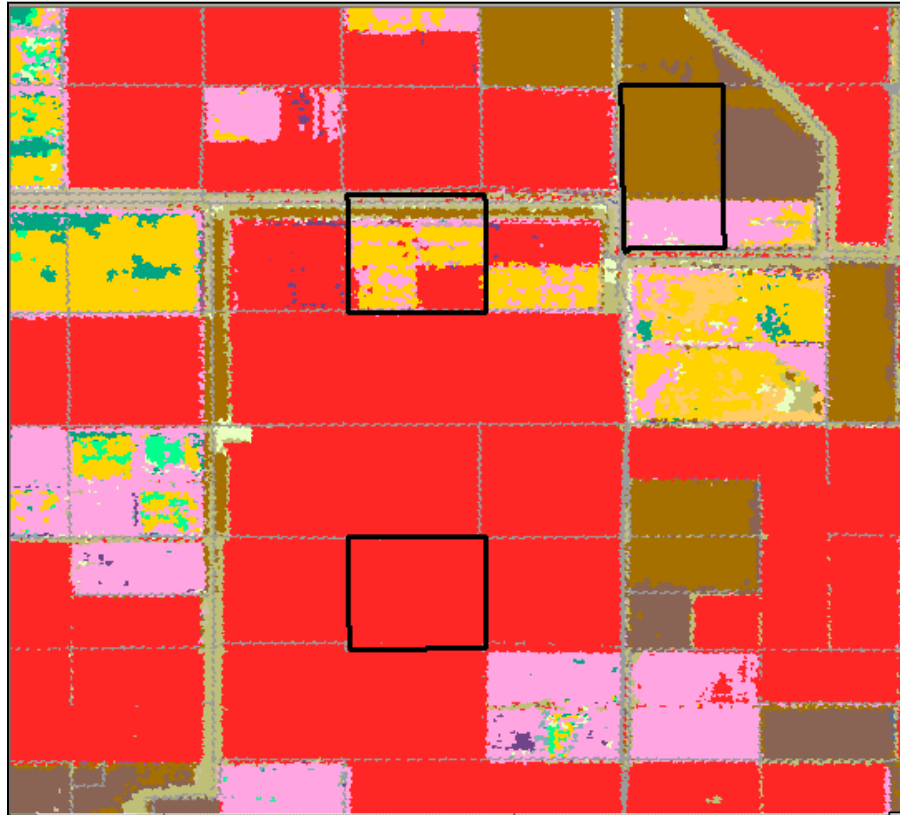
- Crop Acreage Estimation (1997-2012)
- Area Sampling Frame Stratification (2010-2013)
- **Area Sampling Frame Substratification (2013)**



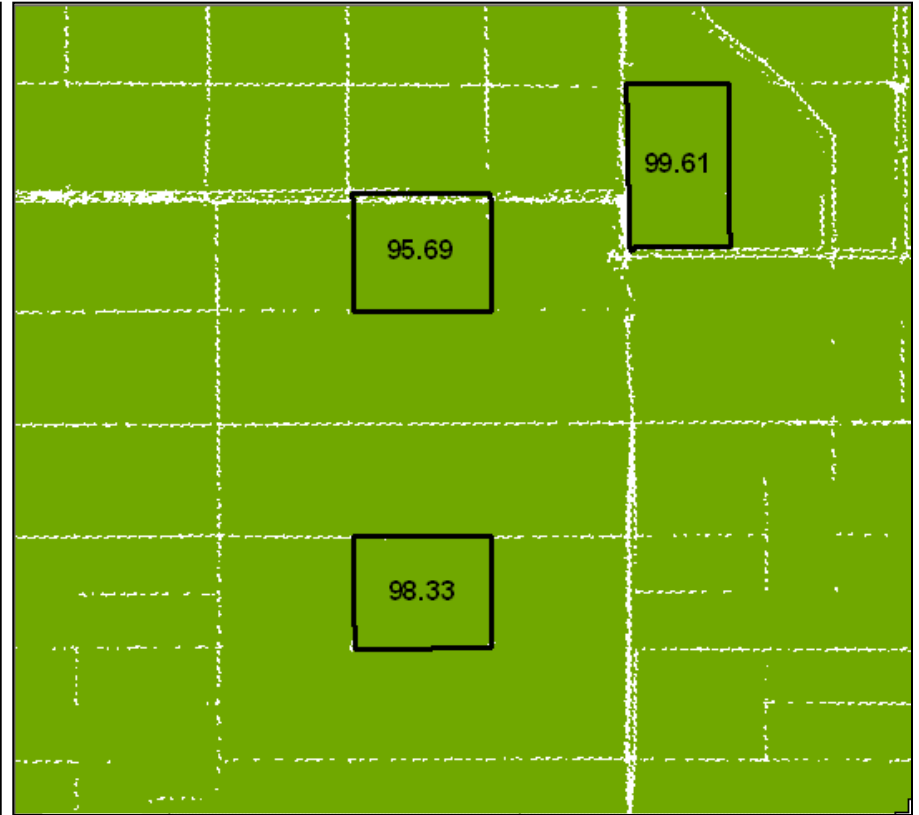
“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”



CDL (2007-2010) covariates at the 2011 JAS segment level

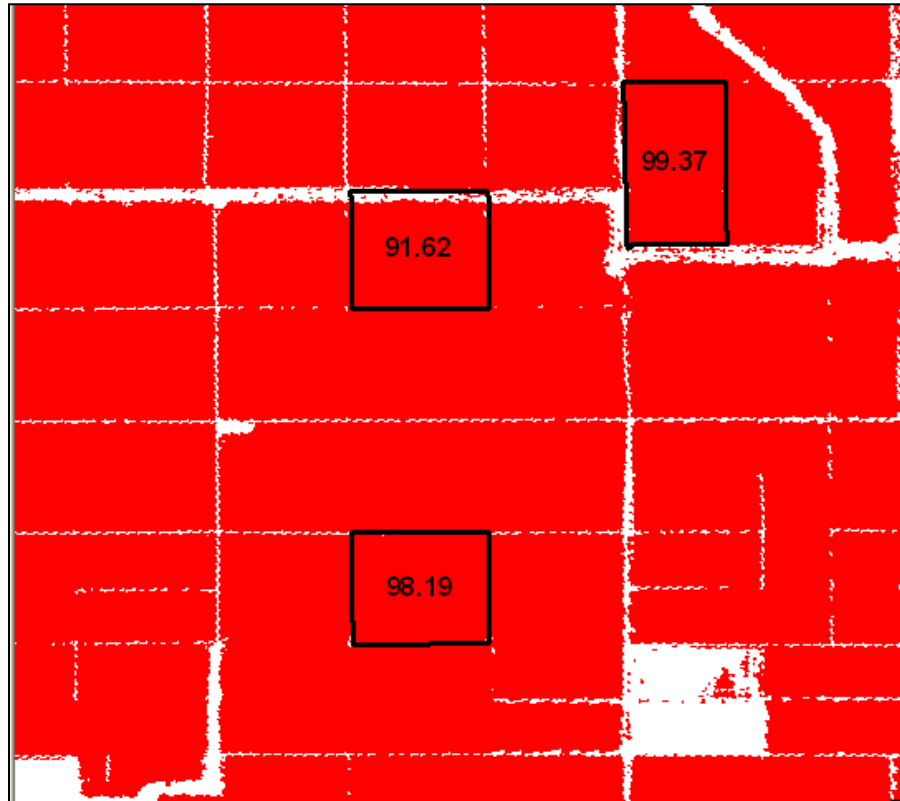


CA11 Cropland Data Layer
with JAS segments

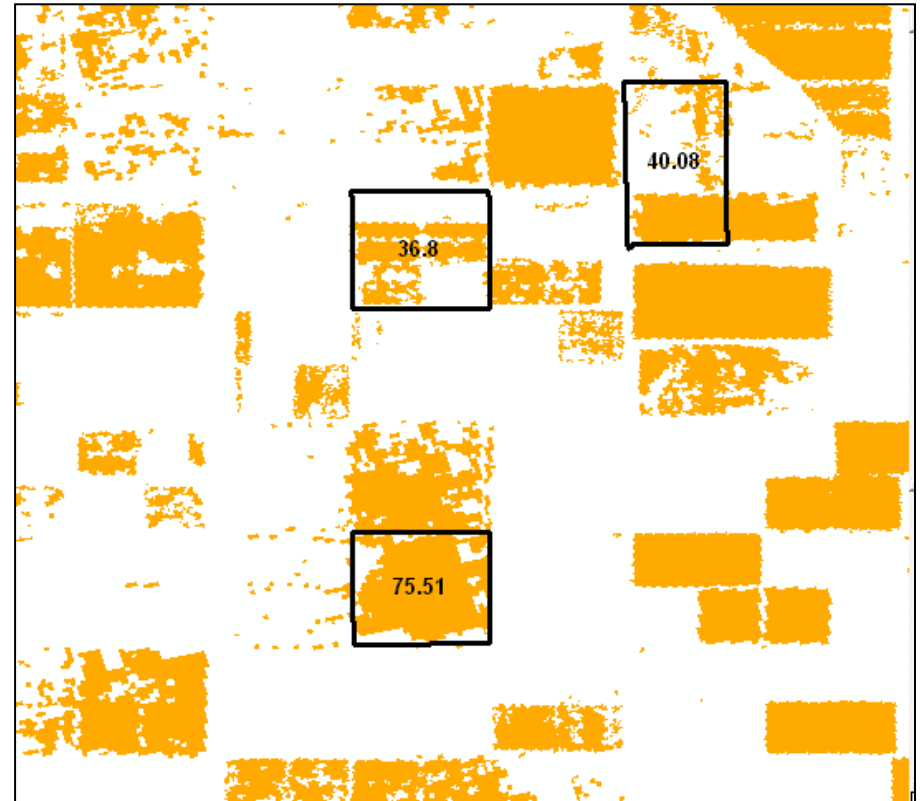


Multi Year (2007-2010) cultivated data set
with JAS segments (percent cultivation
calculated)

CDL (2007-2010) covariates at the 2011 JAS segment level



Multi-year (2007-2010) cotton data set



Multi Year (2007-2010) corn/soy data set

CDL Covariate Predictive Accuracy

	CDL Years	Accuracy	Avg. CDL	Cultivation	Corn/Soy	Wheat	Cotton
California	2007 - 2010	<i>Producer</i>	82.82%	98.95%	52.03%	59.50%	66.73%
		<i>User</i>		95.16%	23.93%	21.06%	36.62%
Indiana	2007 - 2010	<i>Producer</i>	94.82%	96.58%	96.74%	39.88%	N/A
		<i>User</i>		89.08%	86.20%	12.71%	N/A
Mississippi	2007 - 2010	<i>Producer</i>	85.79%	84.11%	93.18%	50.65%	67.55%
		<i>User</i>		93.08%	57.46%	23.08%	36.98%
Nebraska	2007 - 2010	<i>Producer</i>	93.06%	98.45%	94.19%	68.44%	N/A
		<i>User</i>		99.63%	83.76%	25.35%	N/A
Pennsylvania	2008 - 2010	<i>Producer</i>	69.74%	74.16%	83.35%	23.94%	N/A
		<i>User</i>		68.48%	53.11%	8.37%	N/A
Washington	2007 - 2010	<i>Producer</i>	90.27%	89.61%	68.01%	90.04%	N/A
		<i>User</i>		88.78%	27.65%	49.93%	N/A

Validation : 2011 Farm Service Agency Common Land Unit/ NLCD 2006 – cultivation
 2011 Cropland Data Layers - corn/soy, wheat, cotton

Applications of CDL Covariate Data within NASS

(courtesy Jonathan Lisic and Kevin Hunt)

- In the past, commodity information was derived at the county level to infer the agricultural makeup for an entire county.
- CDL covariate data sets provide the opportunity to automatically substratify the NASS Area Frame based on commodity information at the Primary Sampling Unit level.



“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”



Oklahoma 2013 substratification

- Problem: Find the assignment of N_h sampling units to H strata that ***minimizes the sample size***

$$n = \sum_{h=1}^H n_h$$

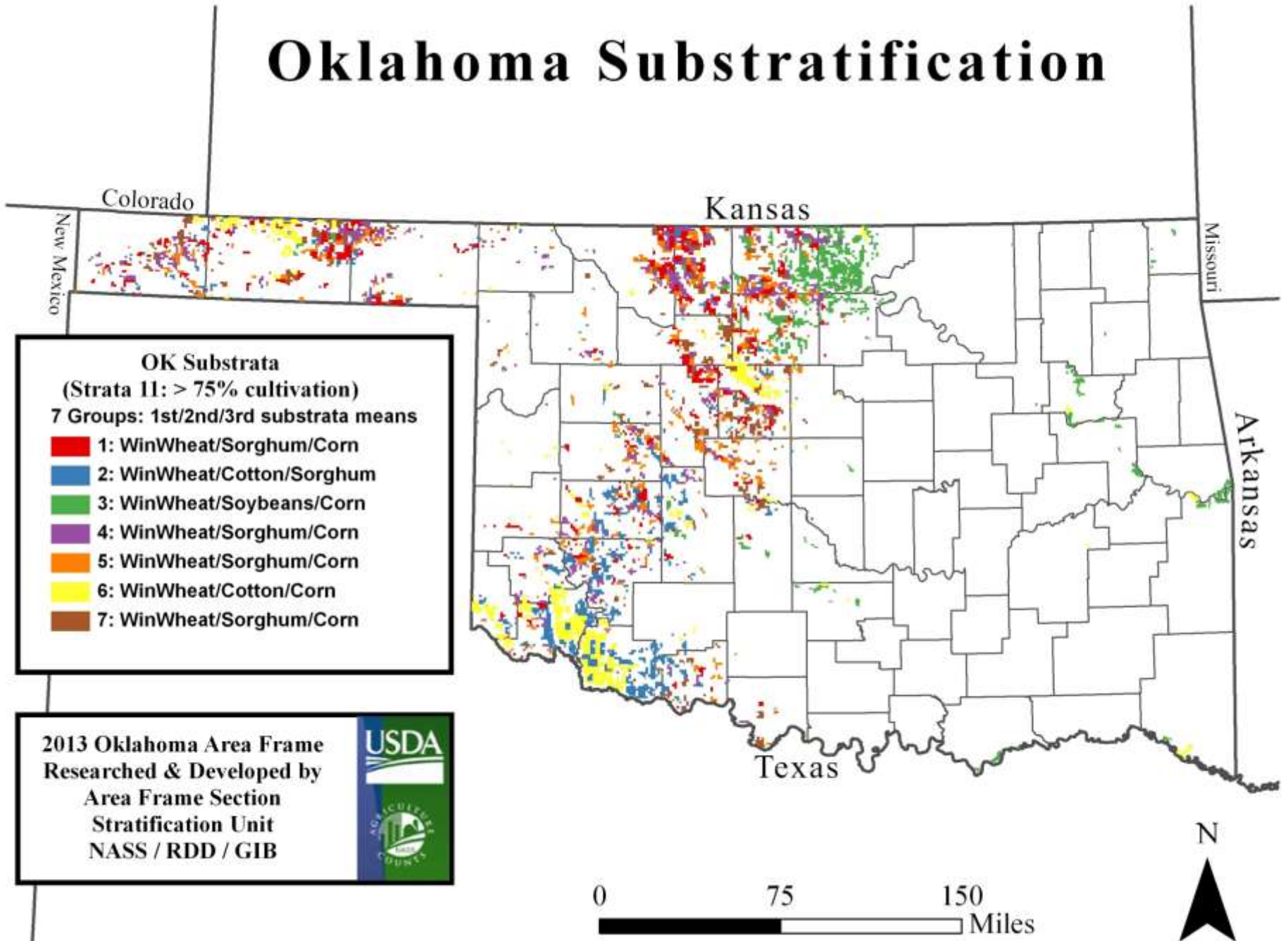
- Subject To

$$T_j \geq \sum_{h=1}^H \frac{N_h^2 S_{h,j}^2}{n_h}$$

- Where T_j is a target variance for commodity j for stratum h in $\{1, \dots, H\}$

$$S_{h,j}^2 = (N_h - 1)^{-1} \sum_{i=1}^{N_h} (x_{i,j} - \bar{x}_j)_{h,j}^2$$

Oklahoma Substratification



Design Effects

- “The **design effect provides a measure of the precision** gained or lost by use of the more complicated design instead of Simple Random Sampling (SRS)” (*Lohr, S., 2010*)
- The **design effect is equal to** the
 - variance (estimator from sampling plan) divided by
 - variance (estimator from a SRS in stratum 11 with the same # of observation units)
- **Design effect values less than 1 indicate an increased precision** (reduced variance) in the estimator

Design Effects

- Comparing prior year design effects using CDL covariate data shows a reasonable overall improvement in substratification efficiency.

Year	Corn	Cotton	Soybeans	Winter Wheat
2012	0.811	0.811	0.773	0.733
2013	0.830	0.683	0.382	0.508

Conclusion

- The strength of the NASS Cropland Data Layer (CDL) product and CDL based stratification method is the **objective and consistent identification of cultivated cropland.**
- Utilizing the CDL data for Area Frame stratification and sub-stratification will **improve the efficiency, reduce the cost and improve the precision of the June Agricultural Survey estimates.**



“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”





Thank you

Questions?

Claire G. Boryan, Geographer
USDA/NASS/RDD
Feb 22, 2012



“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”

