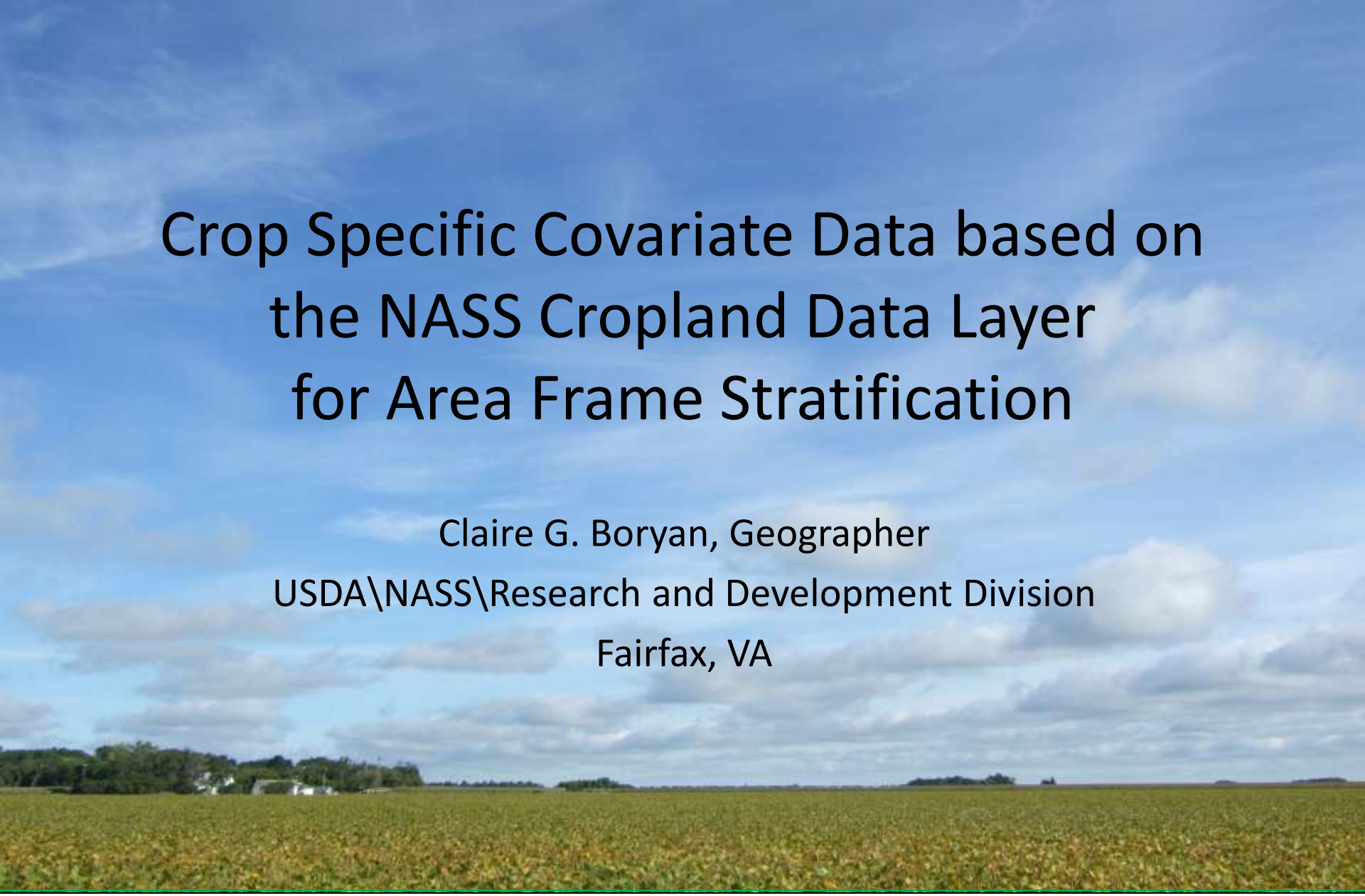
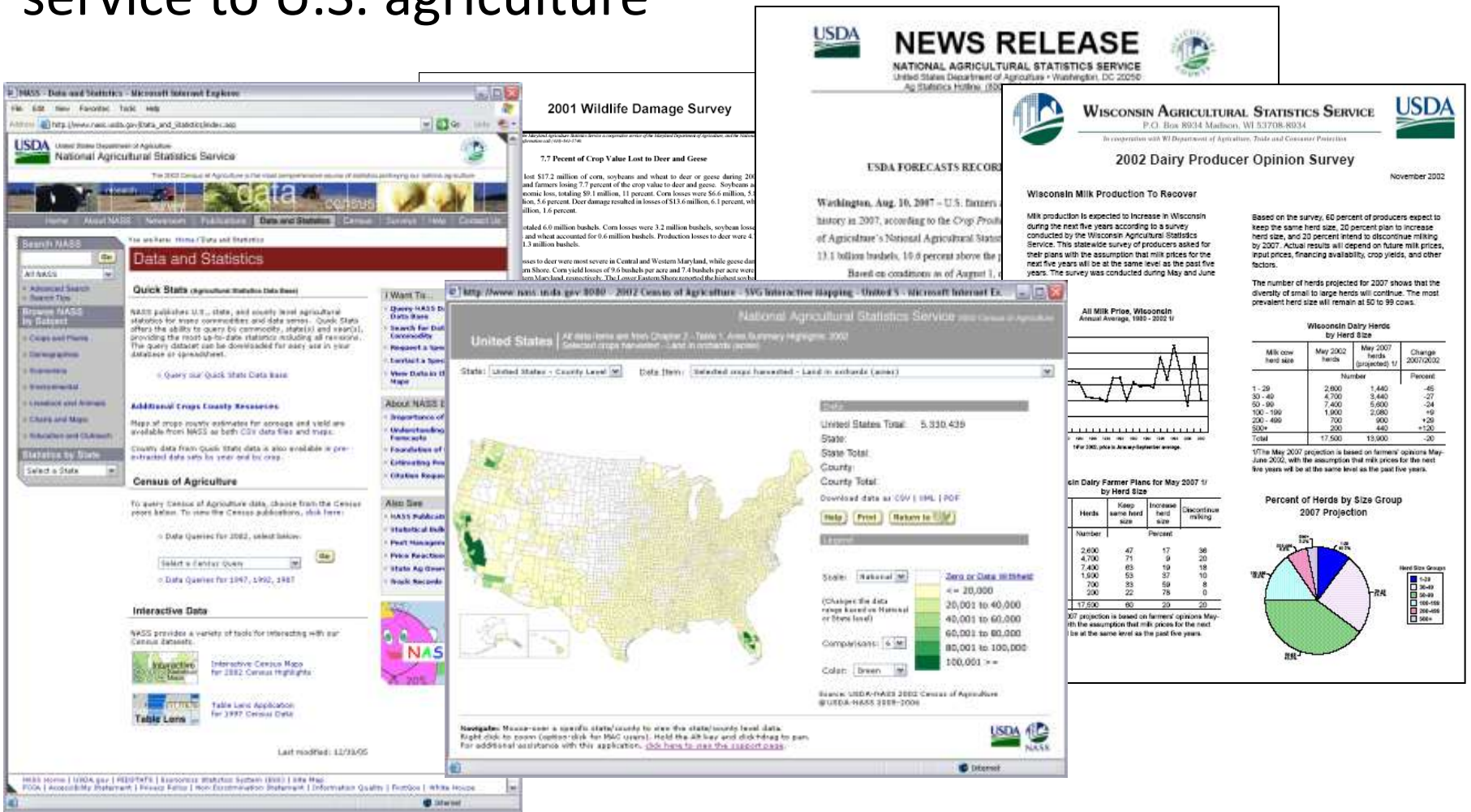


Crop Specific Covariate Data based on the NASS Cropland Data Layer for Area Frame Stratification

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

Background

- NASS provides 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)**

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.

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.

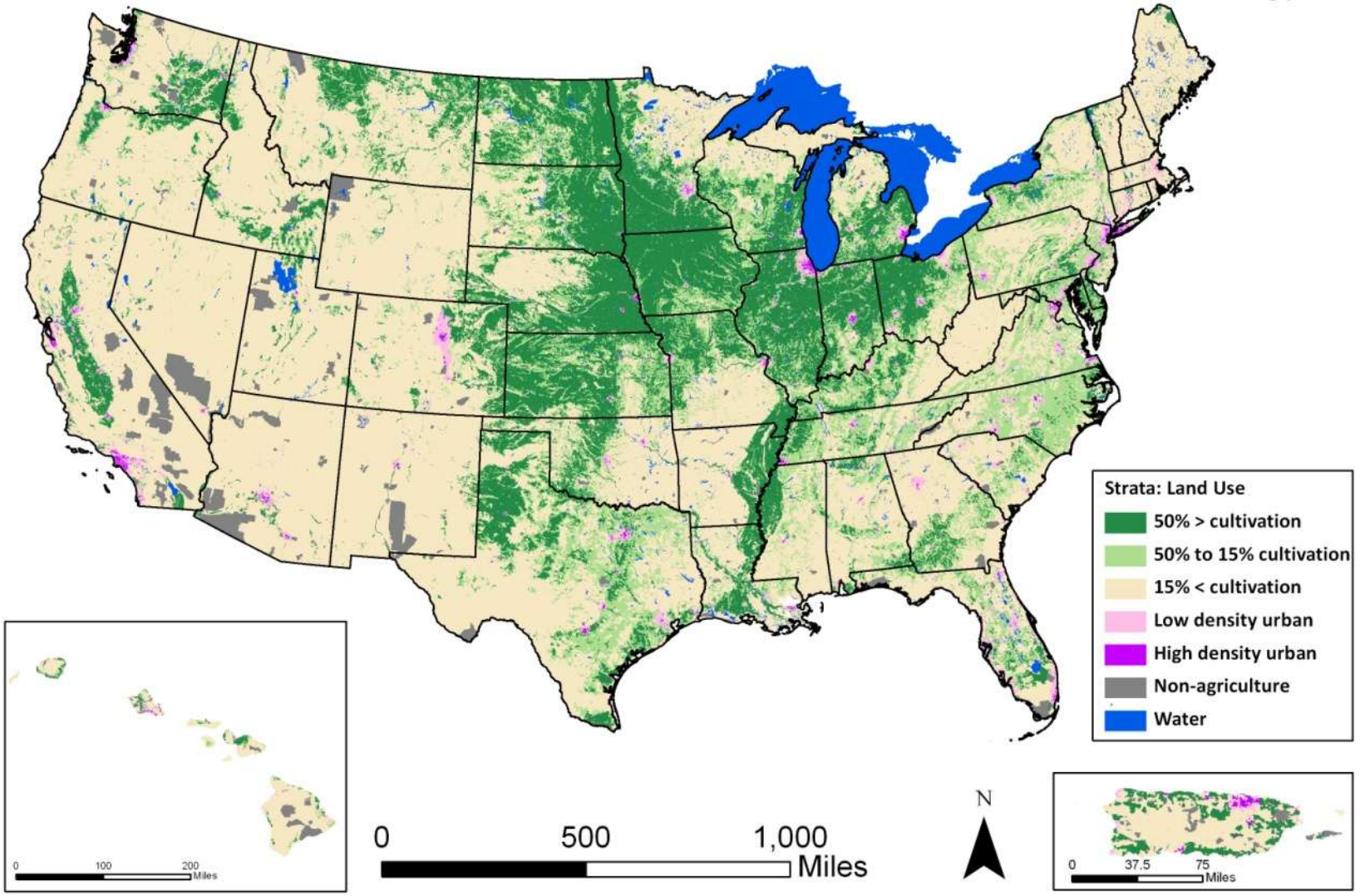
How Is Stratification Performed at NASS?

- 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.

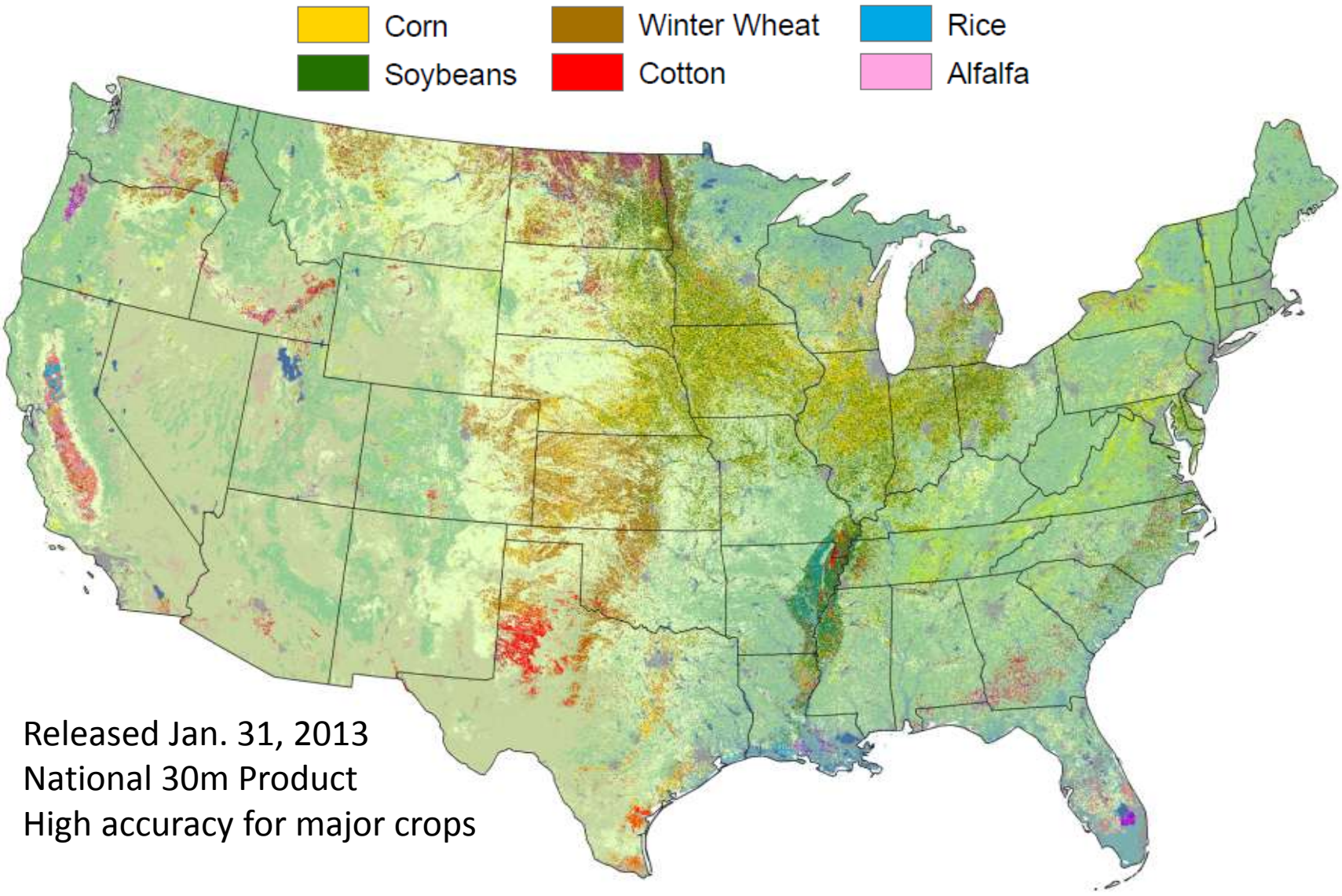
- In the past two years **Cropland Data Layer (CDL) automated stratification** has begun to be implemented.

National Agricultural Statistics Service
Land Use Area Frame



What is the Cropland Data Layer (CDL)?

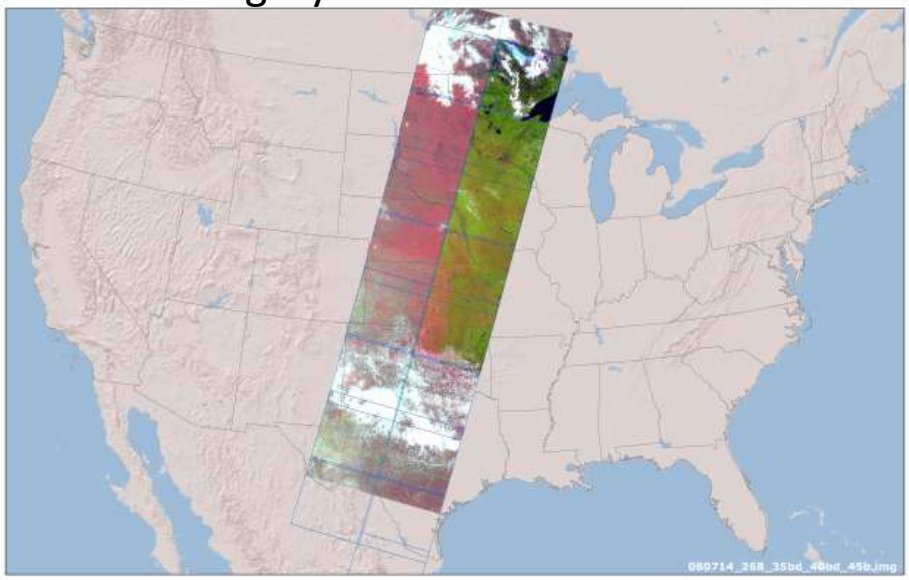
The Cropland Data Layer product is a raster-formatted, crop specific, land cover data set used to generate in-season acreage estimates for 19 crops in 41 states.



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

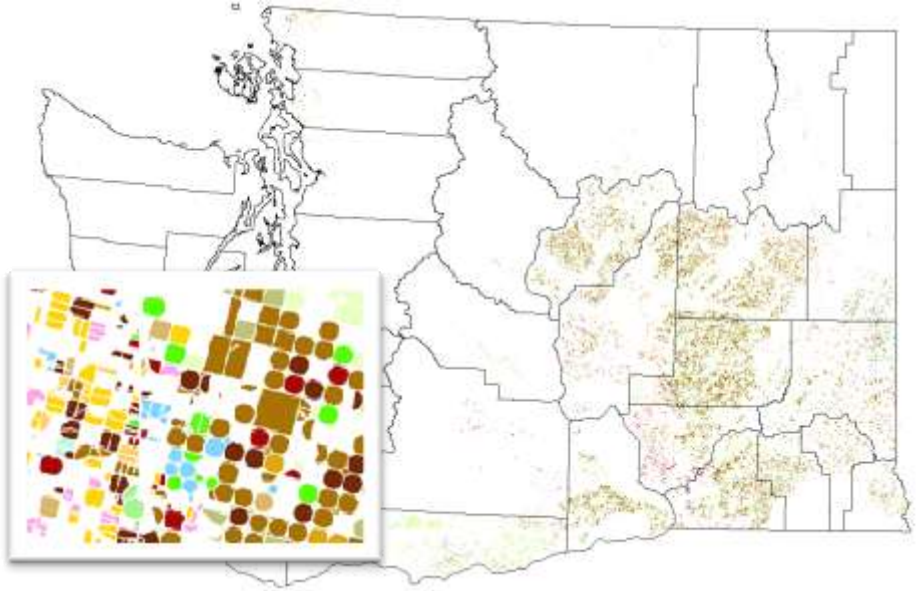
2012 Cropland Data Layer Inputs

Satellite Imagery – Landsat

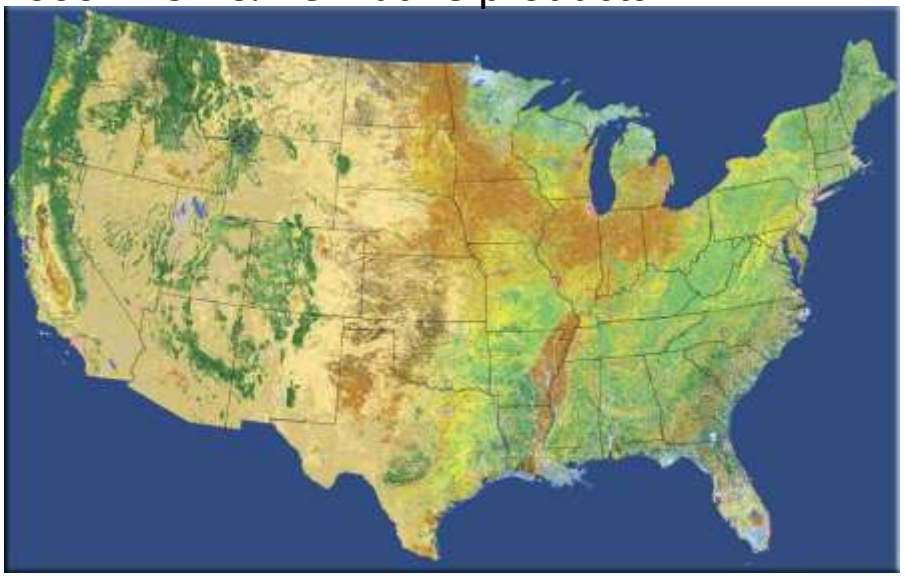


Satellite Imagery – Deimos & UK2

Farm Service Agency: Common Land Unit

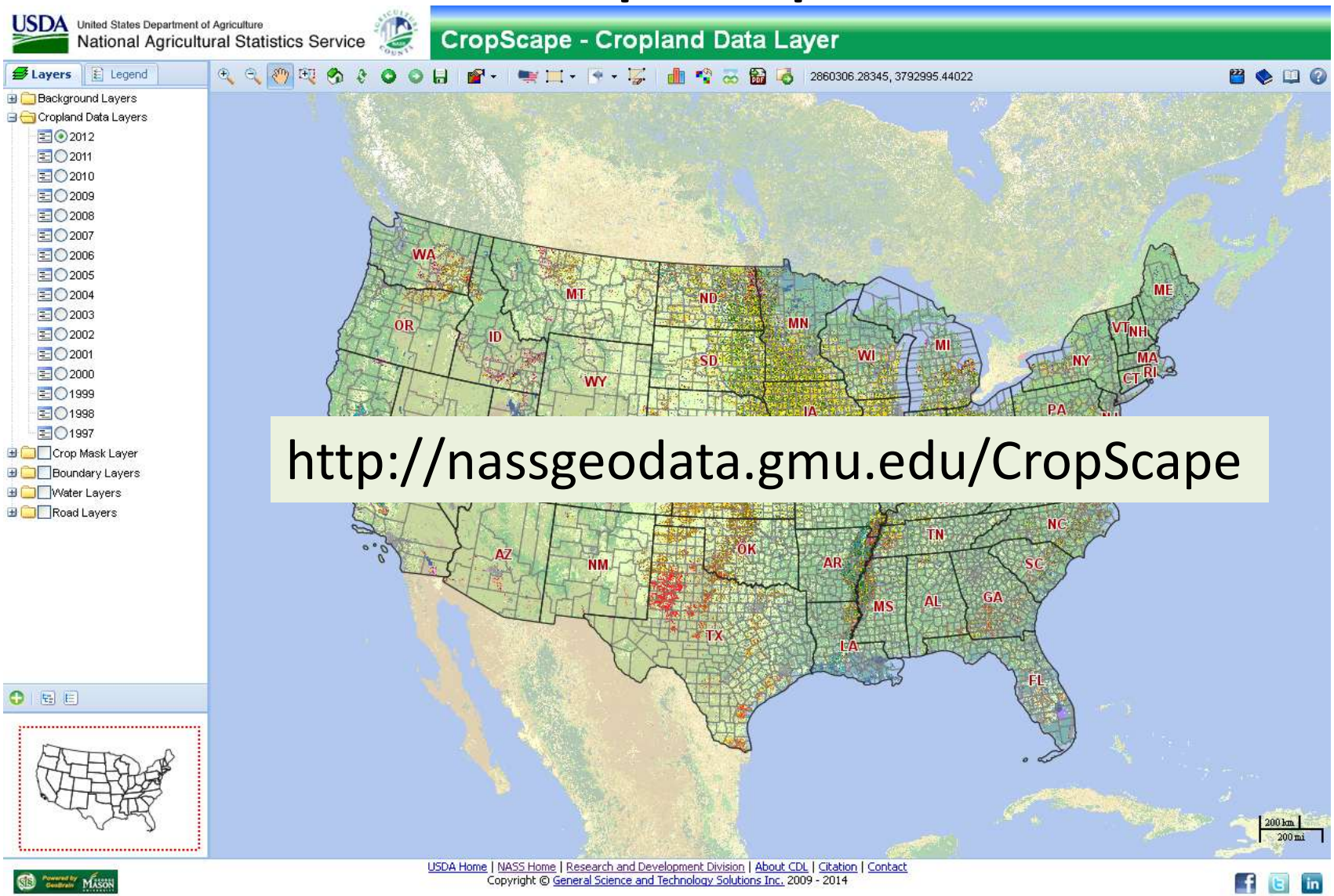


2006 NLCD & Derivative products



National Cropland Data Layer Legend

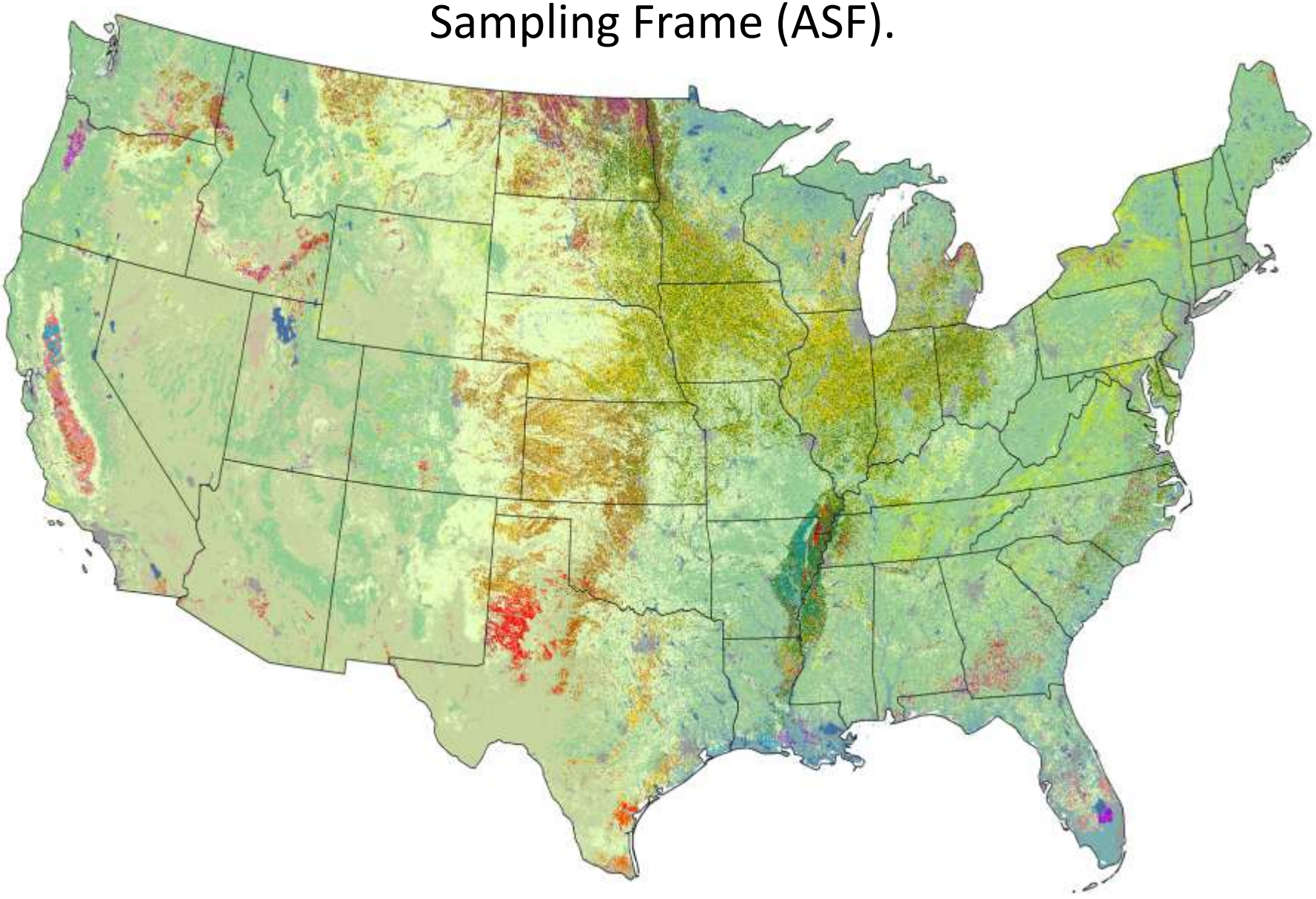
CropScape



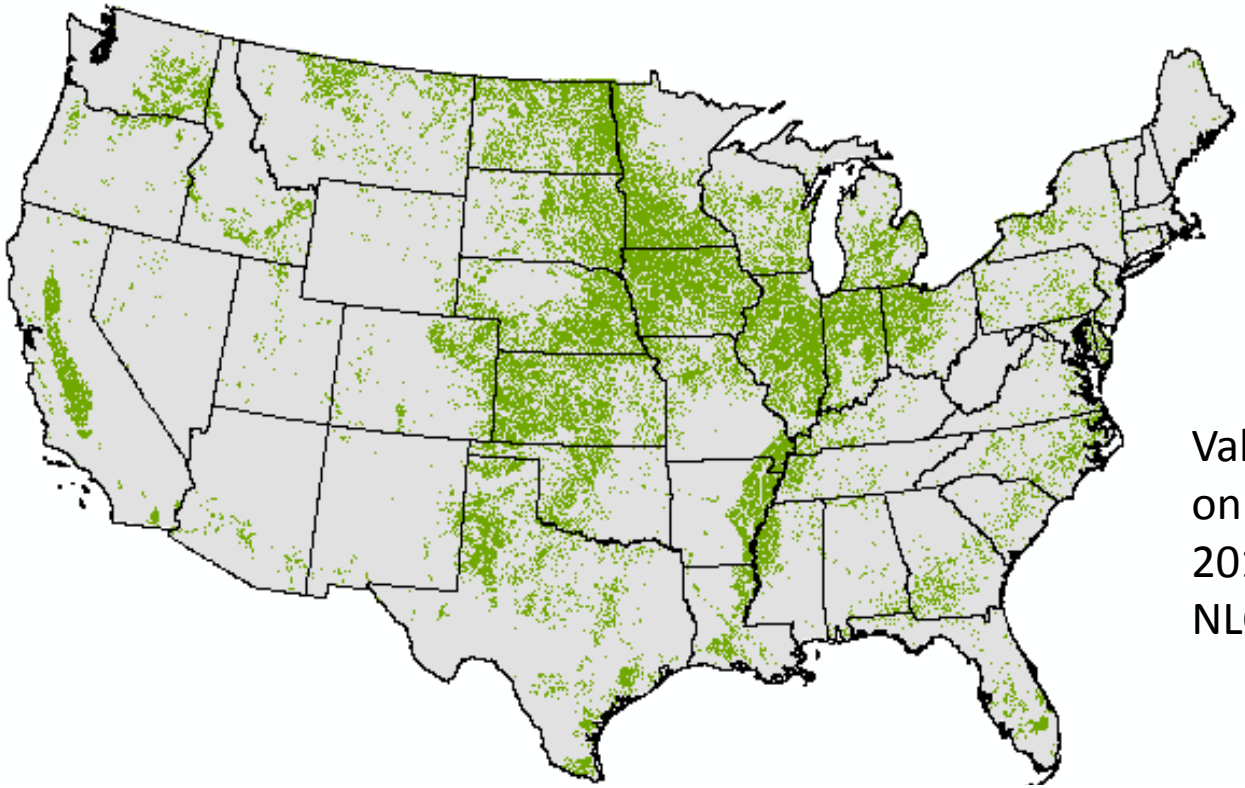
http://nassgeodata.gmu.edu/CropScape

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

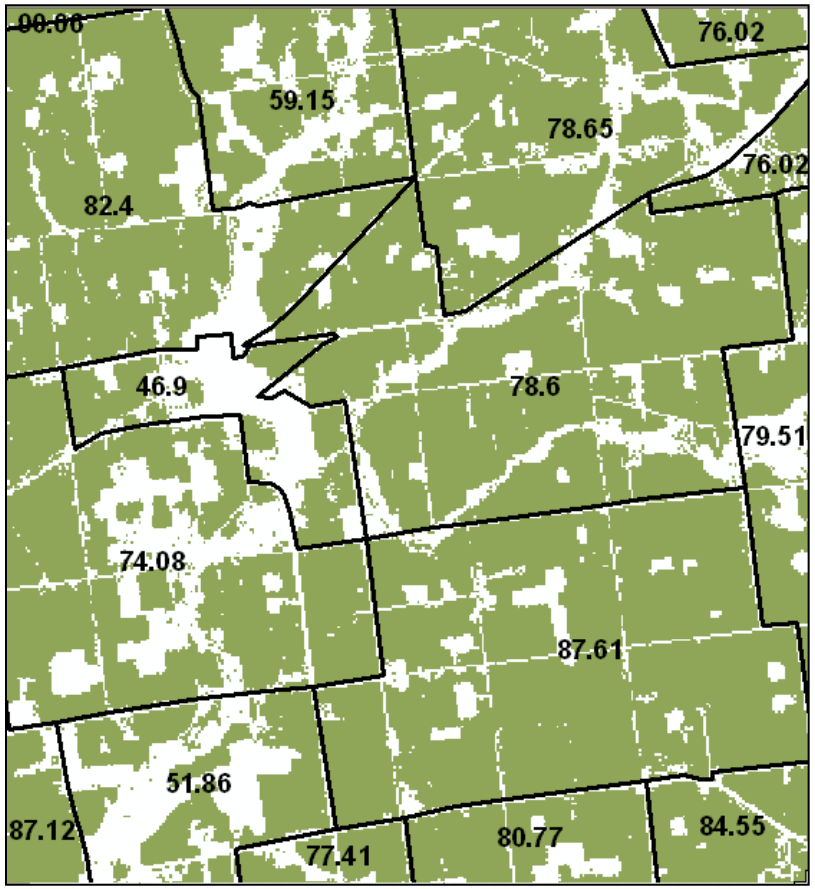
Class Name	Code	Producer Accuracy	Omission Error	Kappa	User Accuracy	Commission Error	Conditional Kappa
Non-Cultivated	1	96.6%	3.4%	0.940	97.2%	2.8%	0.951
Cultivated	2	97.9%	2.1%	0.951	97.4%	2.6%	0.940

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

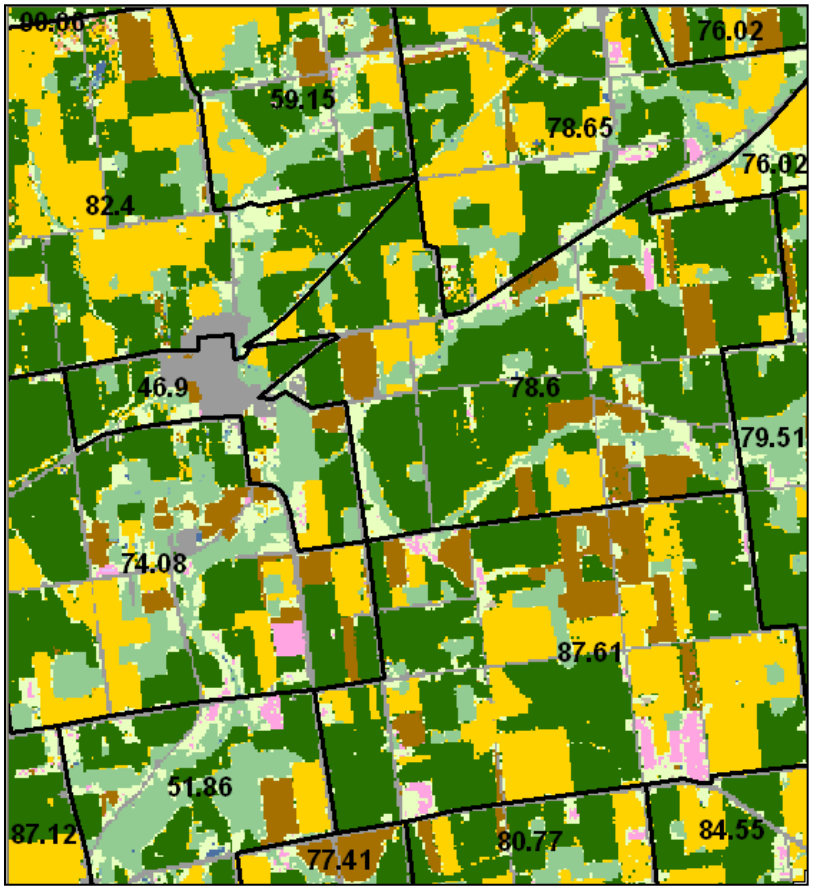
Applications of the Cropland Data Layer within NASS

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

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

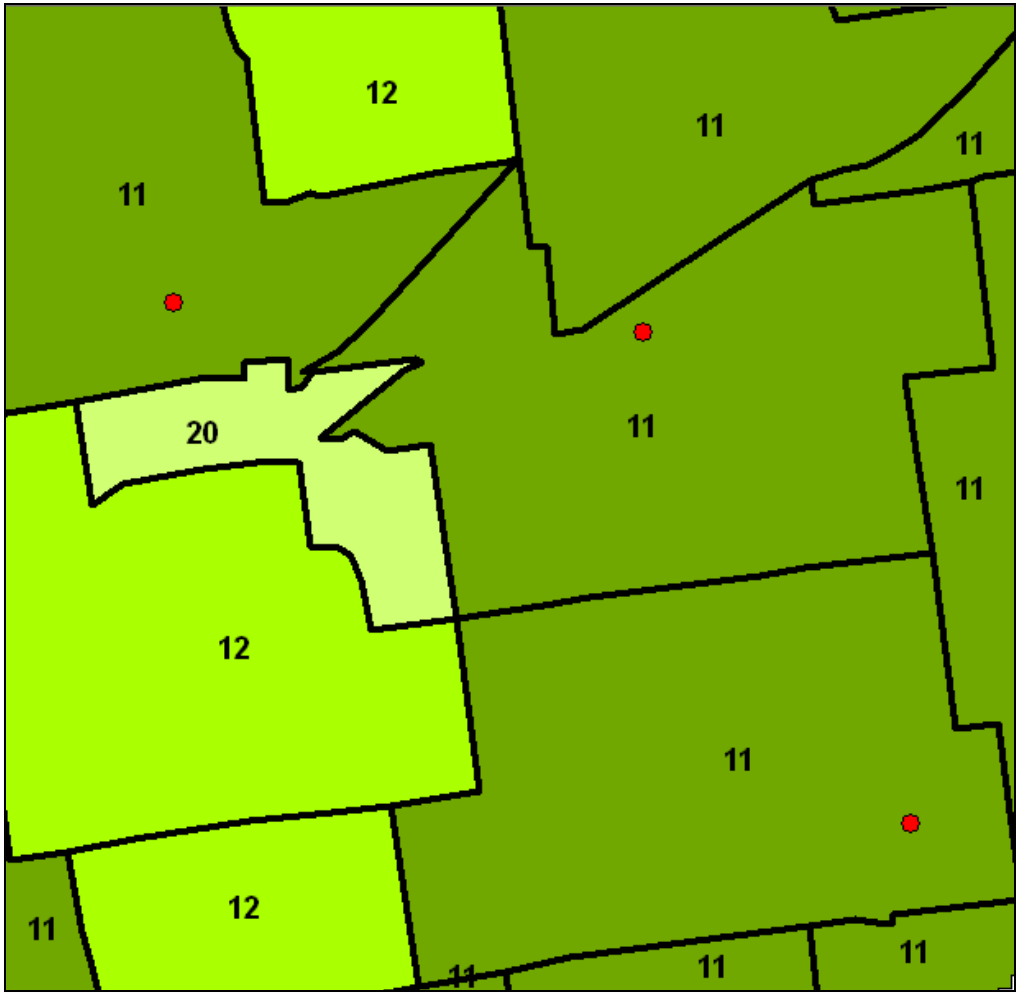


Primary Sampling Units with CDL percent cultivation

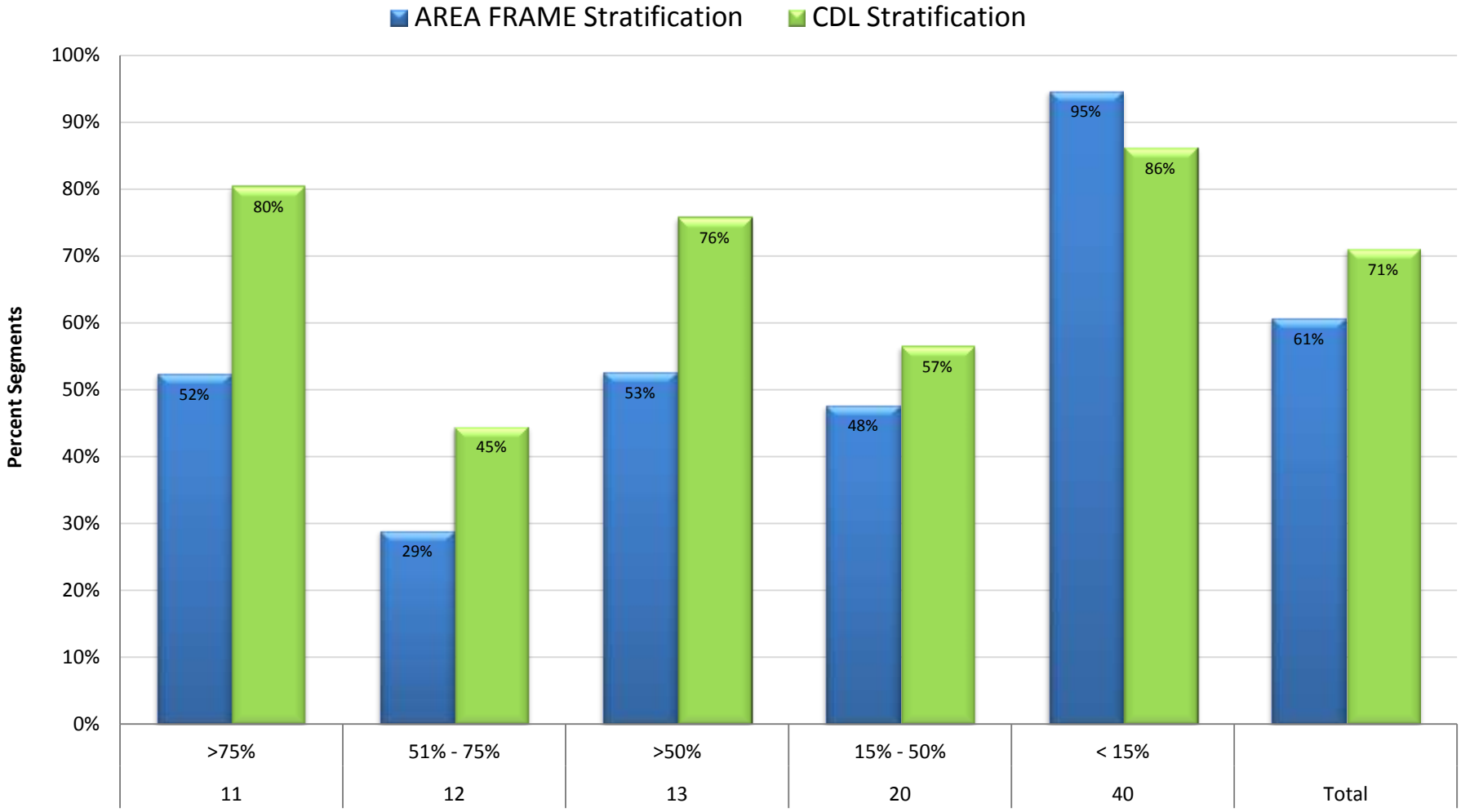


Primary Sampling Units with CDL percent cultivation, overlaying a 2010 CDL image product

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



Area Frame (Traditional) vs CDL Stratification Strata Summary 2010

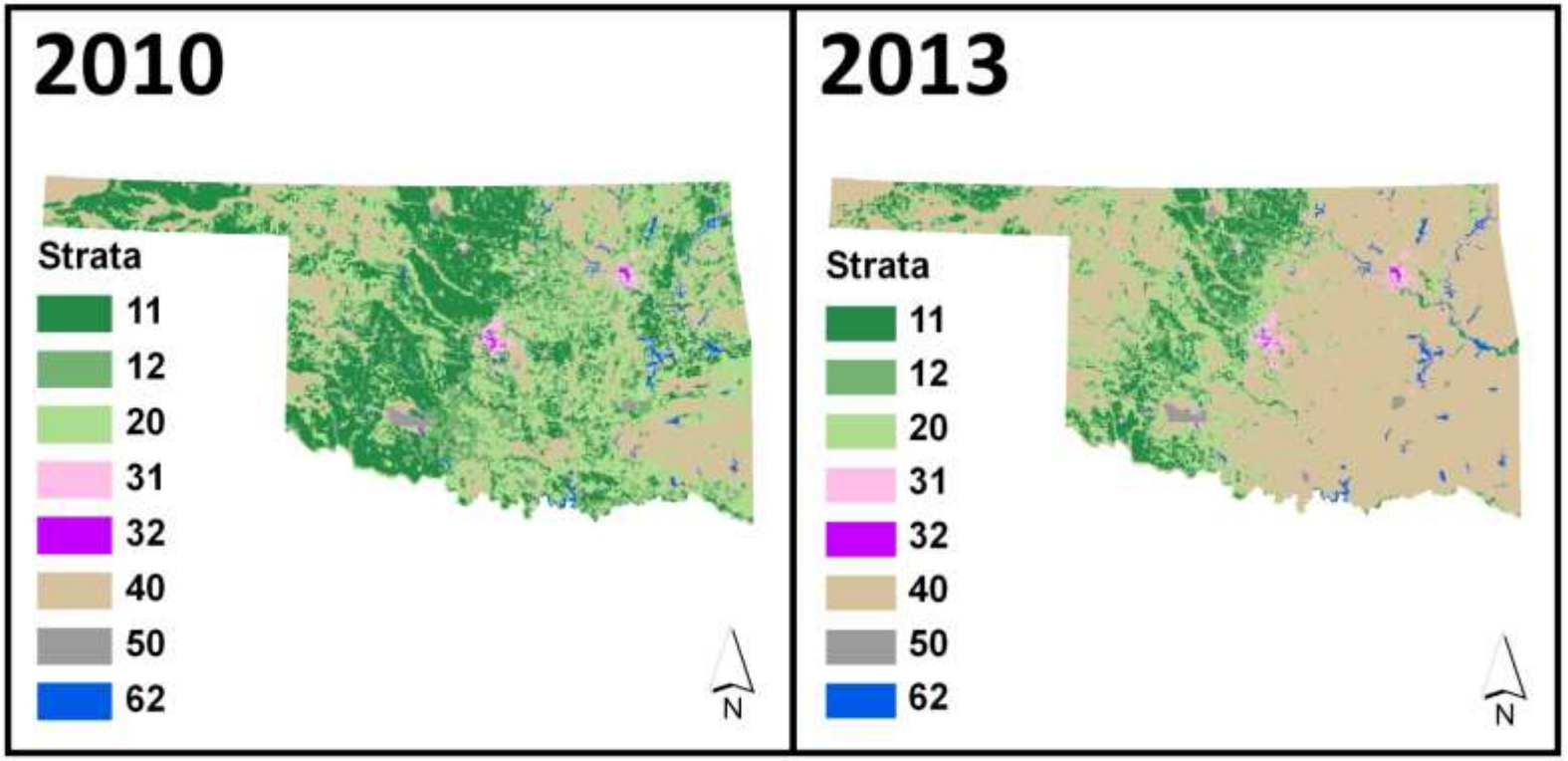


Accuracy based on 2010 June Area Survey in situ data

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*

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

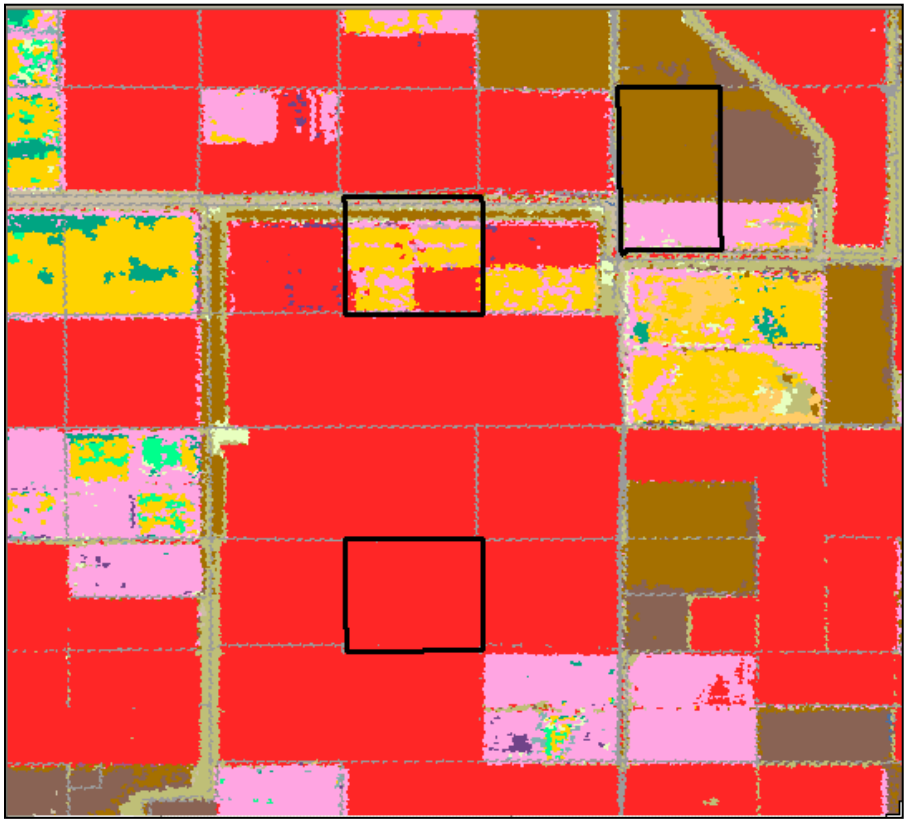
- The 2013 Oklahoma frame (CDL automated method)
  – 1980 employee hours

- 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)

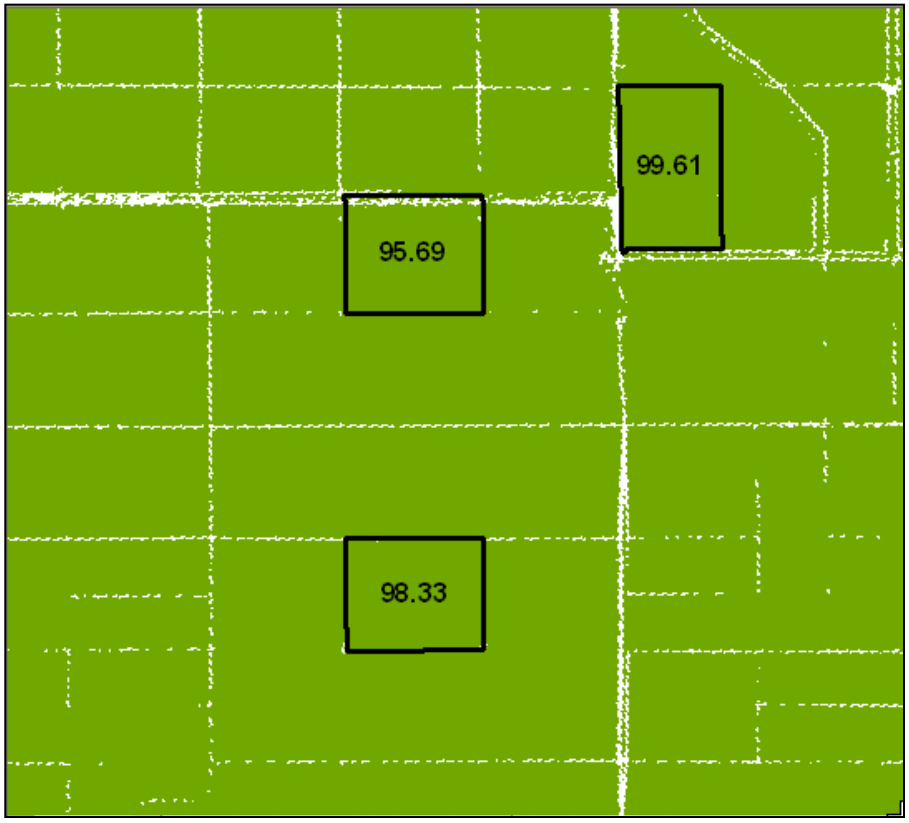
Applications of the Cropland Data Layer within NASS

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

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

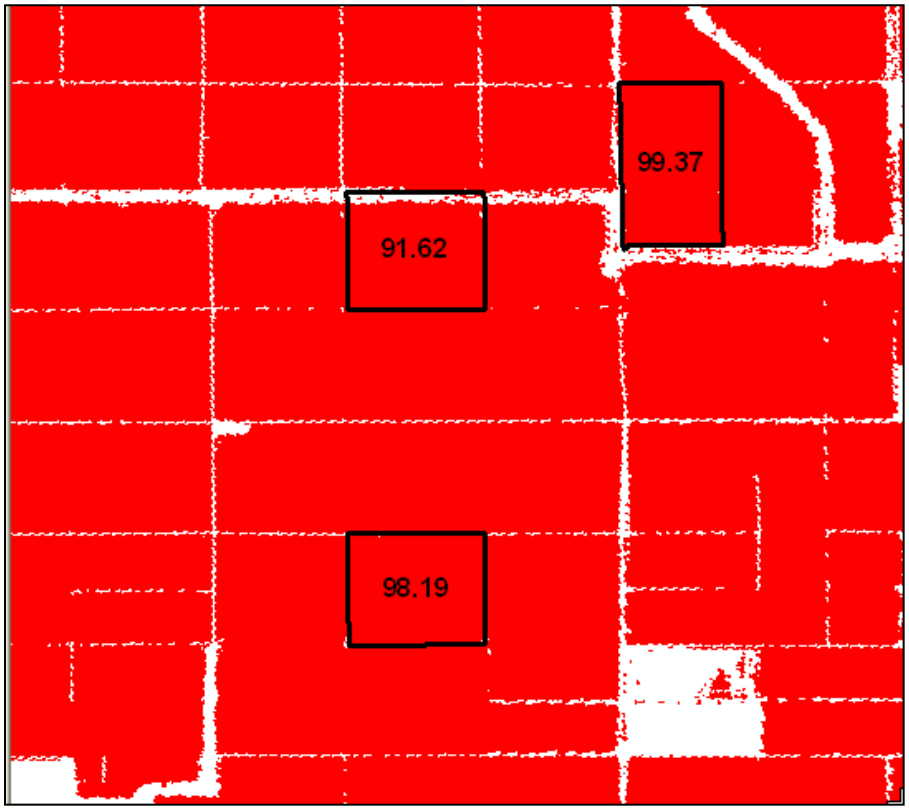


CA11 Cropland Data Layer with JAS segments

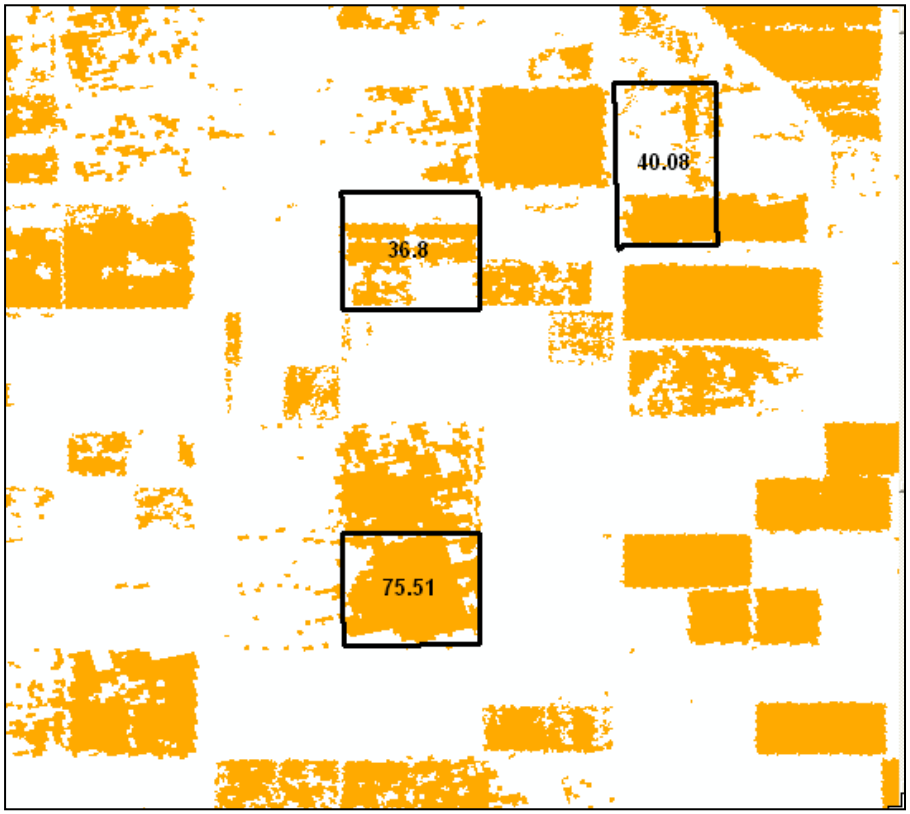


Multi Year (2007-2010) cultivated data set with JAS segments (percent cultivation calculated)

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

Oklahoma 2013 substratification

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

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

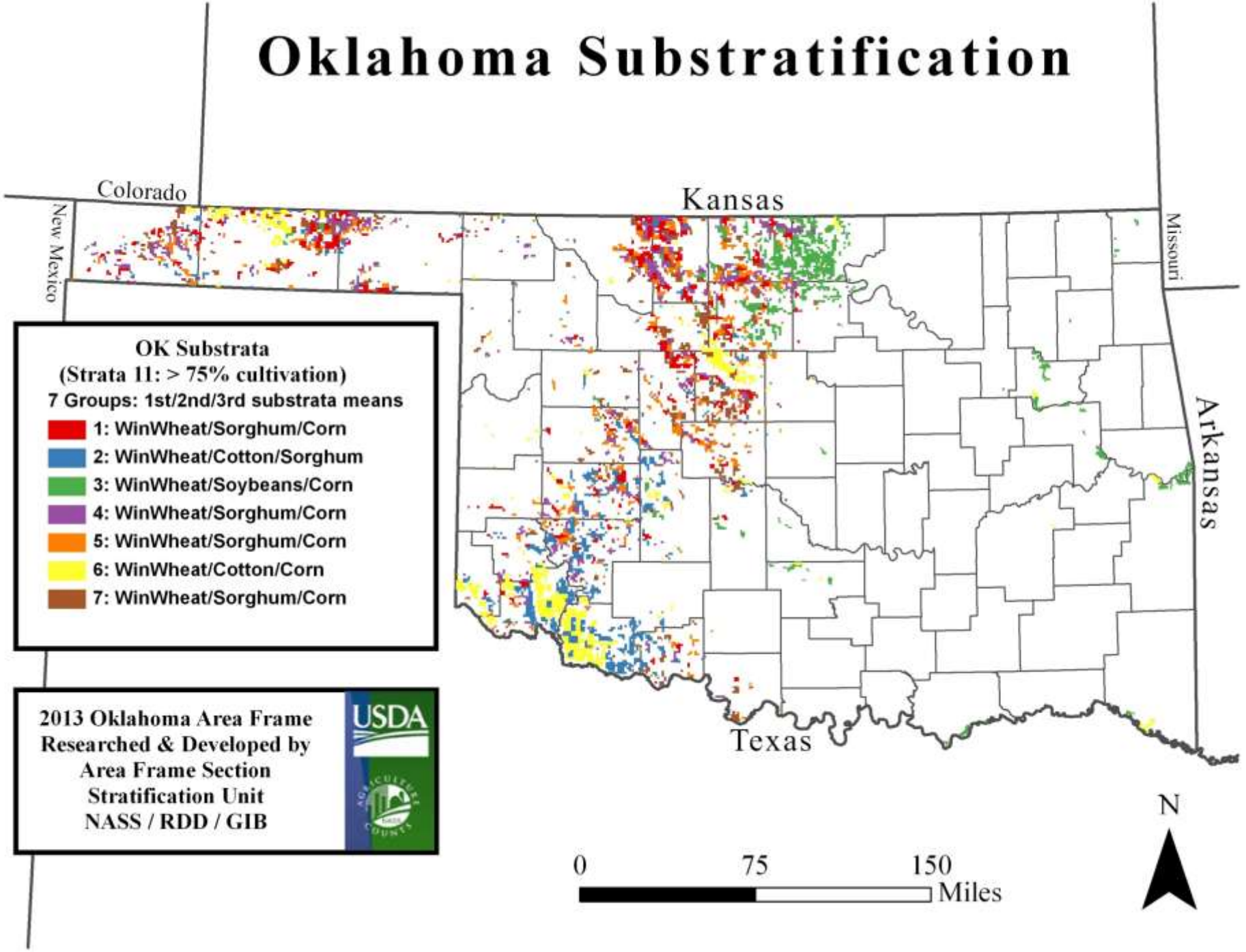
- Subject To

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

- Where T_j is a target variance for commodity j for stratum h in {1,…,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**.

Thank you

Questions?

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