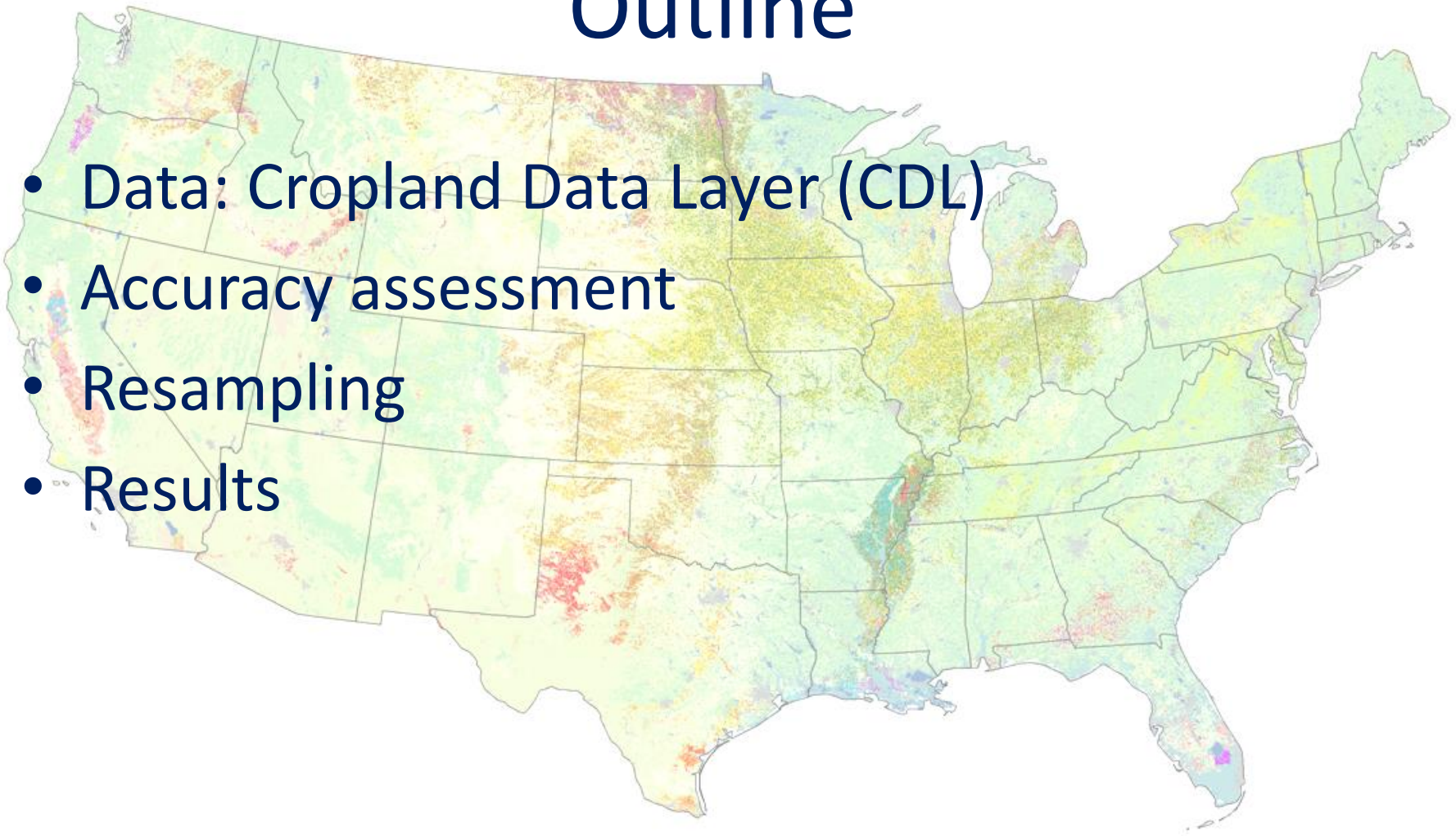
Evaluating the Accuracy Assessment Methods of a Thematic Raster Through SAS® Resampling Techniques and GTL Visualizations

Robert Seffrin

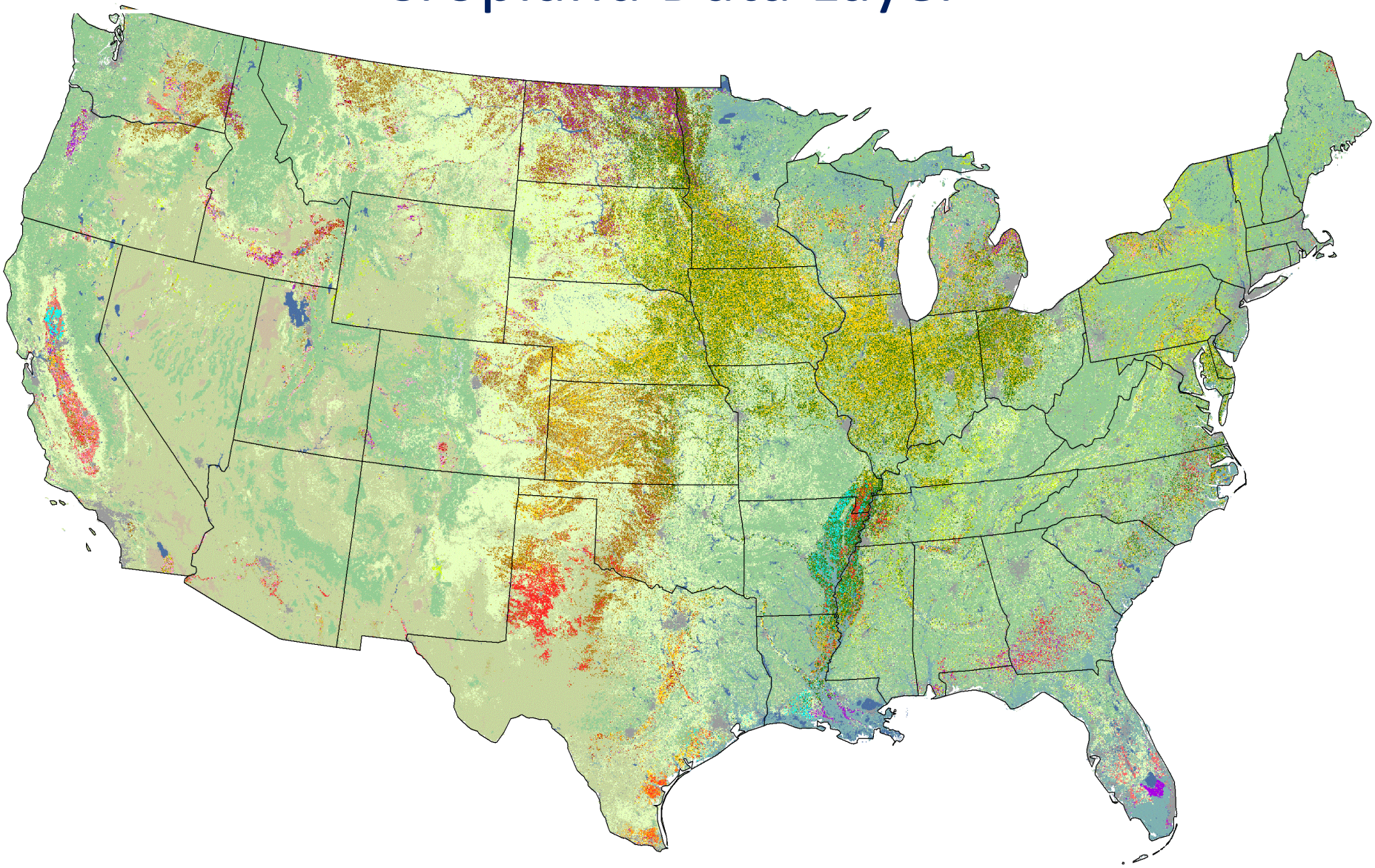
National Agricultural Statistics Service, USDA

Outline



- Data: Cropland Data Layer (CDL)
- Accuracy assessment
- Resampling
- Results

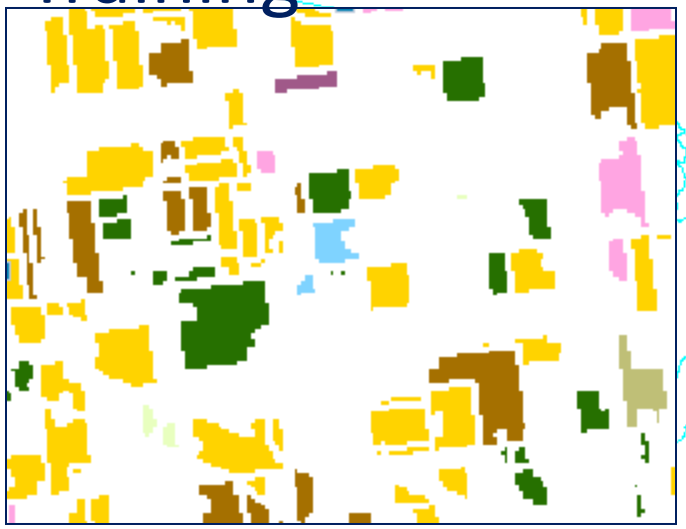
Cropland Data Layer



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

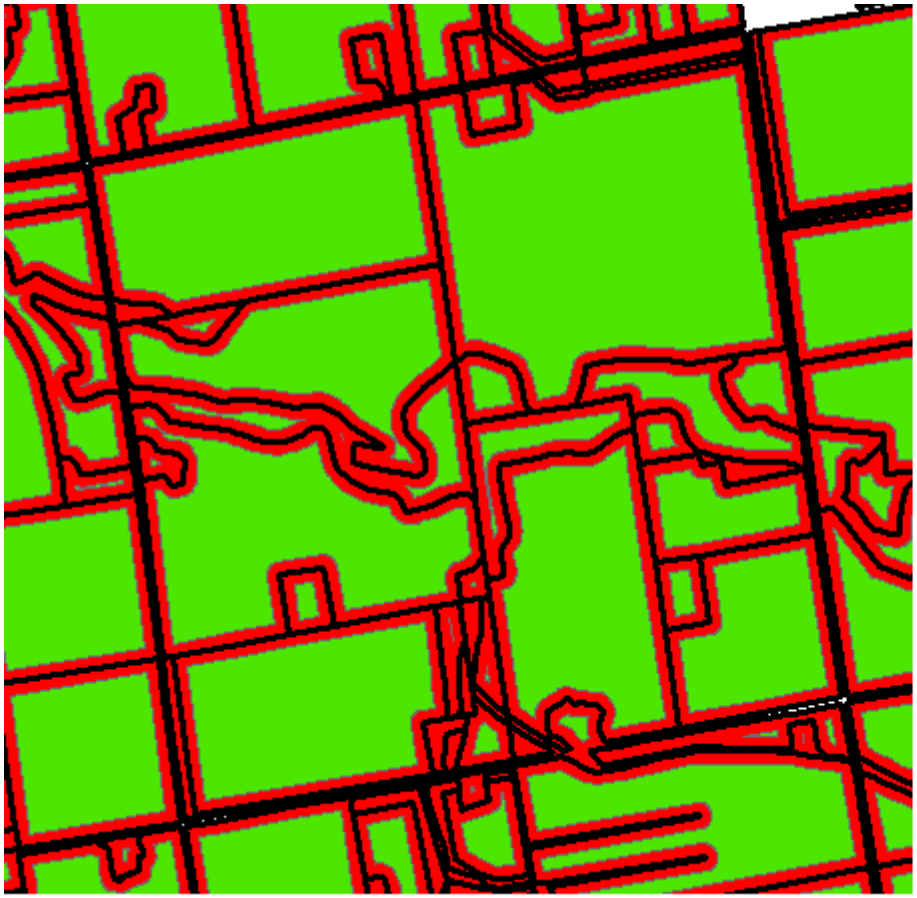
Training and validation data

Training



Validation

Terminology



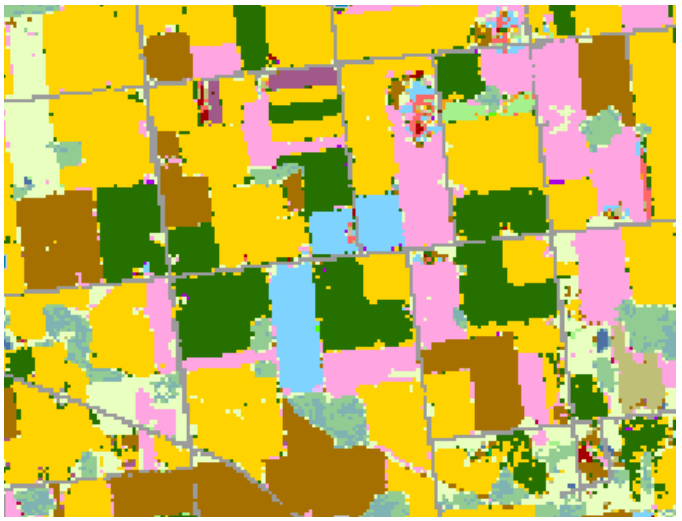
CLU area terminology

CLU boundary

Buffer, outside

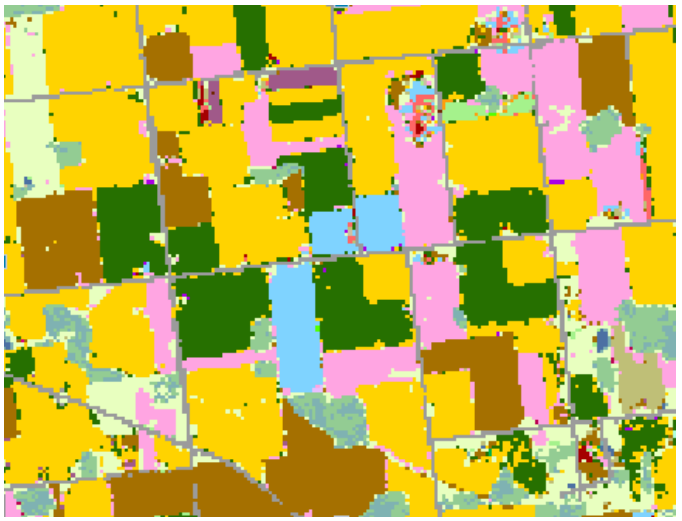
Buffered, inside

Creating the error matrix



	Corn	Sorghum	Soybeans	Sun-flower	Barley	Durum wheat	Spring Wheat	Winter Wheat
Corn	**1,348,923**	204	55,210	172	56	-	-	594
Sorghum	83	**480**	24	-	34	-	-	2
Soybeans	43,520	52	**1,039,576**	301	32	-	-	687
Sunflower	25	9	45	**159**	-	-	-	-
Barley	29	4	57	-	**828**	-	-	13
Durum Wheat	-	-	-	-	-	**-**	-	-
Spring Wheat	-	3	-	-	-	-	**613**	1
Winter Wheat	1,254	26	1,728	-	51	-	3	**277,918**

Creating the error matrix

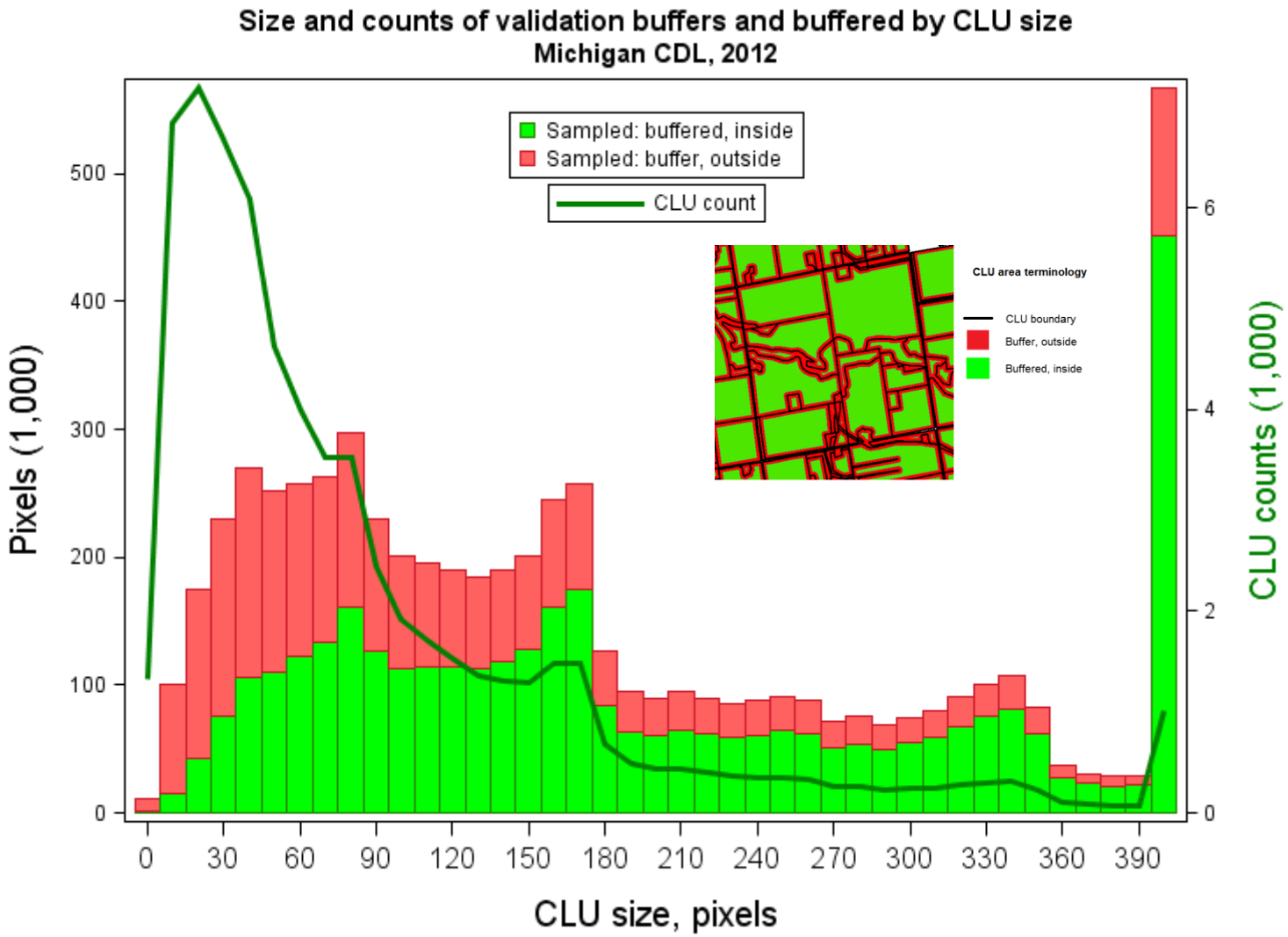


	Corn	Sorghum	Soybeans	Sun-flower	Barley	Durum wheat	Spring Wheat	Winter Wheat
Corn	**1,348,923**	204	55,210	172	56	-	-	594
Sorghum	83	**480**	24	-	34	-	-	2
Soybeans	43,520	52	**1,039,576**	301	32	-	-	687
Sunflower	25	9	45	**159**	-	-	-	-
Barley	29	4	57	-	**828**	-	-	13
Durum Wheat	-	-	-	-	-	**-**	-	-
Spring Wheat	-	3	-	-	-	-	**613**	1
Winter Wheat	1,254	26	1,728	-	51	-	3	**277,918**

Calculate statistics

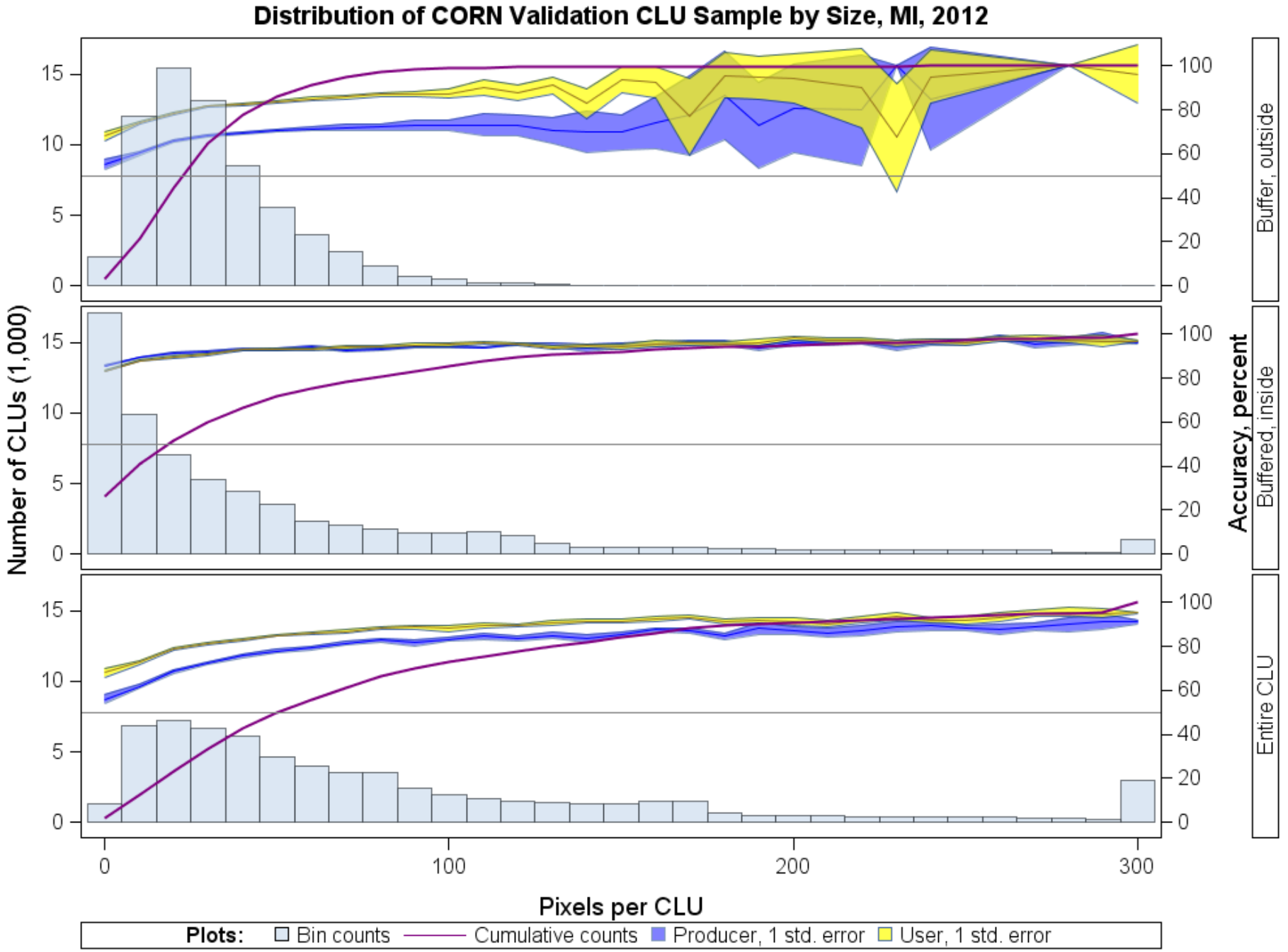
Category	Correct	Total	Producer	Omission	Kappa	Total	User	Commission	Kappa
Corn	1,348,923	1,427,166	94.5%	5.5%	0.915	1,432,870	94.1%	5.9%	0.909
Sorghum	480	1,226	39.2%	60.8%	0.391	819	58.6%	41.4%	0.586
Soybeans	1,039,576	1,124,053	92.5%	7.5%	0.896	1,101,480	94.4%	5.6%	0.922
Sunflower	159	794	20.0%	80.0%	0.200	333	47.7%	52.3%	0.477
Barley	828	2,883	28.7%	71.3%	0.287	1,625	51.0%	49.0%	0.509
Spring Wheat	613	1,231	49.8%	50.2%	0.498	676	90.7%	9.3%	0.907
Winter Wheat	277,918	287,629	96.6%	3.4%	0.964	296,372	93.8%	6.2%	0.933

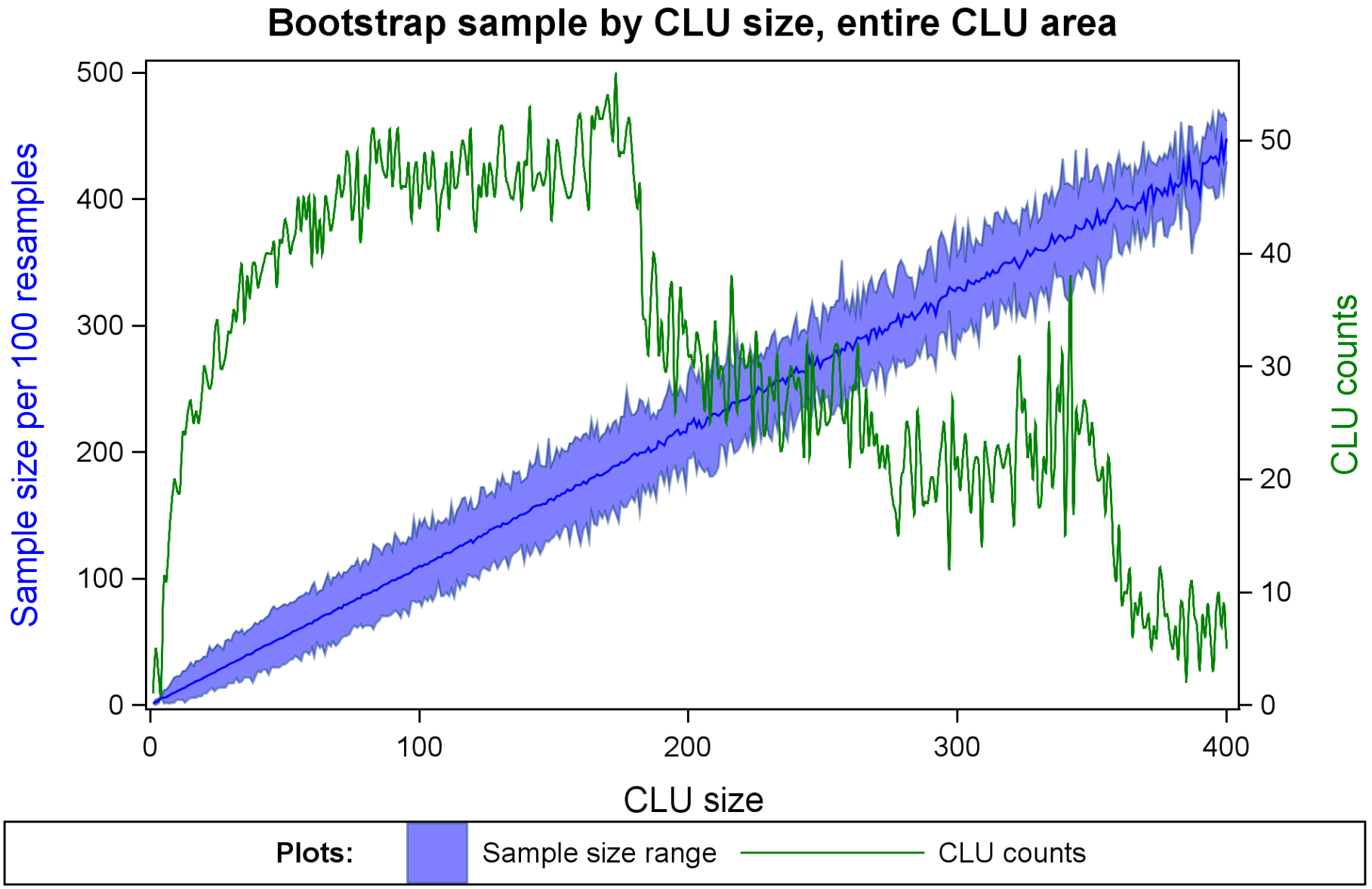
- Data created
- Accuracy assessment
- Issues
  - Spatial autocorrelation
  - Boundaries not represented
  - Cannot calculate confidence interval
- **Explore with resampling in PROC SurveySelect**


Size and counts of validation buffers and buffered by CLU size
Michigan CDL, 2012
Sampled: buffered, inside
Sampled: buffer, outside
CLU count
CLU area terminology
CLU boundary
Buffer, outside
Buffered, inside
Pixels (1,000)
CLU counts (1,000)
CLU size, pixels
0
100
200
300
400
500
0
2
4
6
0
30
60
90
120
150
180
210
240
270
300
330
360
390

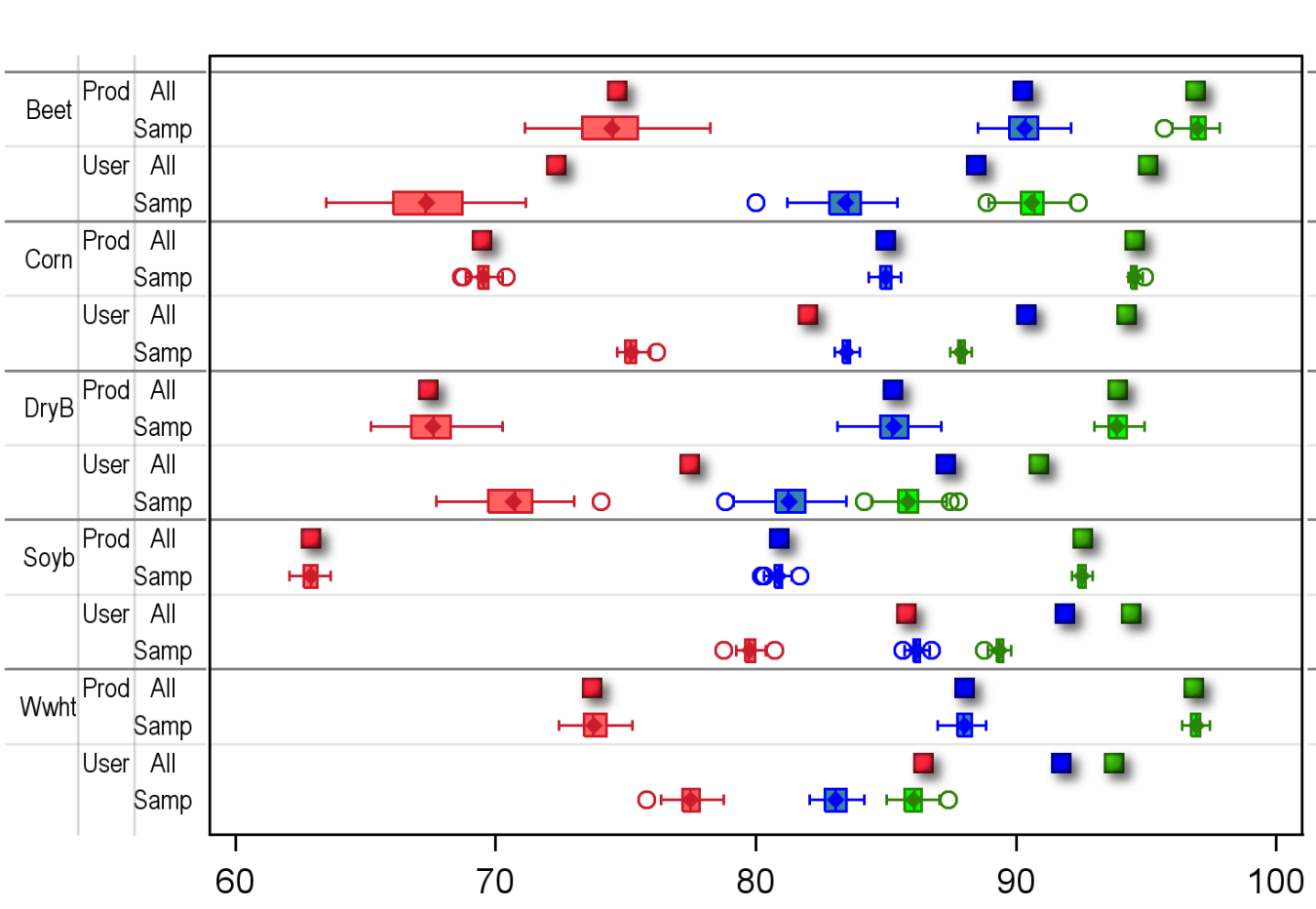
Bootstrap resampling

```
SASFILE MI_CLUs_Valid_State LOAD;
OPTIONS  NoNotes;
PROC SURVEYSELECT
    DATA       = MI_CLUs_Valid_State
    METHOD = SRS
    N          = Temp_Hist_Bin_DATASET
    SEED       = 9999
    OUT        = outFile
    REP        = 20
    NOPRINT ;
    STRATA     buf_Type ;
RUN;
OPTIONS Notes;
SASFILE MI_CLUs_Valid_State CLOSE ;
```


Distribution of CORN Validation CLU Sample by Size, MI, 2012
Number of CLUs (1,000)
Accuracy, percent
Buffer, outside
Buffered, inside
Entire CLU
Pixels per CLU
Plots: Bin counts Cumulative counts Producer, 1 std. error User, 1 std. error


Bootstrap sample by CLU size, entire CLU area
Sample size per 100 resamples
CLU counts
CLU size
0
100
200
300
400
500
0
10
20
30
40
50
Plots:
Sample size range
CLU counts

Distribution of Accuracy Estimate, 100 samples



			Buffer	All	Buffered
Beet	Prod	All	35,111	117,684	**82,573**
		Samp	956	1,283	1,503
	User	All	36,242	120,051	**84,138**
		Samp	1,058	1,389	1,609
Corn	Prod	All	877,621	2,304,310	**1,426,689**
		Samp	23,843	25,145	26,071
	User	All	743,715	2,166,907	**1,431,338**
		Samp	22,031	25,597	28,048
DryB	Prod	All	52,922	162,467	**109,545**
		Samp	1,436	1,771	1,998
	User	All	46,066	158,714	**113,169**
		Samp	1,373	1,859	2,184
Soyb	Prod	All	729,520	1,853,250	**1,123,730**
		Samp	19,807	20,214	20,485
	User	All	535,055	1,631,496	**1,101,338**
		Samp	15,613	18,967	21,209
Wwht	Prod	All	176,651	464,220	**287,569**
		Samp	4,790	5,065	5,254
	User	All	150,585	445,417	**296,943**
		Samp	4,562	5,366	5,917

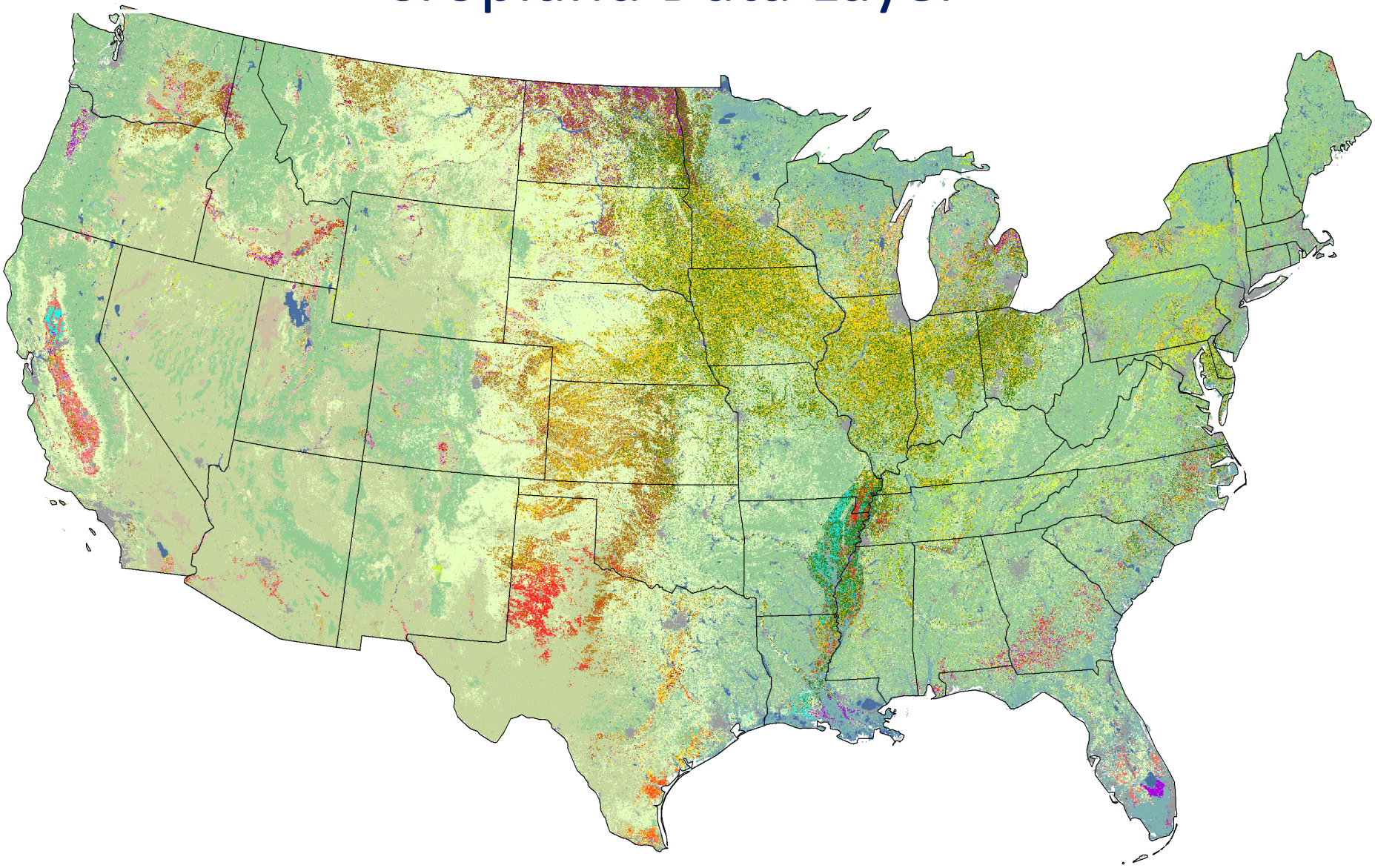
Sample size, pixels

Buffer, outside Buffered, inside Entire CLU

Conclusions

- Current method not statistically defensible
- PROC SurveySelect very useful
- PPS easily incorporated into CDL process
- GTL is very flexible and expressive

Cropland Data Layer



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