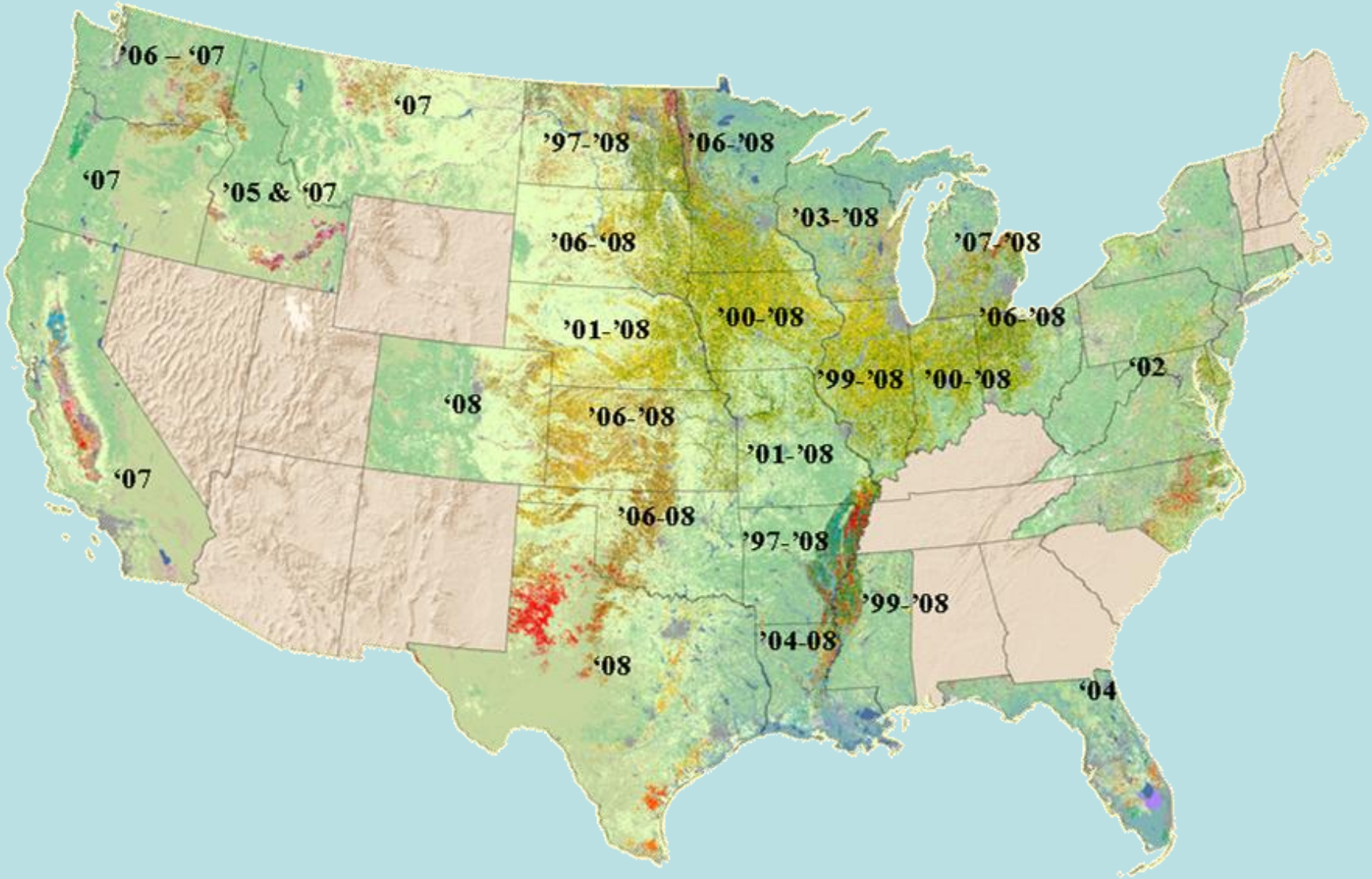
Fractional Factorial Method and Optimization of the Mississippi Cropland Data Layer

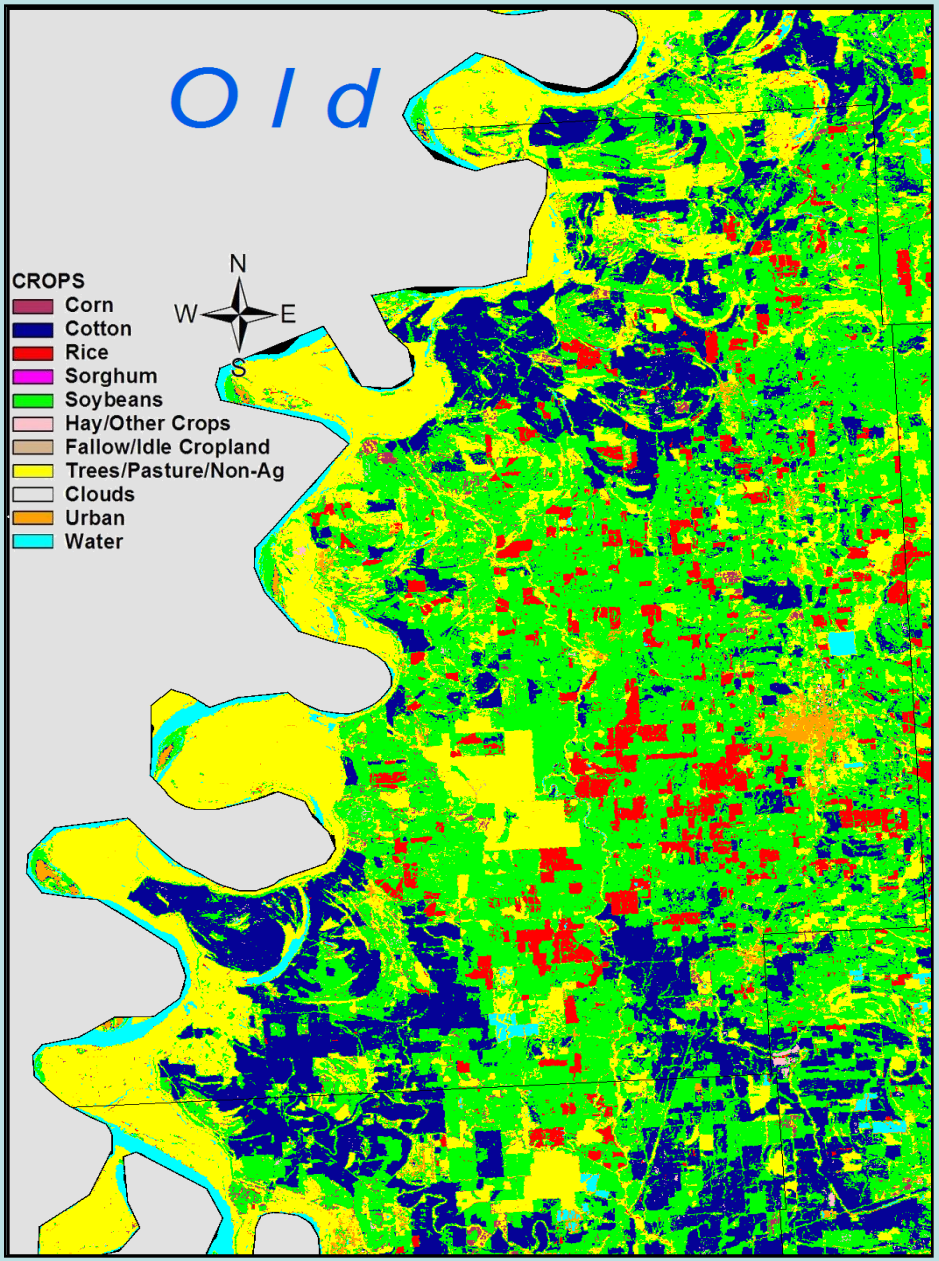
Fred L. Shore, Ph.D.
Mississippi Department of Agriculture and Commerce
Jackson, MS, USA
fred_shore@nass.usda.gov

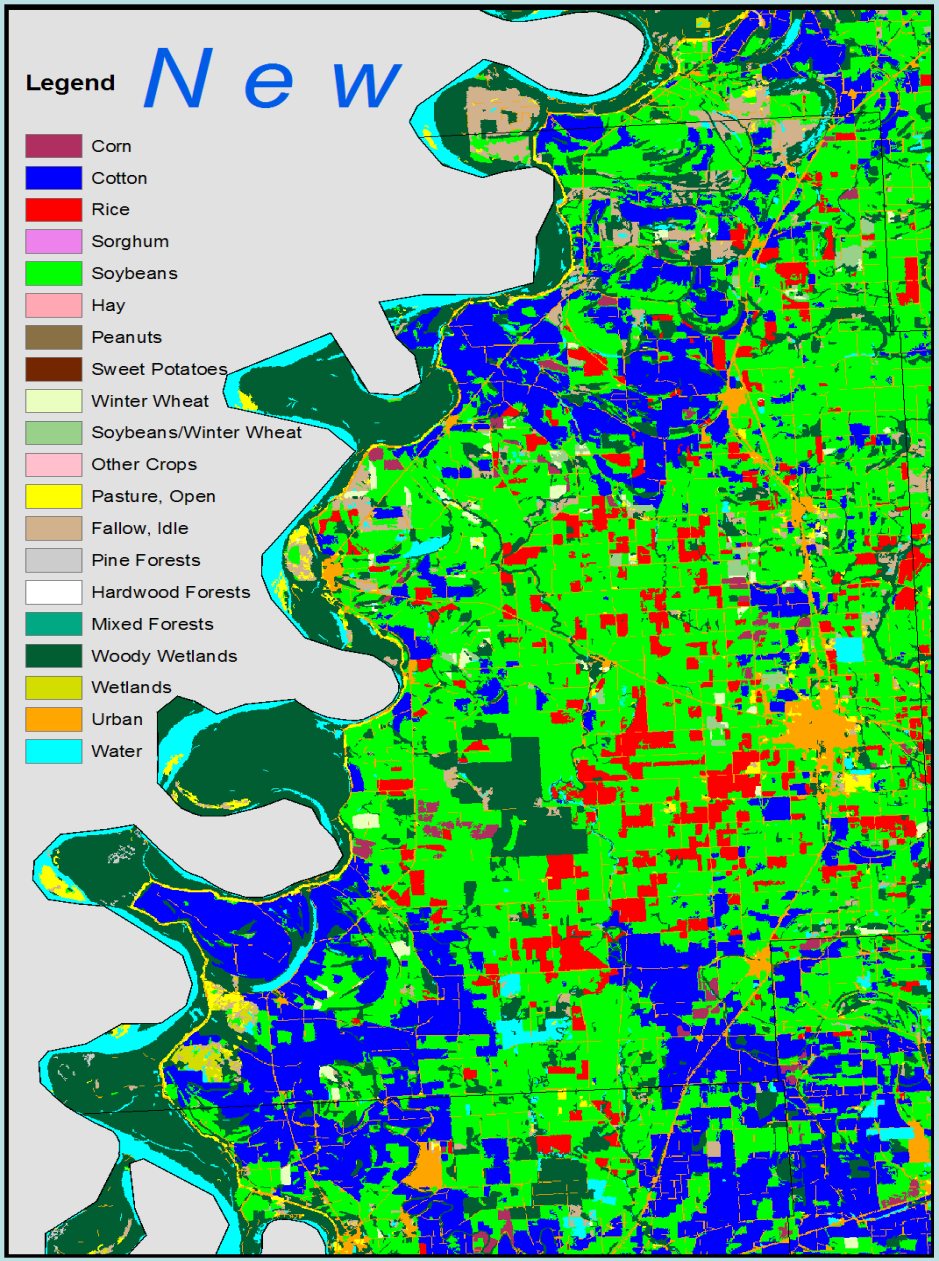
Thomas L. Gregory
National Agricultural Statistics Service
Jackson, MS, USA

Cropland Data Layers
1997 - 2008

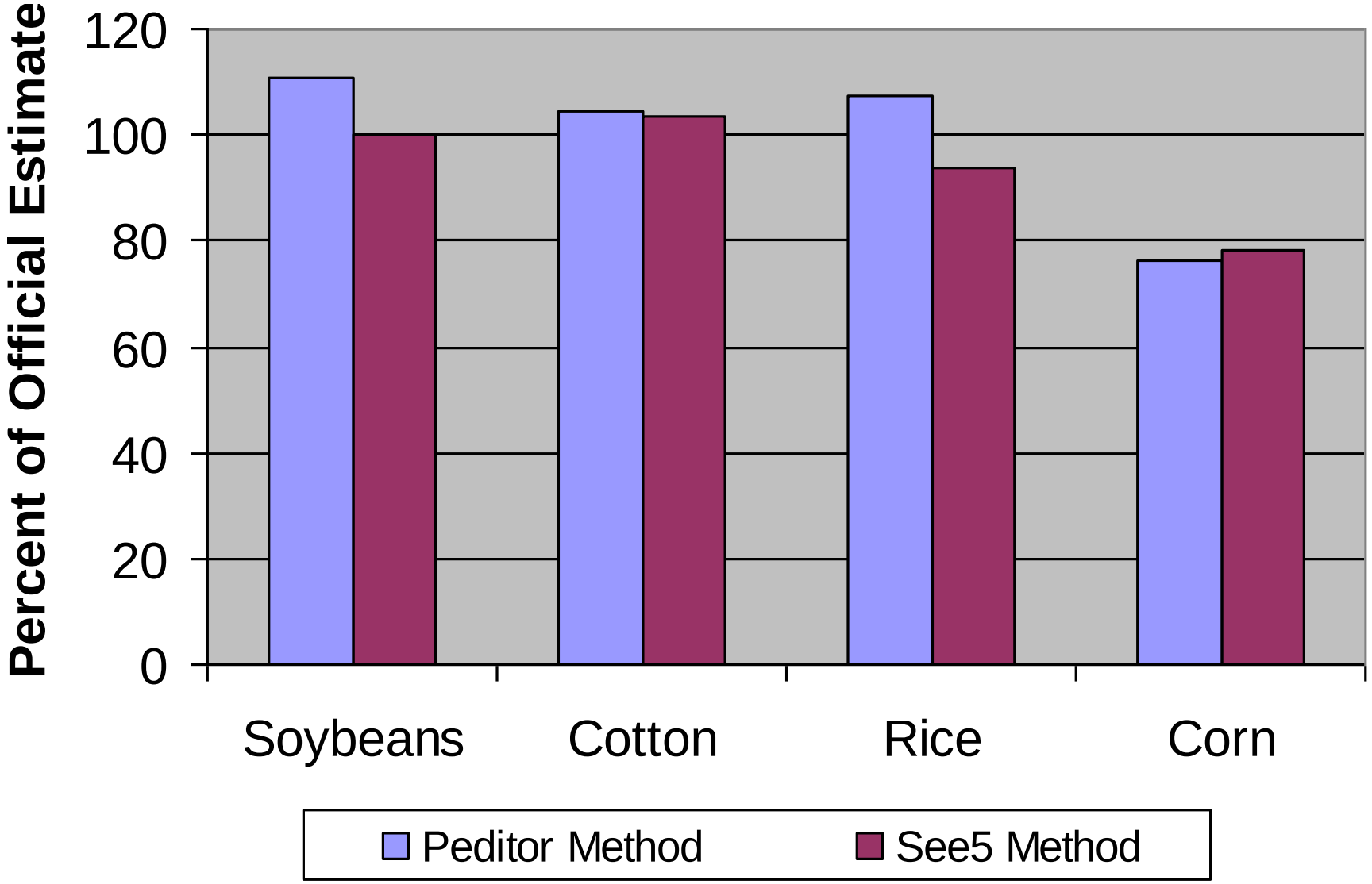


2006 CDL


Old
CROPS
Corn
Cotton
Rice
Sorghum
Soybeans
Hay/Other Crops
Fallow/Idle Cropland
Trees/Pasture/Non-Ag
Clouds
Urban
Water
N
W
E
S


New
Legend
Corn
Cotton
Rice
Sorghum
Soybeans
Hay
Peanuts
Sweet Potatoes
Winter Wheat
Soybeans/Winter Wheat
Other Crops
Pasture, Open
Fallow, Idle
Pine Forests
Hardwood Forests
Mixed Forests
Woody Wetlands
Wetlands
Urban
Water

2006 CDL Acreage Estimates vs. Official Estimates



Fractional Factorial Study, 2007 NASS See5 Method Test

- Test of seven variables.
- Establishes expected range of producer's accuracy.
- Identifies most important variables for optimization.

Fractional Factorial Study, 2007
Variable Levels per Run

Run	Variables						
	MODIS	AWiFS	Percent FSA Training	FSA Sampling Stratified Minimum Samples	Training Sample Points	NLCD Sampling Stratified Minimum Samples	Boost
1	1/2 MODIS	No Clouds	20	5	0.6Mp	10	7
2	MODIS	No Clouds	20	2	0.4Mp	10	10
3	1/2 MODIS	Clouds	20	5	0.6Mp	5	10
4	MODIS	Clouds	20	2	0.4Mp	5	7
5	1/2 MODIS	No Clouds	50	5	0.4Mp	5	10
6	MODIS	No Clouds	50	2	0.6Mp	5	7
7	1/2 MODIS	Clouds	50	2	0.4Mp	10	7
8	MODIS	Clouds	50	5	0.6Mp	10	10

Fractional Factorial Study, 2007
Crop Results (Percent Producer's Accuracy)

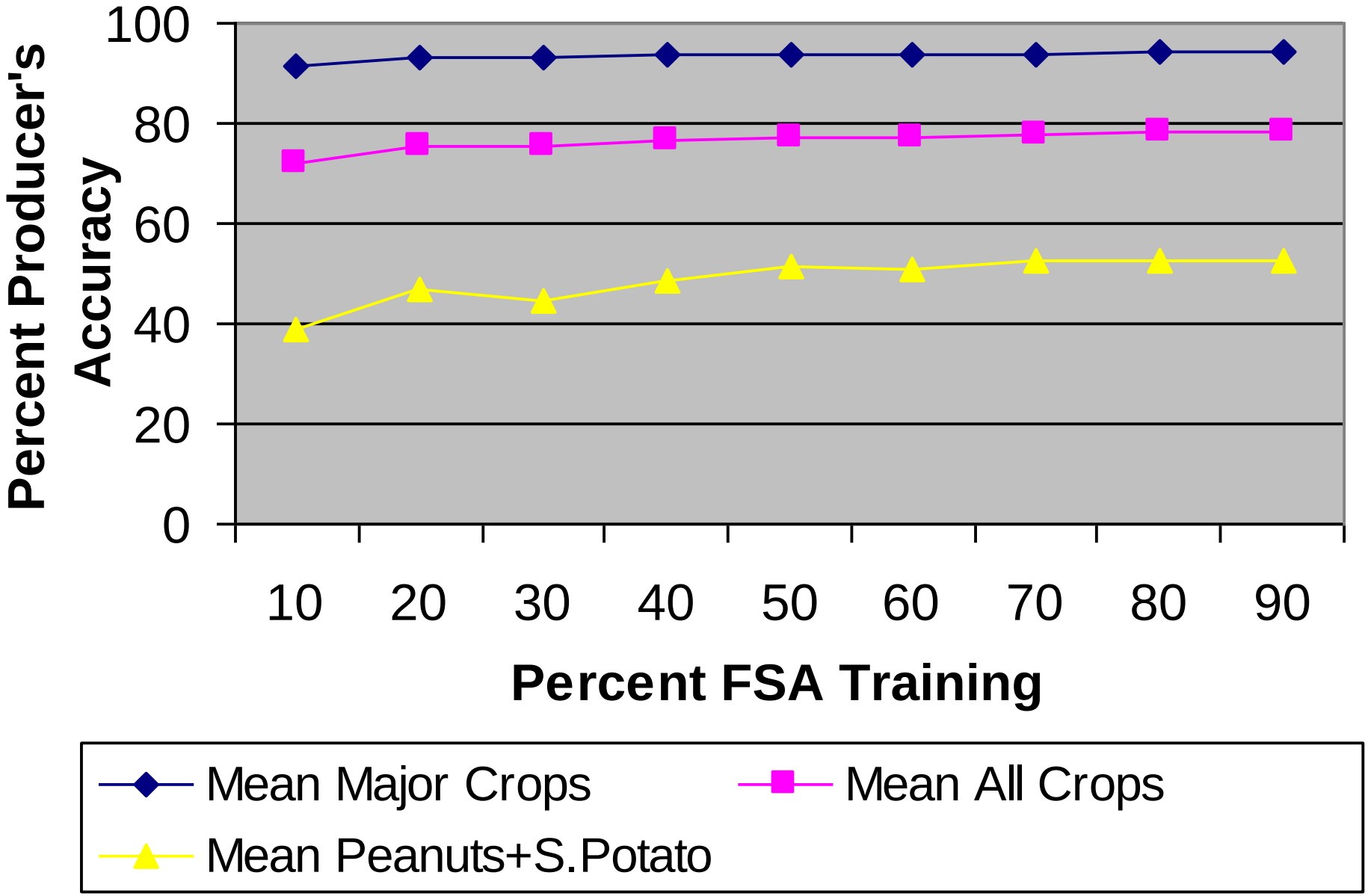
Crop	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Mean	StDev
Corn	89.7	91.6	90.6	91.3	92.7	92.0	92.5	92.5	91.6	0.994
Cotton	93.2	93.9	93.6	93.7	94.9	94.3	94.5	94.7	94.1	0.569
Rice	91.2	91.8	92.9	92.7	92.6	92.4	93.1	93.5	92.5	0.697
Soybean	91.3	92.7	92.0	92.2	94.0	93.1	93.6	93.7	92.8	0.88
Mean Major Crops	91.3	92.5	92.3	92.5	93.6	93.0	93.4	93.6	92.8	0.735
Peanuts	30.3	33.7	29.0	28.3	52.1	39.7	42.0	43.9	37.4	7.933
Sweet potatoes	43.6	45.7	45.6	46.8	51.0	48.5	51.5	48.6	47.7	2.58
Mean Peanut+ S. Potato	36.9	39.7	37.3	37.6	51.6	44.1	46.8	46.3	42.5	5.096
Sorghum	63.5	63.8	62.7	63.1	65.8	65.4	64.3	65.3	64.2	1.085
Winter wheat	61.5	63.3	62.8	61.2	64.9	64.0	63.5	65.0	63.3	1.305
Soybean/ Winter wheat	80.7	82.6	82.2	81.5	84.4	82.9	84.4	84.2	82.9	1.308
Mean All Crops	71.7	73.2	72.4	72.3	76.9	74.7	75.5	75.7	74.1	1.803

Fractional Factorial Study, 2007
Table of Differences

Variables	High Variable Levels				Low Variable Levels				Means Difference
	1	2	3	4	1	2	3	4	High-Low
MODIS	39.7	37.6	44.1	46.3	36.9	37.3	51.6	46.8	-1.238
AWiFS	37.3	37.6	46.8	46.3	36.9	39.7	51.6	44.1	-1.108
% FSA Training	51.6	44.1	46.8	46.3	36.9	39.7	37.3	37.6	9.313
FSA St. Min. Samples	36.9	37.6	51.6	46.3	39.7	37.3	44.1	46.8	1.128
NLCD Training Sample Points	36.9	37.3	44.1	46.3	39.7	37.6	51.6	46.8	-2.743
NLCD St. Min. Samples	36.9	39.7	46.8	46.3	37.3	37.6	51.6	44.1	-0.233
Boost	39.7	37.3	51.6	46.3	36.9	37.6	44.1	46.8	2.353

Fractional Factorial Study, 2007

Optimization of FSA Training



Fractional Factorial Study, 2007
FSA Training Study, Percent Correct

Crops	Range of Percent FSA Training		
	10-90	20-90	50-90
Mean Major Crops (StDev)	93.5 (0.902)	93.7 (0.422)	94.0 (0.148)
Mean All Crops (StDev)	76.4 (1.979)	76.9 (1.176)	77.7 (0.383)
Mean Peanuts+ S.Potatos (St Dev)	48.8 (4.711)	50.0 (3.041)	52.0 (0.917)

Fractional Factorial Study, 2007 Results

- Over the extreme range settings for the seven parameters varied, the mean producer's accuracy for the major crops was 92.8% with a standard deviation of 0.735.

- The most important parameter is the FSA training.

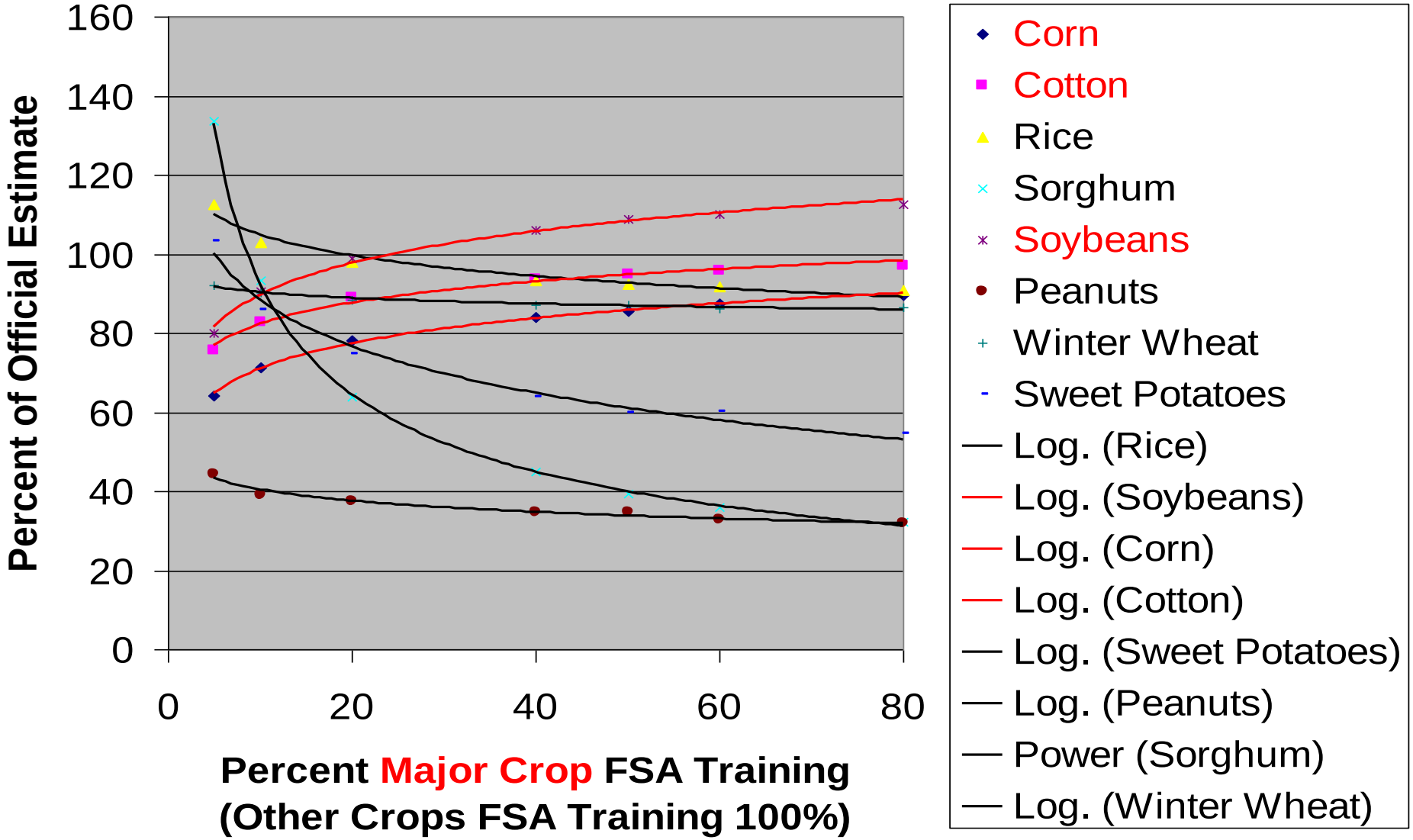
- Optimization of the FSA training gave an improved mean producer's accuracy of 94.0% and a standard deviation of 0.148 for the major crops.

- Minor crops require further optimization.

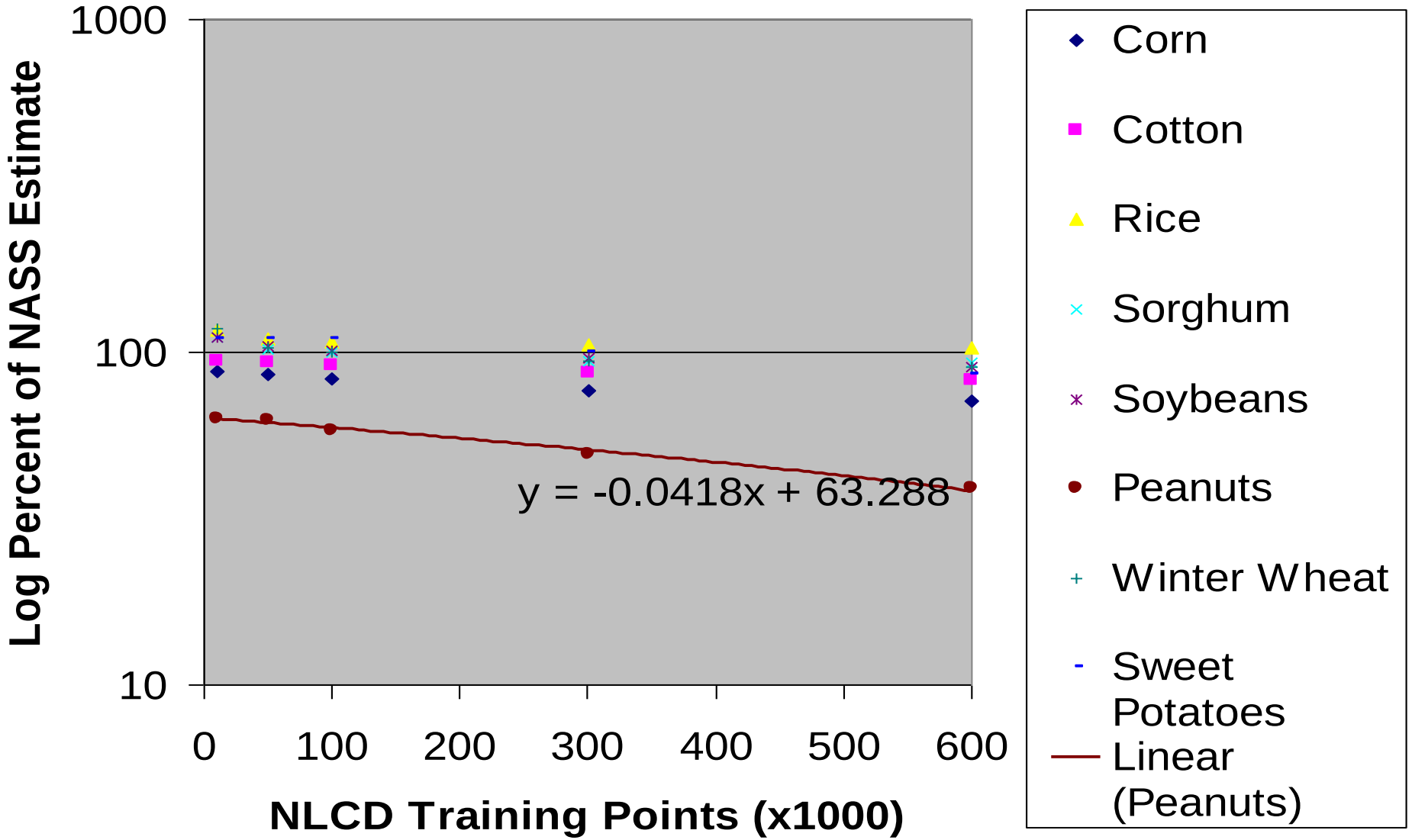
CDL 2007 Crop Acre Accuracy Retrospective Study vs. Official Results

- Pixel acre comparisons vs. official estimates.
- Study of optimization of parameters, starting with % FSA training.
- Main goal of the study is to improve performance of the classifier for low acreage crops and obtain an optimized CDL representative of all crops.

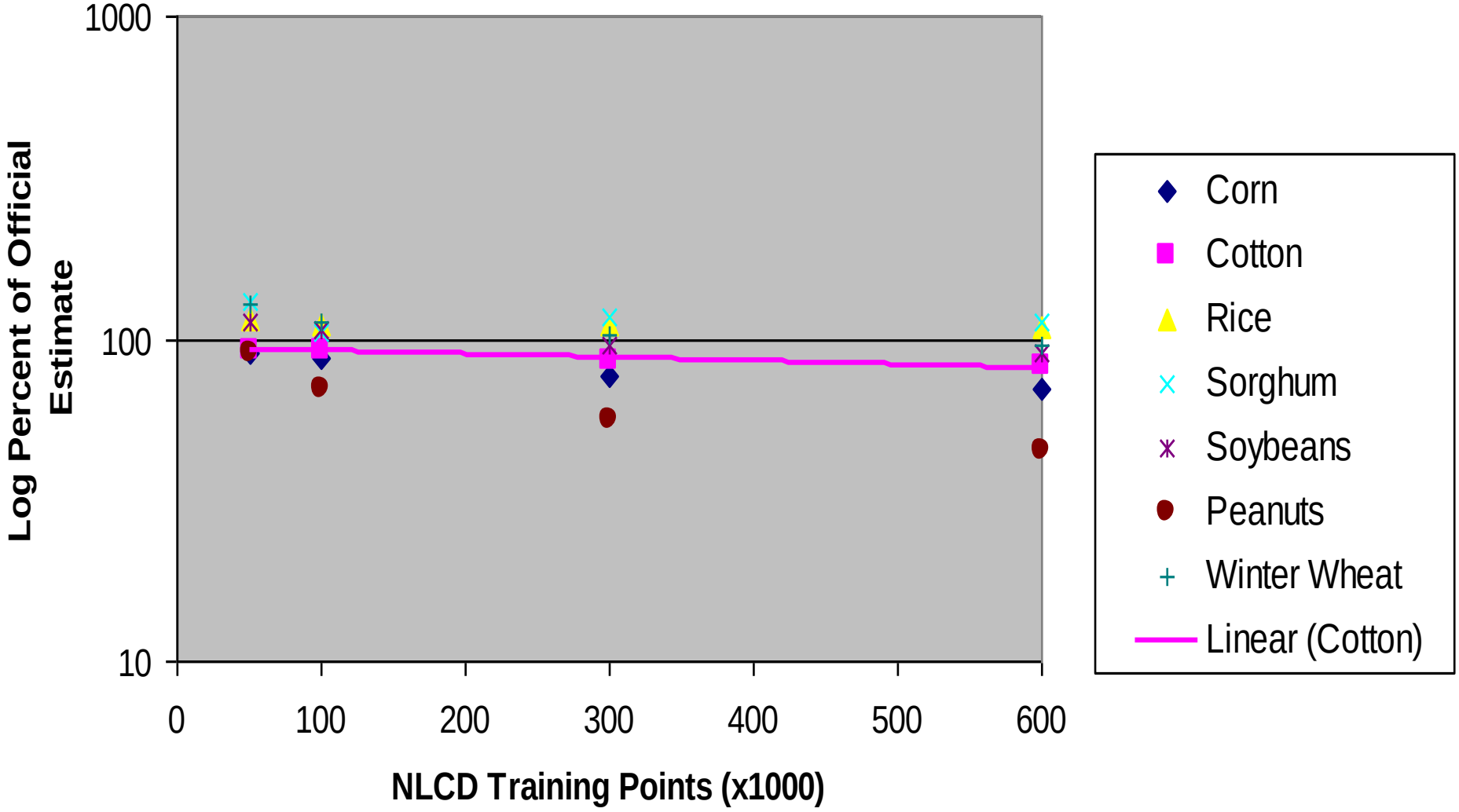
CDL 2007 Crop Acre Accuracy vs. Major Crop Training Data



CDL 2007 Crop Acre Acuracy vs. NLCD Training Points



CDL 2007 Crop Acre Accuracy vs. NLCD Training Points (Deleting Selected FSA Categories)



CDL 2007 Crop Acre Accuracy Results

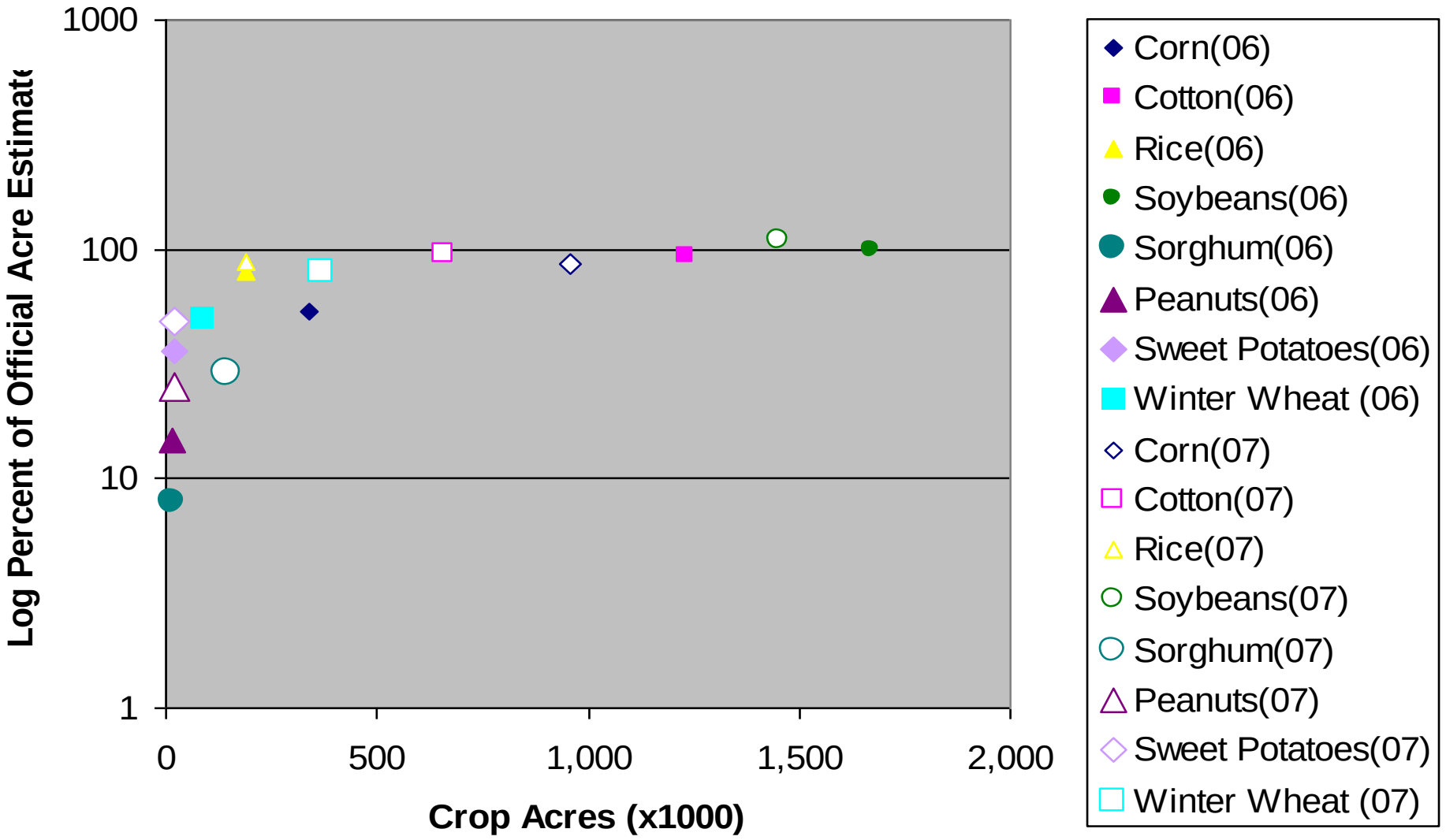
- Splitting the FSA training allows reduction of the major crops training relative to the minor crops and (with the exception of peanuts) improves the estimates for minor crops.

- Improvement for peanuts resulted from reducing the number of NLCD training points.

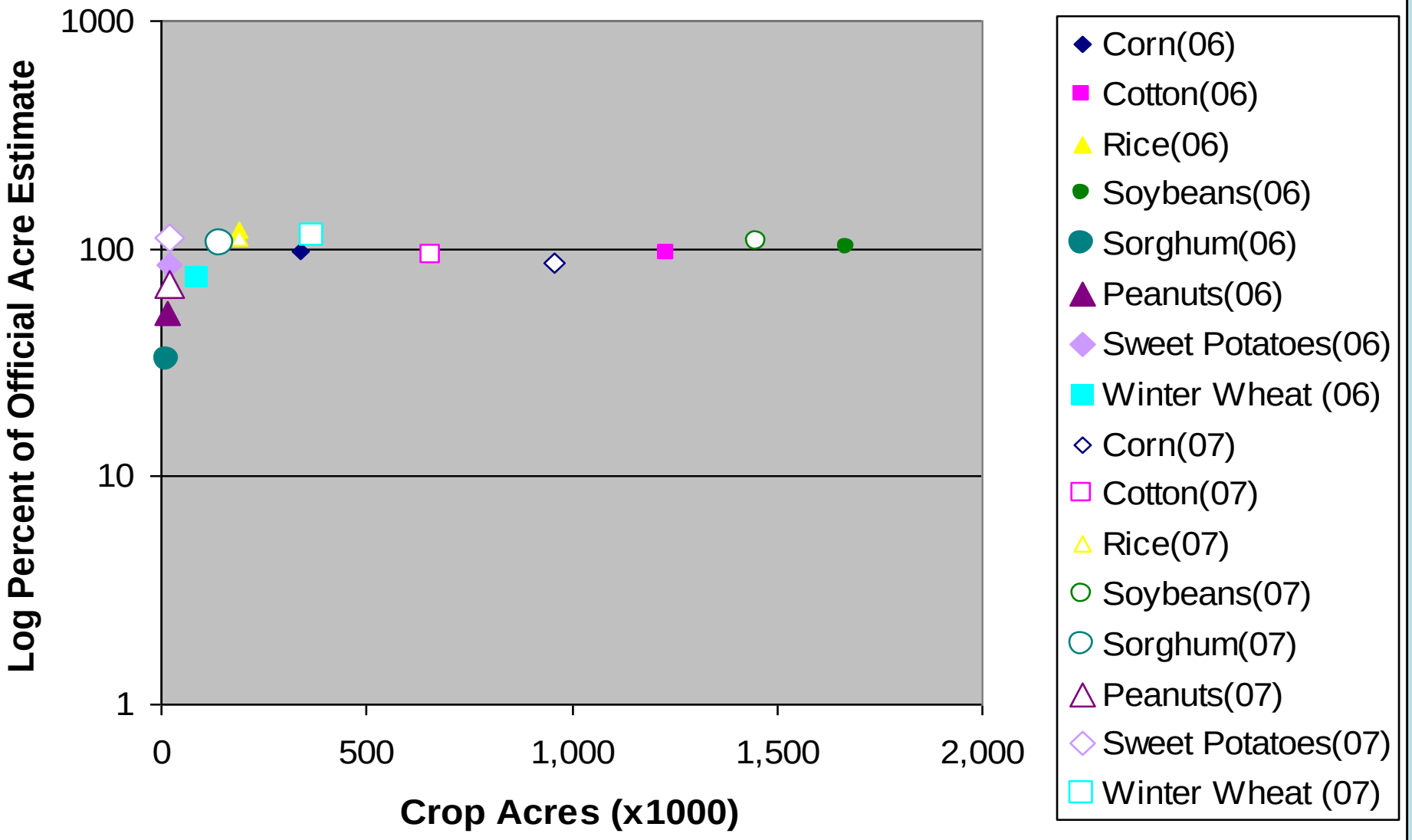
- The best overall results were obtained by reducing (a) the NLCD training points, (b) the major crop percent FSA training, and (c) the FSA categories.

- The final optimized 2007 CDL gave a mean result for pixel acres 100% of the official estimate with a standard deviation of 15.5. The peanut estimate was improved to 70% of the estimate. An optimized 2006 CDL was also produced using a similar method.

Crop Acre Accuracy vs. Acres
Published 2006/2007 Mississippi CDLs



Crop Acre Accuracy vs. Acres
Optimized 2006/2007 Mississippi CDLs



Winter Wheat Study, 2008
The First Mississippi Early Summer Remote Sensing Estimate

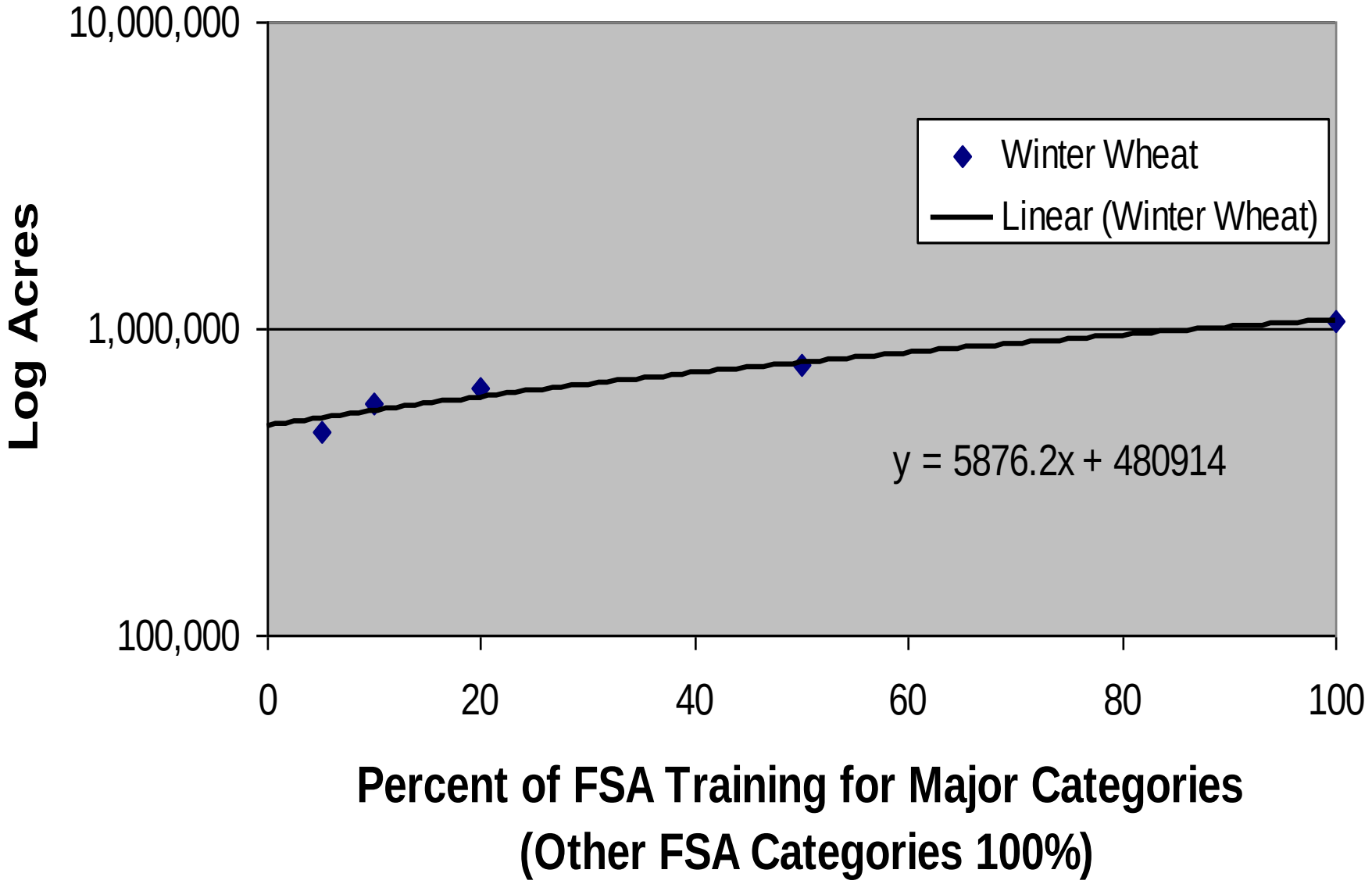
- Great AWiFS imagery covering the entire state twice in 3 mosaics. Very early dates for most crops (last date May 19).

- Great winter wheat FSA training data. FSA training data for other crops was very sparse and is absent for the north 1/3 of the Delta.

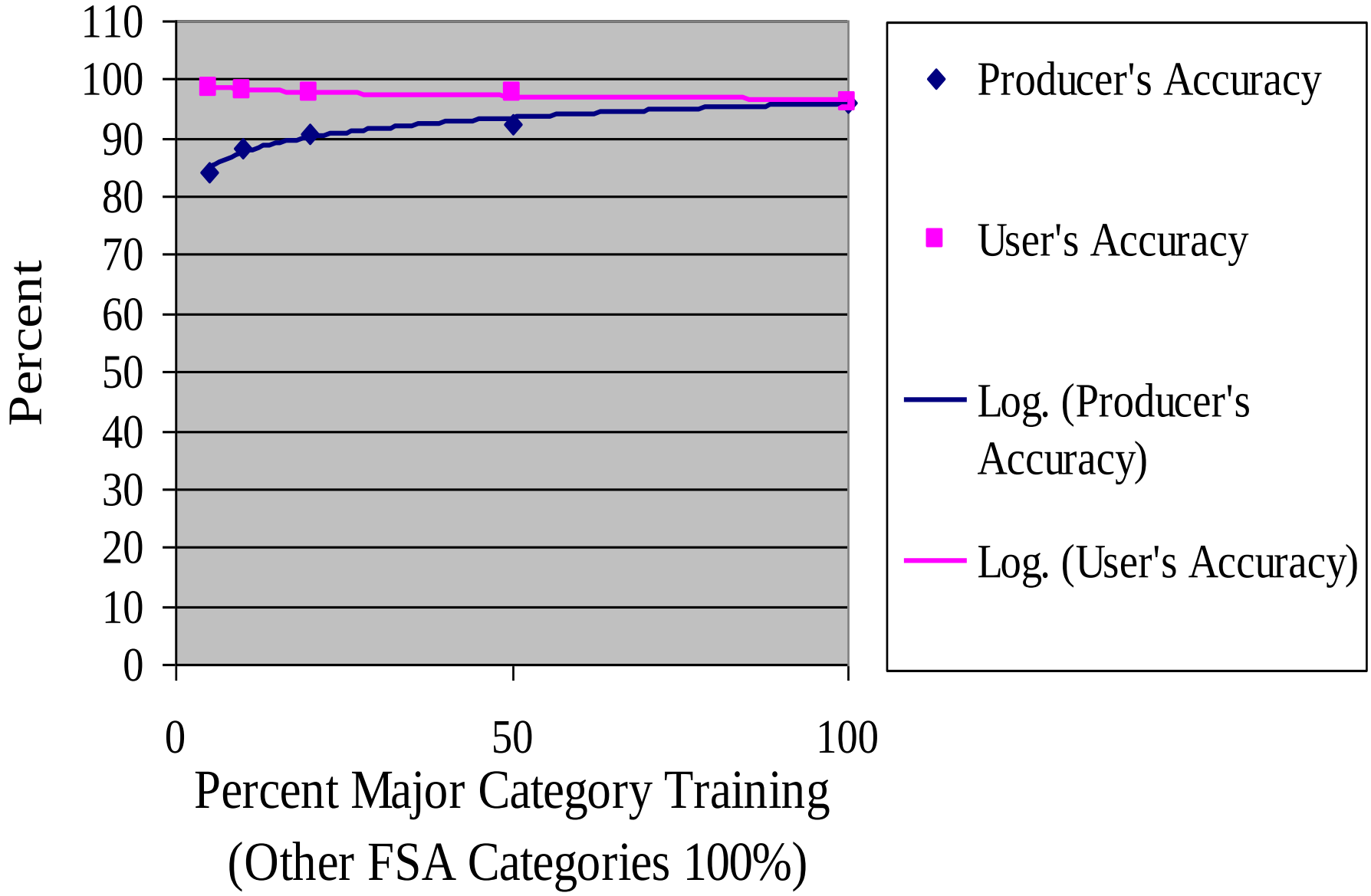
- Study varied the FSA training and used extrapolation to zero training acres.

- Tested producer's and user's accuracy.

Winter Wheat Study, 2008
Acres vs. FSA Training Data for Major Categories



Producer's and User's Accuracy for 2008 Winter Wheat Study



Winter Wheat Study, 2008 Results

- The estimate was prepared by June 12, 2008, before the June release of the official estimate.
- Extrapolation to zero FSA winter wheat training gave a planted acre estimate of 96.2% of the official estimate.
- The wheat acres decreased in a logarithmic fashion by a factor of >2 as the FSA major category split training was decreased to zero.
- The producer and user accuracies changed in opposite directions as the major category FSA training was lowered with an average accuracy always greater than 90%.

Discussion

- Testing the NASS See5 Method with a fractional factorial study over a wide range of parameter settings, generally gives excellent producer and user accuracy and acreage estimates for the major crops.

- The FSA training is most important in improvement and can be modified by splitting the training and reducing the categories with the most training pixels.

- The NASS See5 Method has been optimized with a method to include accurate minor crop estimates using adjustment of (a) the percent FSA training, (b) the NLCD training, and (c) the FSA categories used.

- An additional extension of the NASS See5 Method was required for winter wheat estimation. Training data at the required early date was not available for major crops causing a serious over estimate by the standard method. However, extrapolation of a semi-log plot to zero FSA training gives a winter wheat acreage result unbiased by overfitting.

Acknowledgements

Commissioner Lester Spell, Jr., D.V.M., Mississippi Department of Agriculture and Commerce, and Melissa J. Mixon, Interim Director, Mississippi Cooperative Extension Service, were critical to the success of this project. Also, thank you to Rick Mueller, Claire Boryan, Dave Johnson, and other members of USDA-NASS, Spatial Analysis Research Section, Fairfax, VA for training and assistance.