

JACIE 2007

Evaluation of AWIFS Classifiers  
for Crop Acreage Estimation

Michael E. Craig  
USDA / NASS

[Mike\\_Craig@nass.usda.gov](mailto:Mike_Craig@nass.usda.gov)

# Evaluation of AWIFS Classifiers for Crop Acreage Estimation

- Comparisons created for the following:
  - Max. Likelihood vs. Regression Tree
  - Use/Level of Smart Eliminate smoothing
- Statistics available for above evaluations
  - Kappa Values
  - Regression R-square & Slope

# Evaluation of AWIFS Classifiers for Crop Acreage Estimation

- Mississippi River Delta Region
  - State of Arkansas, 2006 AWIFS imagery
- “Old” Standard – PEDITOR approach
  - Maximum Likelihood classification
  - Supervised ISODATA clustering
- “New” Approach – See5 with ERDAS
  - CART – Regression Tree classification
  - Smart Eliminate (S.E.) smoothing available
  - Burn-in of ‘administrative’ data (NLCD)

## Arkansas 2006 Ground Data

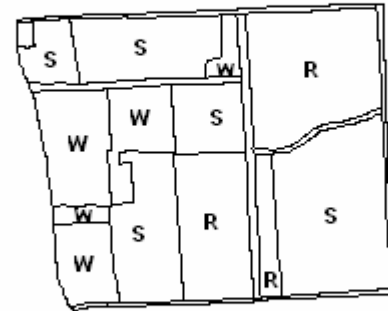
Stratum	# Segs Popn	# Segs Sampled	Expans. Factor
11*	11669	232	50
21*	2718	32	85
31	1308	4	327
32	418	2	209
42	18571	56	332
50	35	2	18
State	34719	328	

\* = approximately 1 square mile each

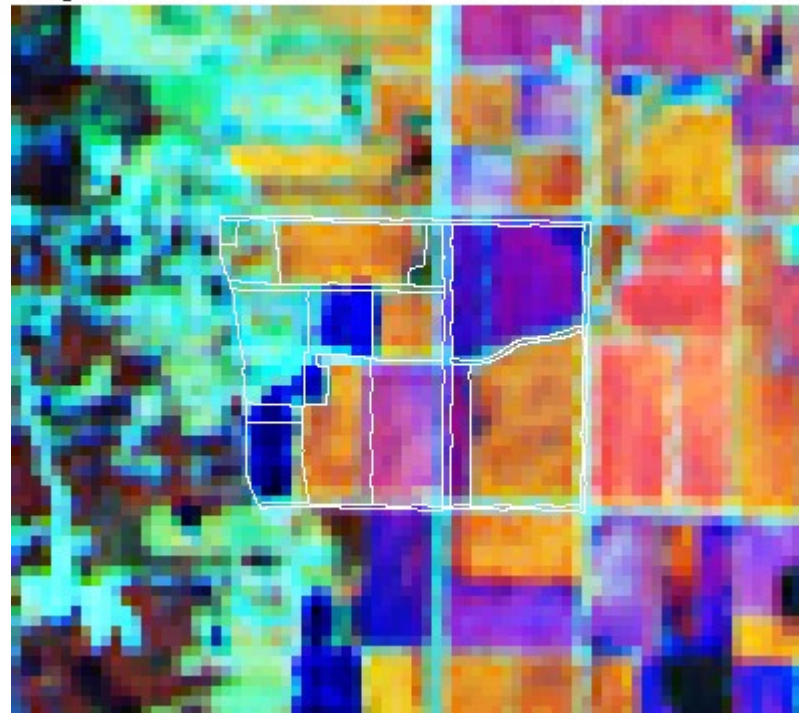


Segment 136

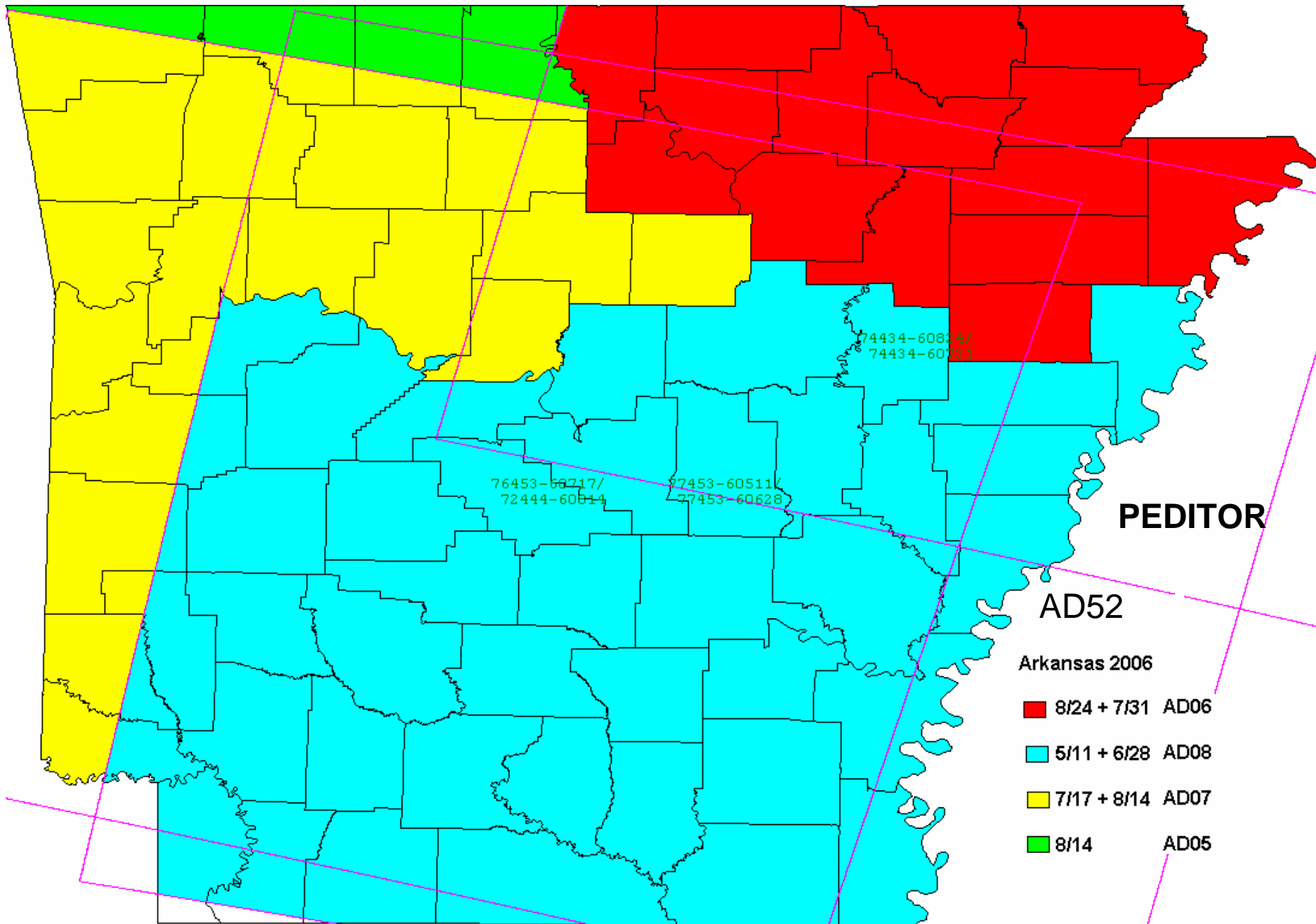
R = Rice  
 S = Soyb  
 W = Waste/FS



Segment 136 R=0.00 C=0.80 74434-60824/74434-6073



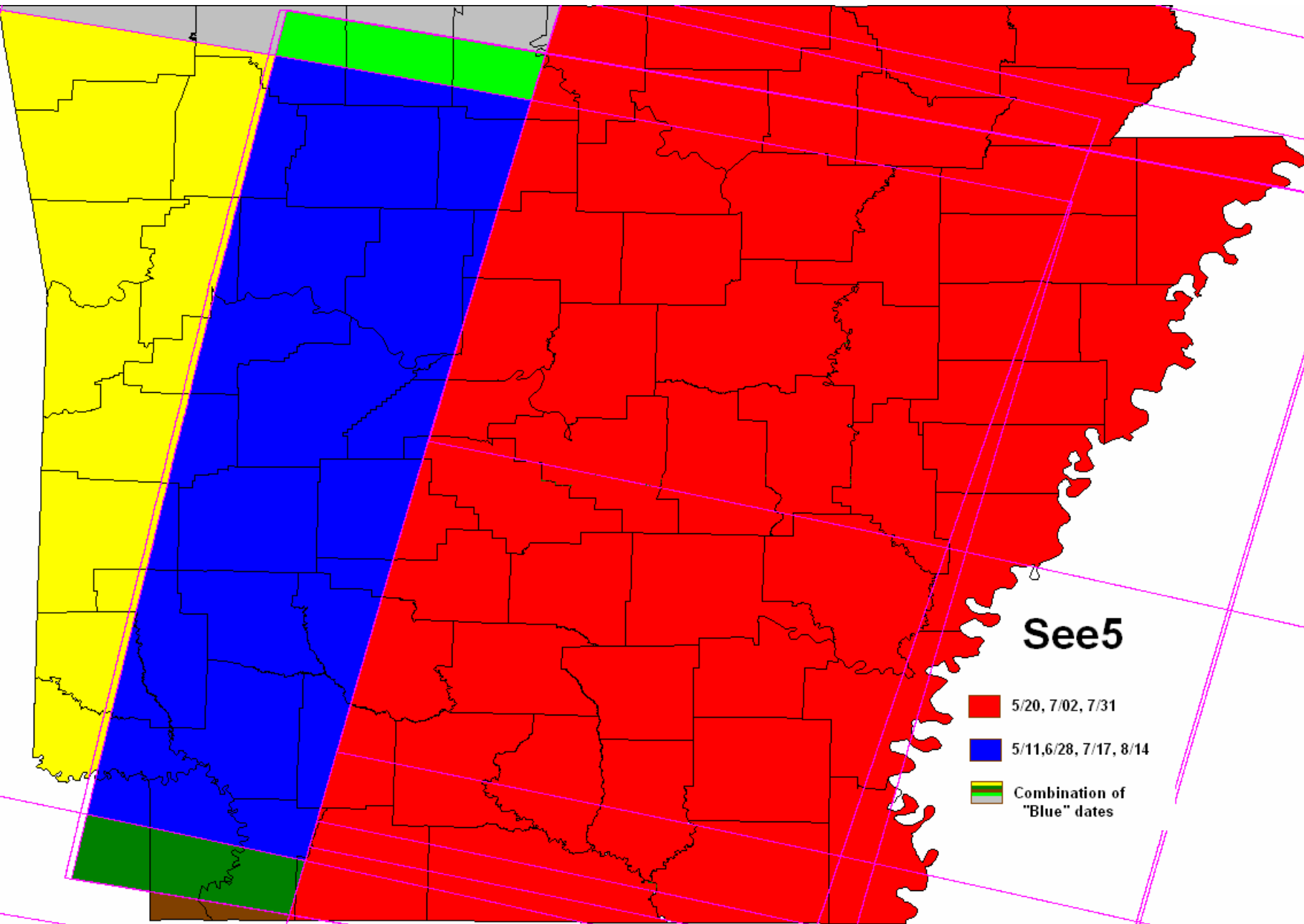
Crop Type	Y	
	Enumerated JAS Acres	
Rice	227.0	
Soybean	337.0	



ARO6 3 scenes from scene files

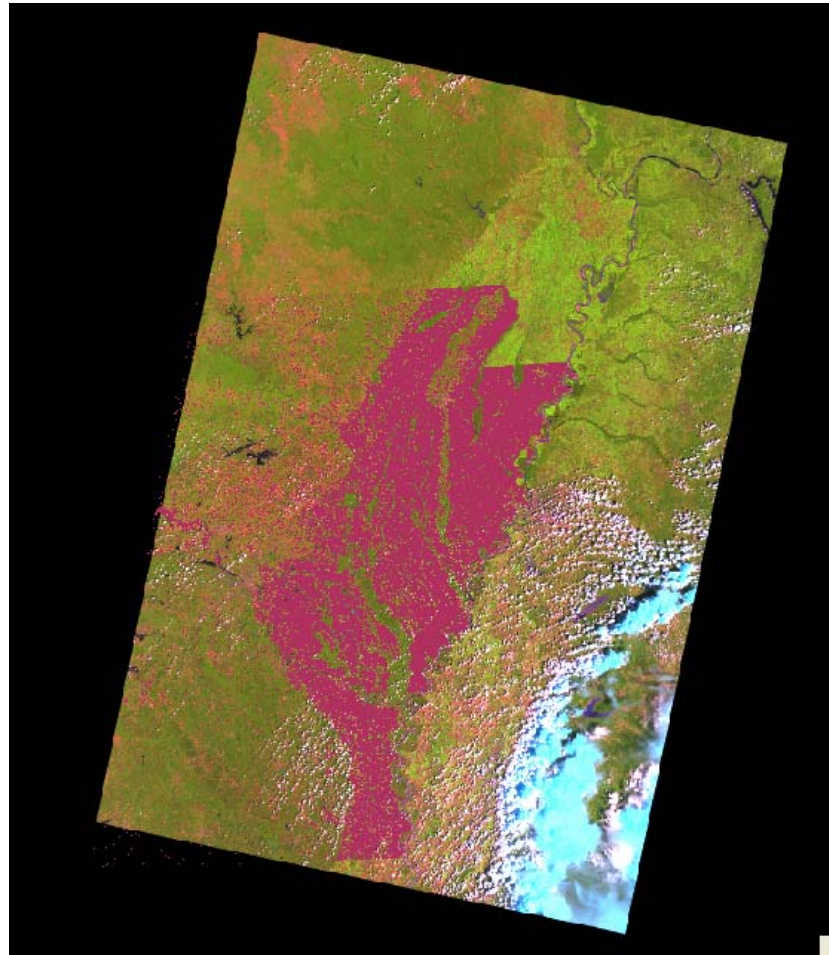
# AD51 See5 Specifications

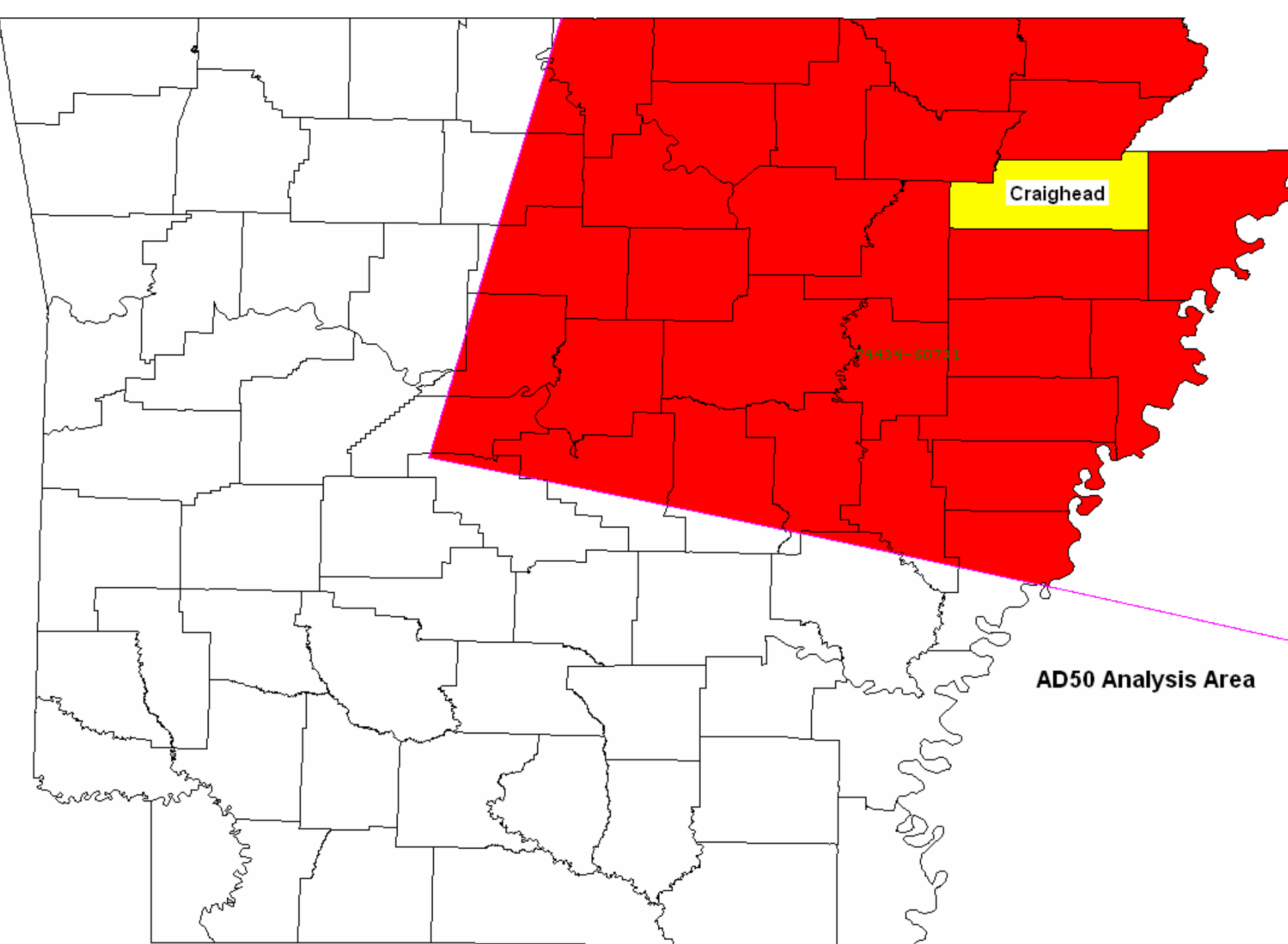
- Entire State Coverage
  - Used 10 scenes in 2 ‘runs’ (priority Run 1)
    - Run 1: 2 scene mosaics of dates 5/20, 7/02, 7/31
    - Run 2: 4 separate dates: 5/11, 6/28, 7/17, 8/14
  - Crop mask generated from 7 previous CDL’s
    - Pixels having 3 or more ‘crop’ classes saved
    - See5 classification performed in crop mask only
  - Areas outside of crop mask ‘burned in’
    - NLCD 2001 (or 1992) Landcover Data





# AD51 – Run 1 Area (7/31) AR Crop Mask in Pink





Craighead

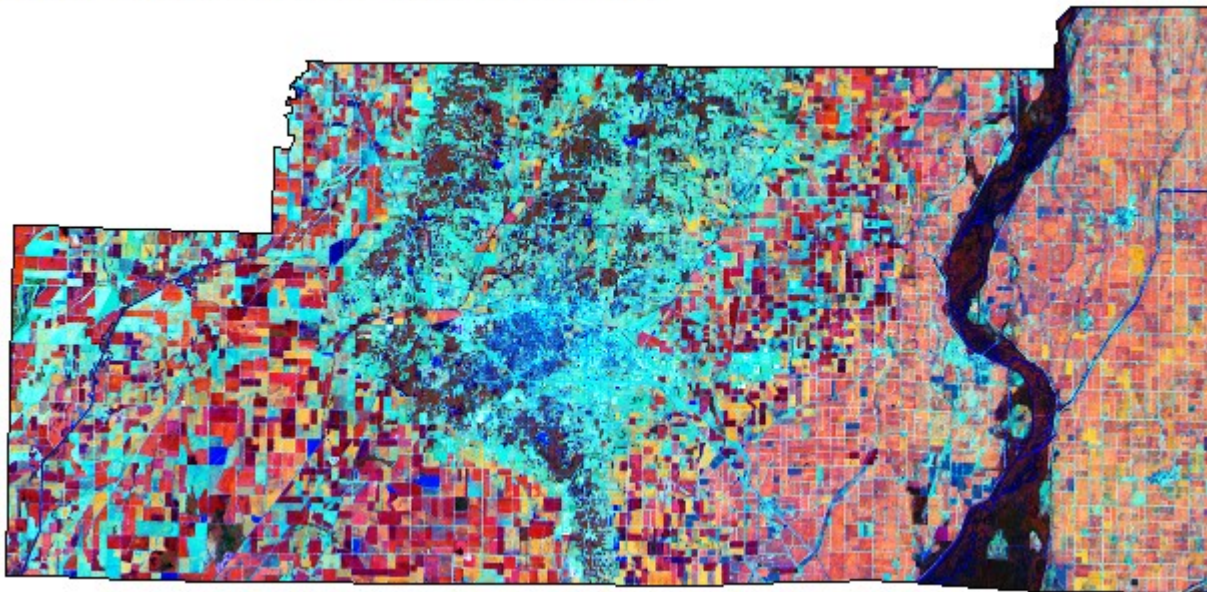
4434-60751

AD50 Analysis Area

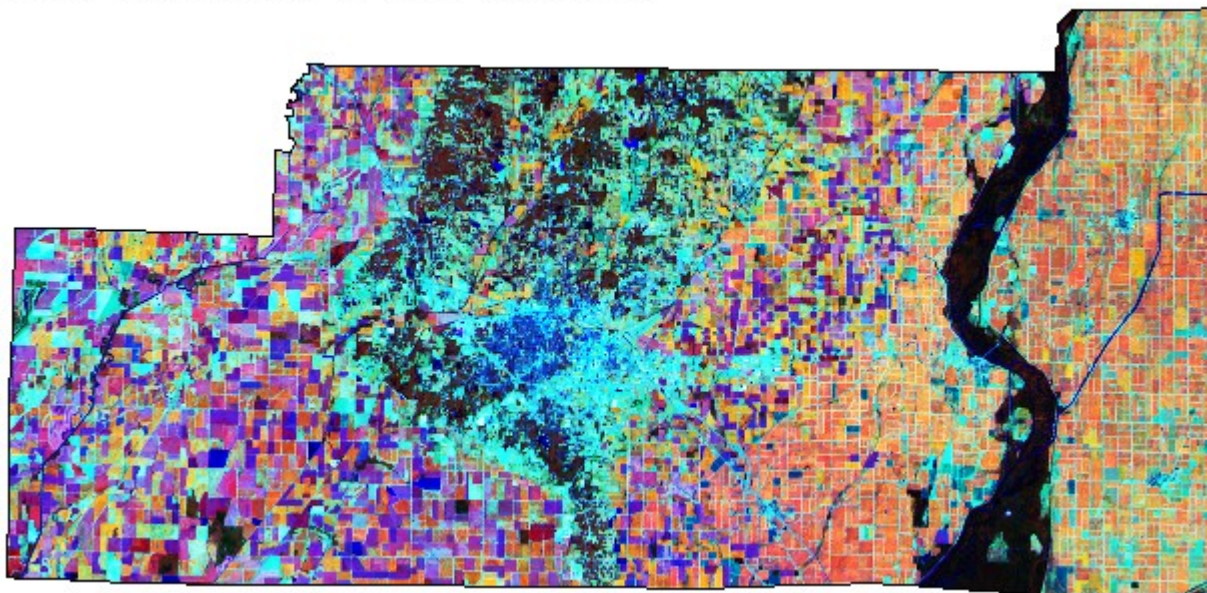
# (9) AD50 Datasets Available

- (3) 8-bit data, no crop mask, no NLCD burn-in
  - See5 with no S.E., S.E.=20, S.E.=90
- X(3) 10-bit data, no crop mask, no NLCD burn-in
  - See5 with no S.E., S.E.=20, S.E.=90
- (1) 8-bit data, no crop mask, no NLCD burn-in
  - See5 with no S.E., has additional NLCD'92 training
- (1) 8-bit data, w/crop mask & NLCD burn-in
  - See5 with S.E.=45 (approximate AD51 subset)
- (1) PEDITOR AD52, matching area subset

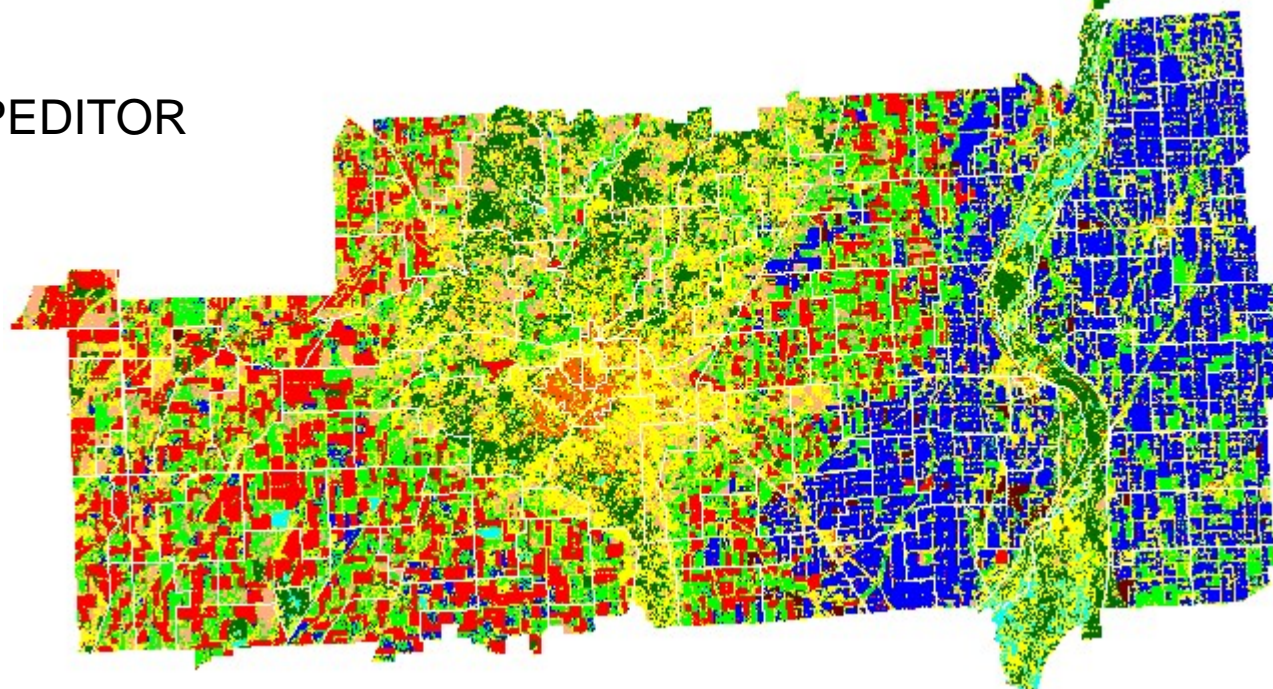
County CRAIGHEAD(31) AR Scene 74434-60731



County CRAIGHEAD(31) AR Scene 74434-60824



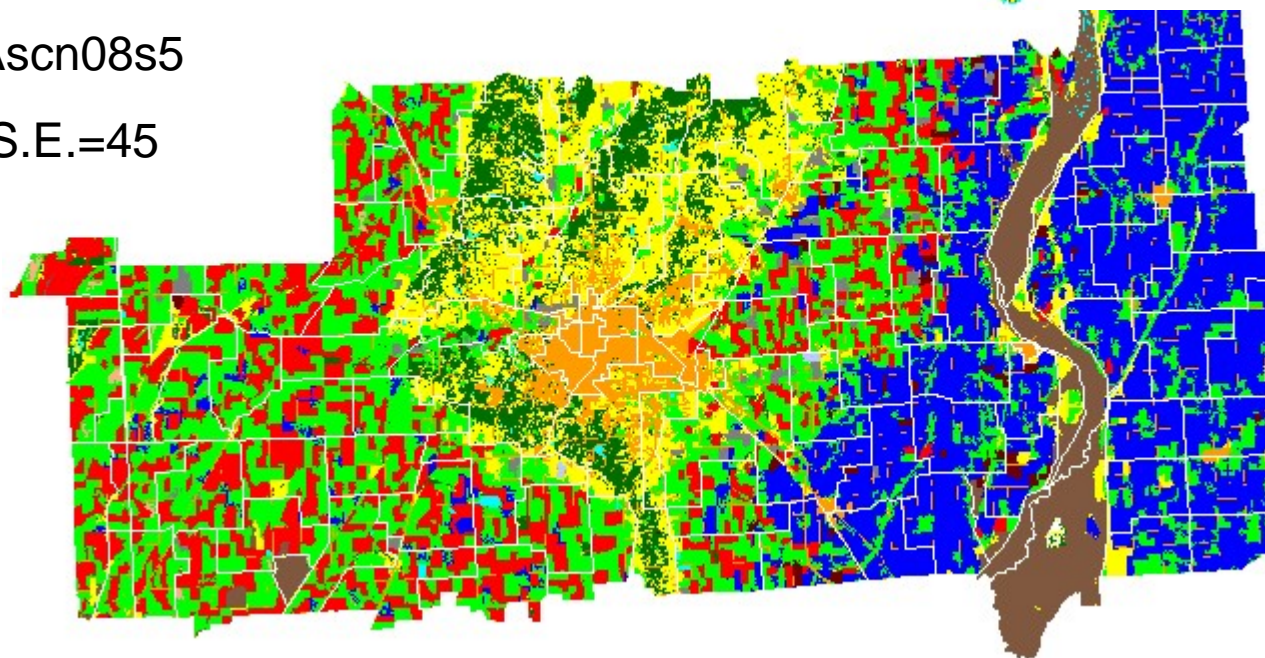
PEDITOR



- COTTON
- RICE
- WATER
- SORGHUM
- WIN WHEAT
- Non Ag
- URBAN
- IDLE CROP
- CLOUDS
- WOODS
- CORN
- OTHER GRN

Asc08s5

S.E.=45



- RICE
- WATER
- SORGHUM
- WIN WHEAT
- Non Ag
- Rock Sand
- Urban
- IDLE CROP
- Wetlands
- WINWSOYB
- CLOUDS
- WOODS
- CORN
- OTHER GRN

# PEDITOR vs See5 Kappa's AD50 & Statewide

- AD50 Kappas:

CSN	CORN	COTT	RICE	SOYB	OCRCP	NONAG	WOODS	OVERALL
3scn0800	80.14	87.44	93.14	86.21	52.35	64.08	88.60	83.57
3scn08cg	81.91	87.93	93.26	86.26	52.84	66.54	83.53	83.25
Asc08s5	71.34	91.59	89.17	91.67	44.18	56.68	47.86	75.49
Asc08pe	62.10	68.01	80.30	53.42	24.08	61.47	59.67	62.67

- Statewide Kappa's:

CSN	CORN	COTT	RICE	SOYB	OCRCP	NONAG	WOODS	OVERALL
Asc05no	71.94	85.03	85.46	83.16	49.66	61.49	42.85	69.85
Asc0520	75.73	90.14	89.31	89.76	50.38	62.04	42.42	72.86
Asc05se	71.41	90.62	87.45	91.21	47.36	59.70	42.41	72.39
Asc08pe	61.68	71.03	80.00	55.47	25.09	61.09	65.00	64.40

# Review of Kappa Values for PEDITOR versus See5

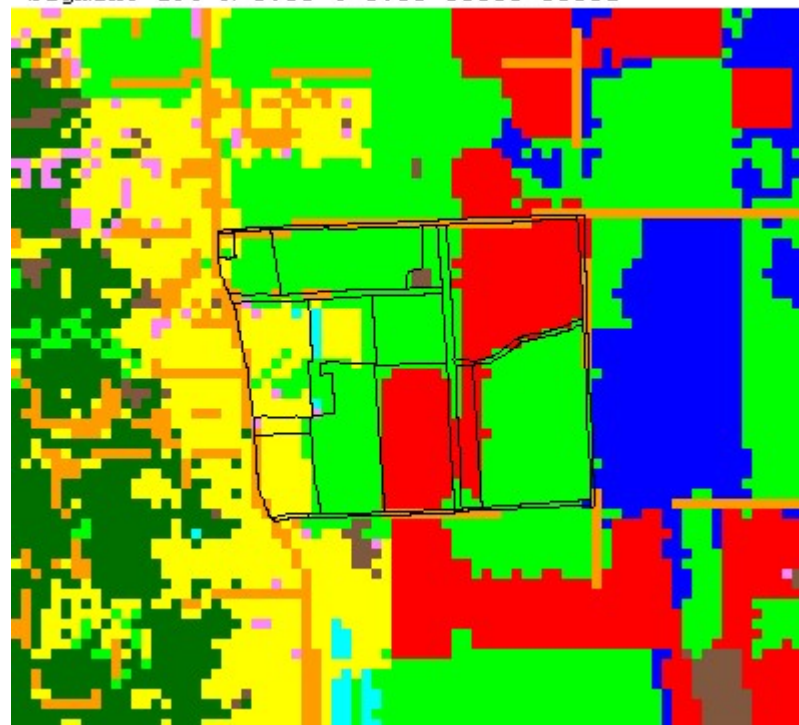
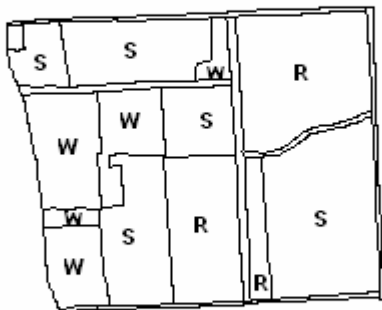
- AD50
  - Almost all See5 values > PEDITOR values
    - Exception Asc08s5 (S.E.=45) for 'woods' & 'non-ag'
      - These two categories are burnt-in from NLCD!
    - See5 'Overall' values always greater, significantly so!
- Statewide (AD51 vs AD52)
  - Almost all See5 values > PEDITOR values
    - Exception 'woods' in all cases, 'non-ag' in 1 case
    - Although See5 'Overall' values are still greater, there is not so much disparity between them as was seen in AD50

# Regression Estimator (Evaluated at the 'Segment' Level)

- Regression used to relate categorized pixel counts to the ground reference data
  - Independent variable - satellite data - pixels
  - Dependent variable - JAS acreage estimate
- Satellite data - lower variance than with only JAS
- Outlier segment detection - correction or removal from regression analysis

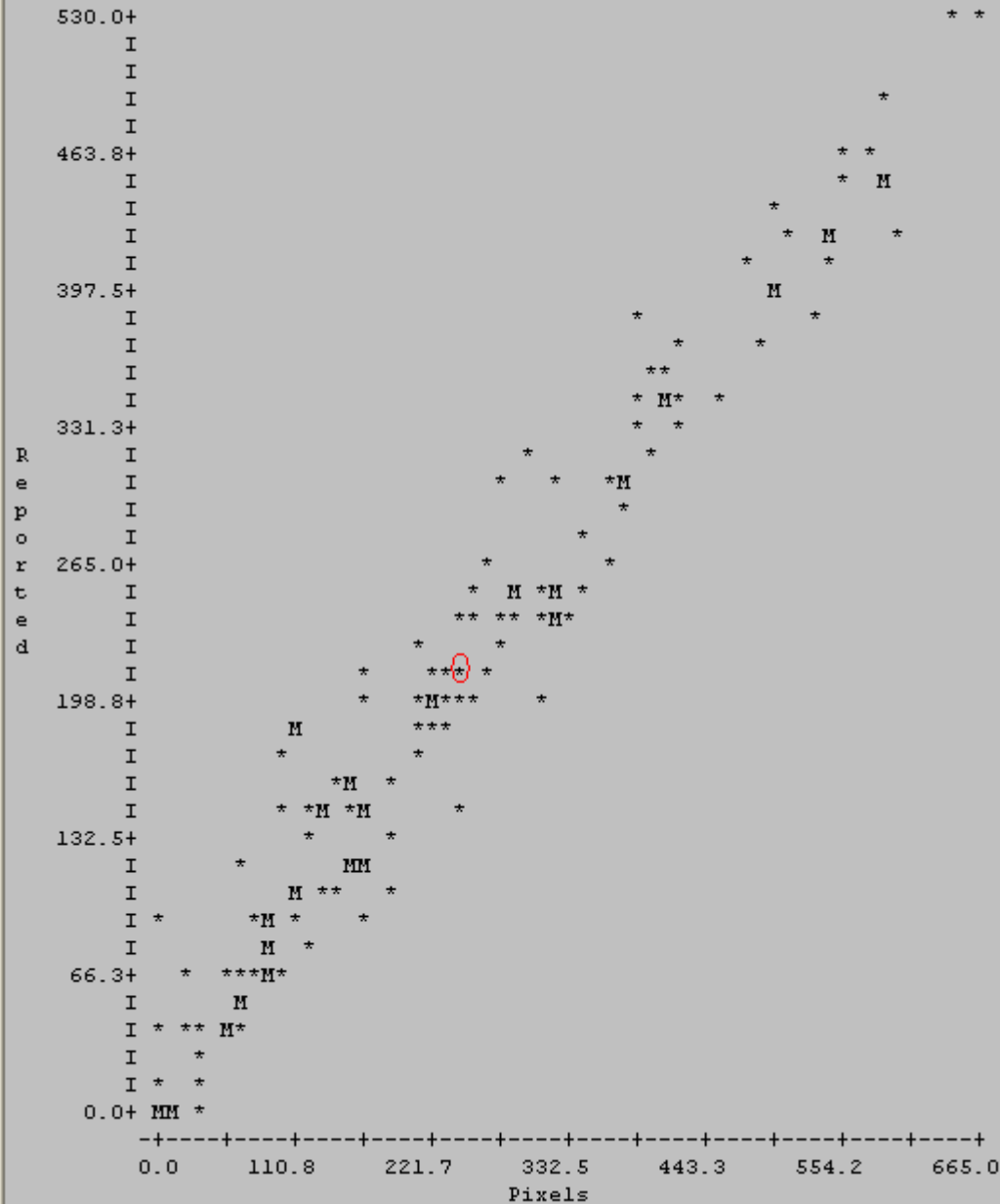


R = Rice  
 S = Soyb  
 W = Waste/FS



Crop Type	Y	X
	Enumerated JAS Acres	Classified Pixels
Rice	227.0	273
Soybean	337.0	541

Plot of RICE Pixels(X) Reported(Y)  
Strata 11



$$R^2 = 0.971$$

$$a = \text{intercept} = 7.11$$

$$b = \text{slope} = 0.802$$

-----

Linear Regression

$$y = a + bx$$

-----

Seg 136 (x=273, y=227)

$$y = 226.11$$

# R2 Values Before & After Outlier Deletes (“Best” See5 v. PEDITOR)

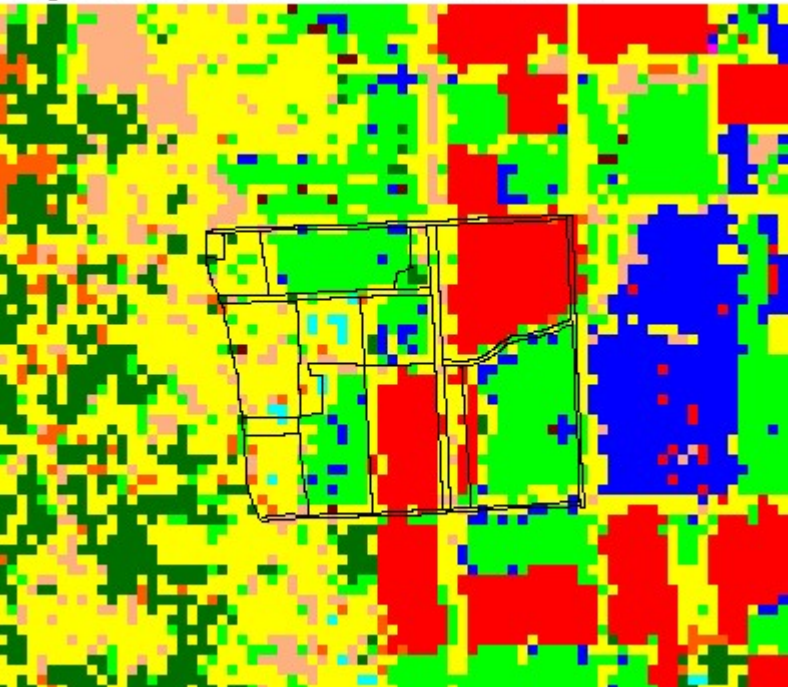
		AD50		Statewide	
		Before	After	Before	After
<b>Rice</b>	See5	.942	.981	.936	.971
	Pedt	.908	.958	.894	.948
<b>Cotton</b>	See5	.981	.991	.970	.984
	Pedt	.845	.909	.838	.920
<b>Soybean</b>	See5	.860	.897	.848	.901
	Pedt	.674	.720	.661	.673
<b>Corn</b>	See5	.957	.978	.915	.974
	Pedt	.753	.734	.653	.734

# R2 Comments

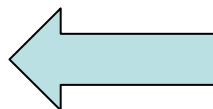
## PEDITOR vs See5

- Any See5 approach used is always better than the corresponding PEDITOR
  - “Before” deletes, at AD50 level, the one with combined NLCD & JAS is  $\geq$  any other See5
    - A similar statement can be made for S.E.=20 at the state level (i.e. better than no S.E. or S.E.=45)
  - “After” deletes, at AD50 level, there is no clear favorite among See5 approaches
    - At the state level, one and usually both smart eliminate approaches (S.E.=20,45) are better than the no smart eliminate value

Segment 136 R=0.00 C=0.00 55555-00002



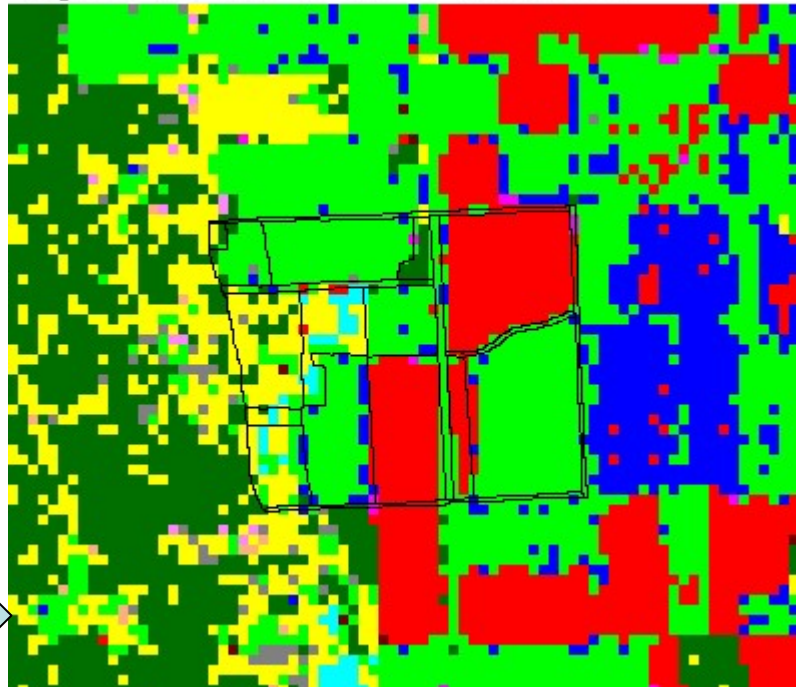
PEDITOR



AD50 See5  
Comb Gtruth

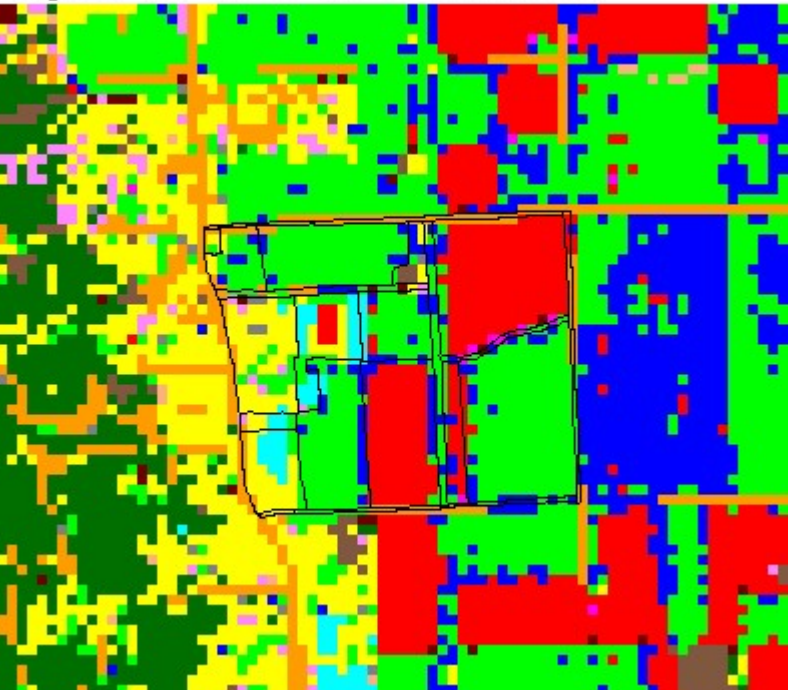


Segment 136 R=-0.20 C=0.80 74434-00005

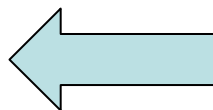


Segment 136 R=0.00 C=0.80 55555-00001

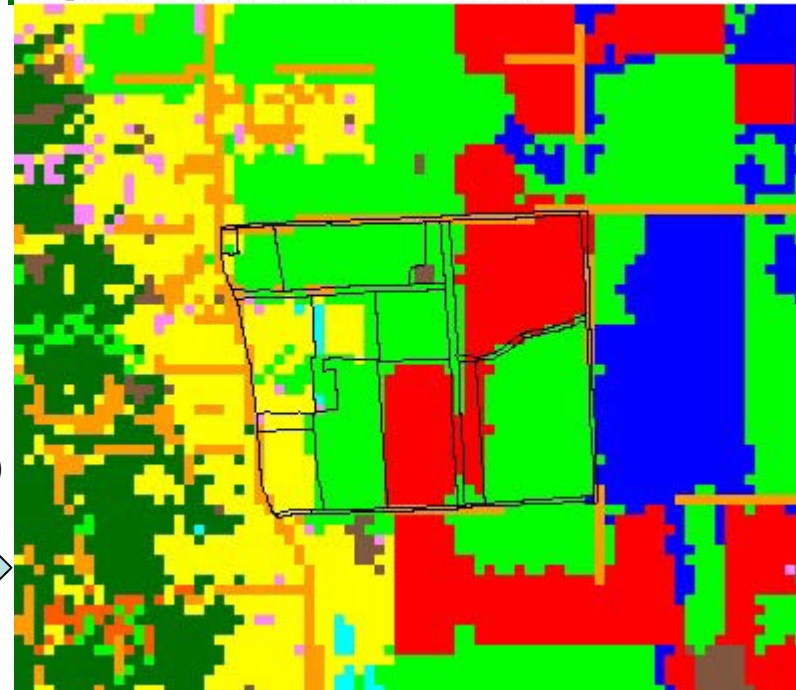
Segment 136 R=0.00 C=0.80 55555-00001



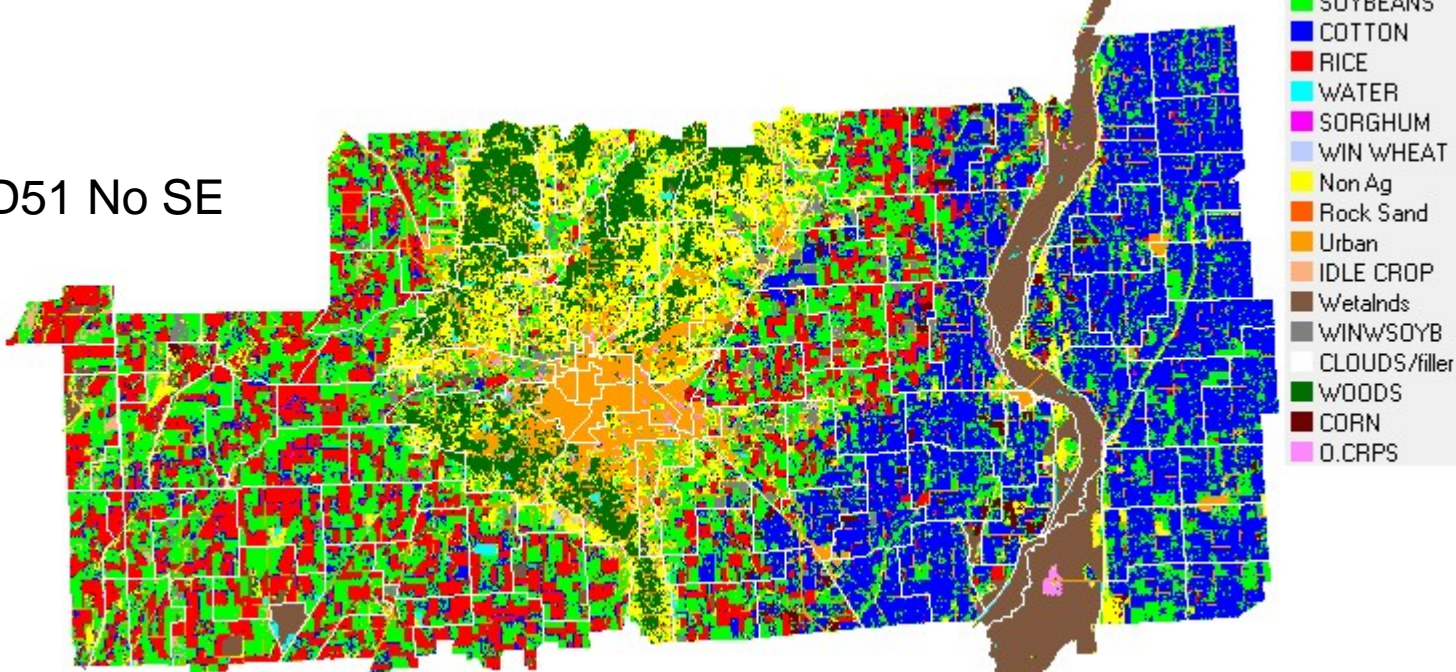
AD51 See5  
SE=0,NLCD



AD51 See5  
SE=20,NLCD

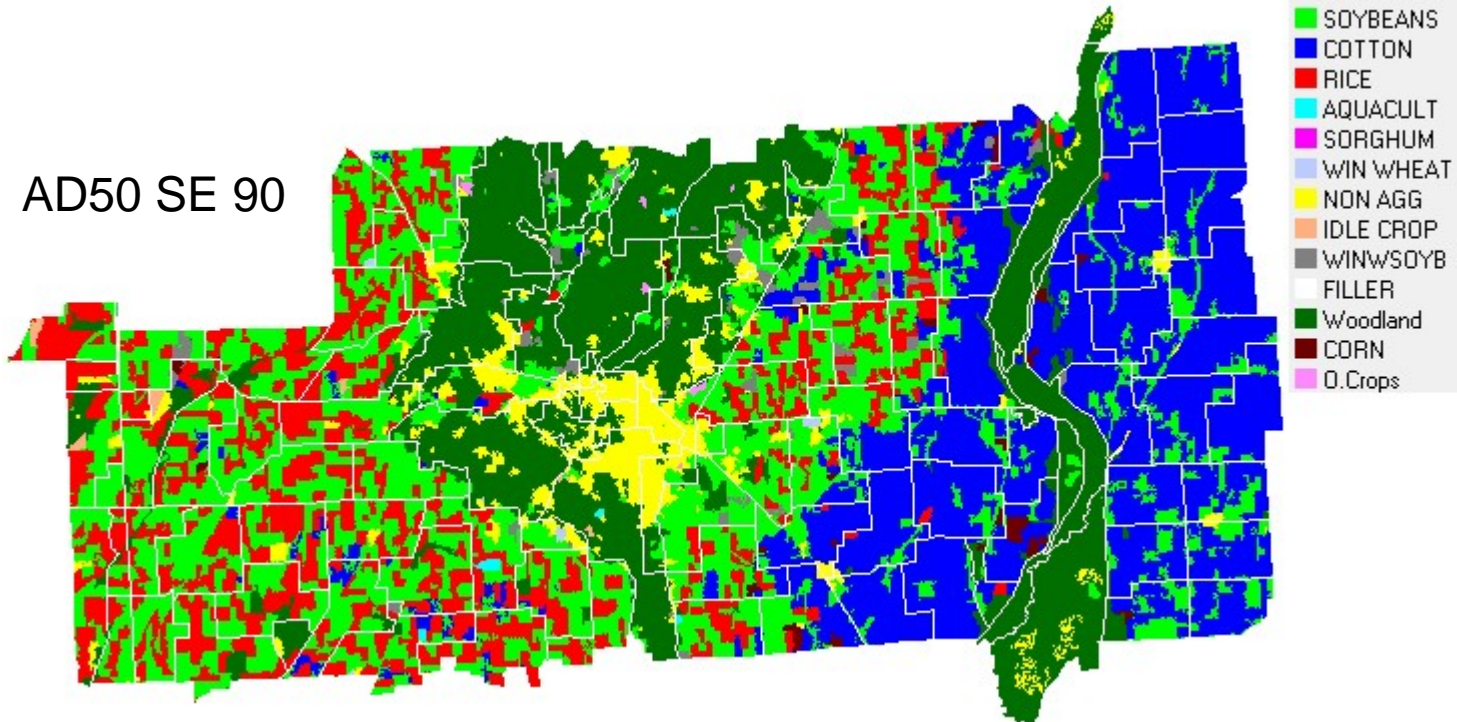


AD51 No SE



County CRAIGHEAD(31) AR Scene 74434-00005

AD50 SE 90



# Using See5's Smart Eliminate Kappa Values

## AD50 Subset

CSN	CORN	COTT	RICE	SOYB	OCRCP	NONAG	WOODS	OVERALL
3scn0800	80.14	87.44	93.14	86.21	45.93	62.56	88.60	83.41
3scn0820	77.57	92.23	94.20	94.06	44.96	59.80	94.20	86.81
3scn0890	58.67	90.97	87.10	93.71	15.32	53.75	91.88	82.52
3scn1000	78.26	86.76	94.12	86.30	46.36	63.16	87.68	83.18
3scn1020	77.75	91.61	93.77	93.80	42.40	59.15	93.43	86.33
3scn1090	67.36	89.74	89.15	93.72	14.97	52.55	90.45	82.50

## Statewide

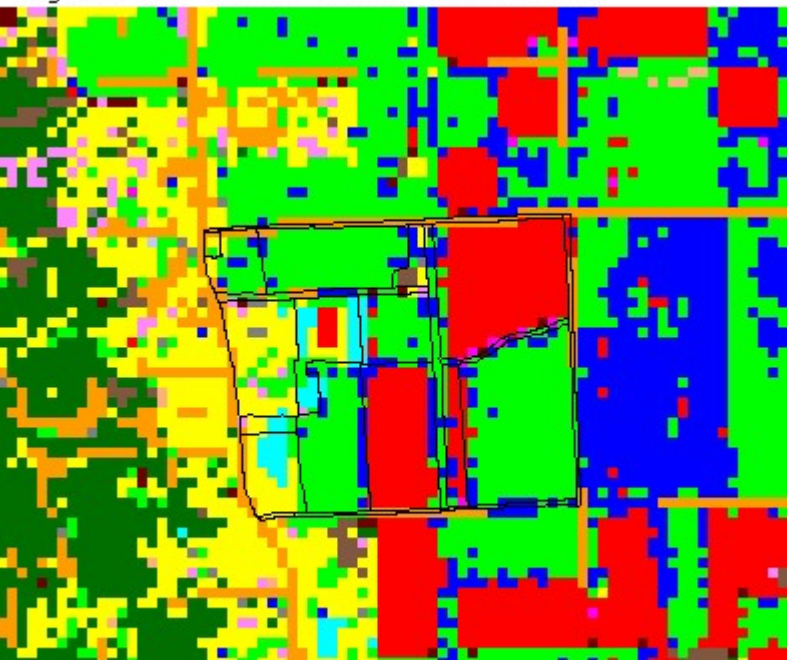
CSN	CORN	COTT	RICE	SOYB	OCRCP	NONAG	WOODS	OVERALL
Ascns5no	71.94	85.03	85.46	83.16	44.26	57.50	42.85	69.39
Ascns520	75.73	90.14	89.31	89.76	50.38	62.04	42.42	72.86
Ascns5se	71.41	90.62	87.45	91.21	39.60	54.10	42.41	71.70

# Using See5's Smart Eliminate Kappa Values

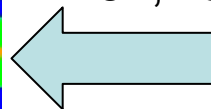
- AD50 Subset
  - (8 covers) x (2 bit types) = 16 comparisons
  - No SE wins 7, SE=20 wins 9, SE=90 wins 0
    - Using only “Overall” SE20 wins over both bit types
- Statewide
  - Only one measure per each of 8 covers
  - No SE wins 1, SE=20 wins 5, SE=45 wins 2
    - SE=20 wins the “Overall” measure



Segment 136 R=0.00 C=0.80 55555-00001



AD51, No SE

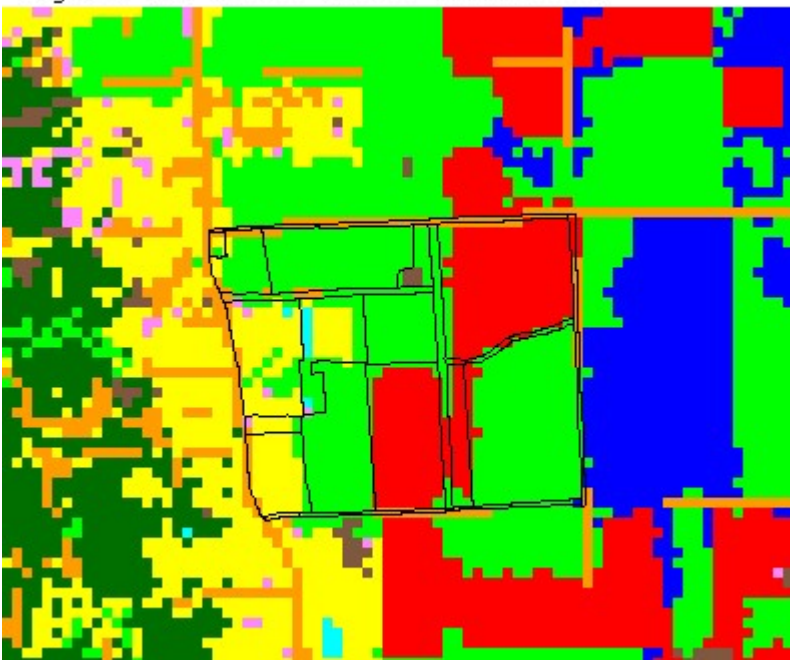


AD51, SE20



Segment 136 R=0.00 C=0.80 55555-00001

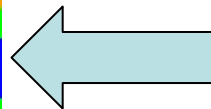
Segment 136 R=0.00 C=0.80 55555-00001



Segment 136 R=-0.20 C=0.80 74434-00005

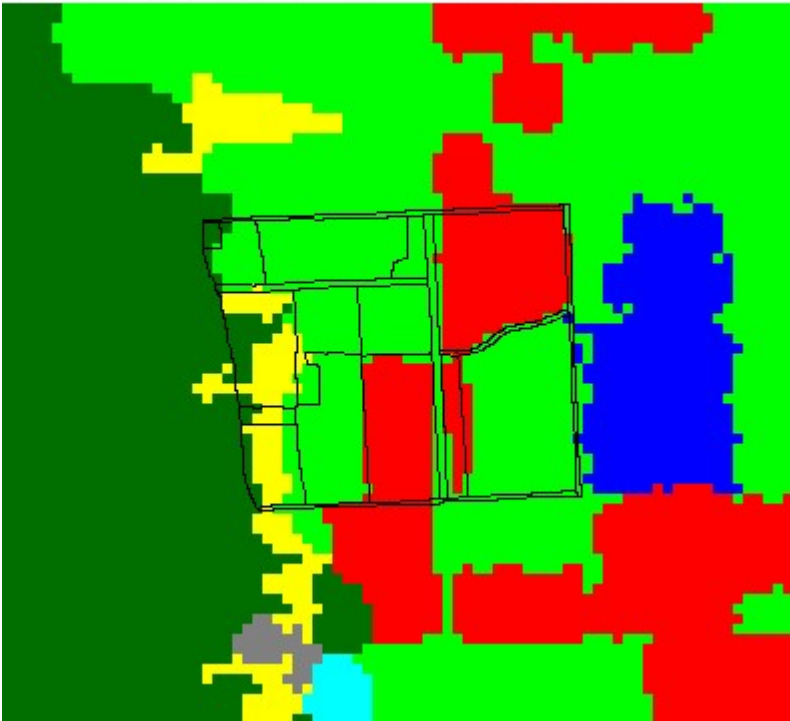


AD51, SE45



AD50, SE90

No NLCD

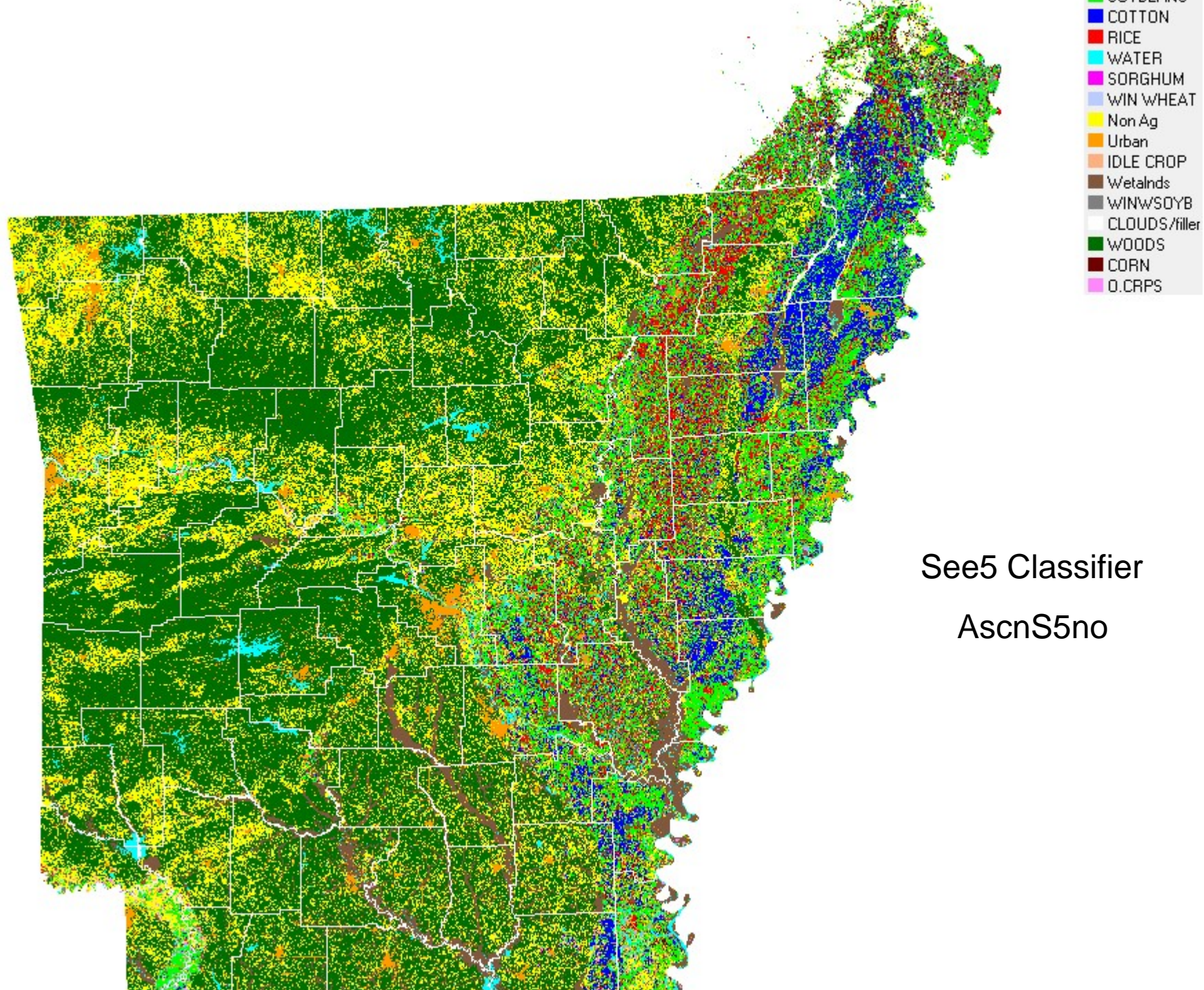


# Using See5's Smart Eliminate R<sup>2</sup> Values

Highest R <sup>2</sup> After Outlier Deletes						
	AD50 Analysis			Statewide Analysis		
	NO SE	SE 20	SE90	No SE	SE 20	SE 45
Corn	0.0	0.0	2.0	0.0	1.0	0.0
Cotton	0.0	1.0	1.0	0.0	0.0	1.0
Rice	1.0	1.0	0.0	0.0	1.0	0.0
Soybean	1.0	1.0	0.0	0.0	1.0	0.0
Sum	2.0	3.0	3.0	0.0	3.0	1.0

# Conclusions

- PEDITOR versus See5
  - See5 wins almost all comparisons easily
    - Need to 'beef up' See5 training: non-ag & woods
    - Added gain: use of FSA & any other data (not analyzed)
- 8-bit imagery versus 10-bit imagery
  - No clear winner, slight edge to 8-bit data
- Application of Smart Eliminate in See5
  - Definite edge to using SE at some level
    - "Best" level might be between 20 and 45
      - Similar research in another state chose 26 meters!
    - SE = 90 is too much
- Need to determine how & when to add NLCD



# Comments on Woods & NonAg

- Woods and NonAg are much more prevalent statewide per unit area than they are in the highly intensive AD50 area.
  - PEDITOR seems to do better for these cover types due to the efforts to get training for 'extra' signatures
- We need to come up with an approach to add similar 'extra' training to See5 in addition to just burning these covers in.