University of Illinois
Institute of Natural Resource Sustainability
William Shilts, Executive Director

ILLINOIS NATURAL HISTORY SURVEY
Brian D. Anderson Director
1816 South Oak Street
Champaign, IL 61820-6964
217-333-6830

Assessment and Potential of the 2007 USDA-NASS Cropland Data Layer for Statewide Annual Land Cover Applications

Don Luman and Tari Tweddale

Joint project between the
Illinois State Geological Survey and Illinois Natural History Survey

Prepared for: Tom Heavisides
Illinois Department of Natural Resources
Office of Realty and Environmental Planning

INHS Technical Report 2008 (49)
Date of issue: 03 December 2008

ILLINOIS NATURAL HISTORY SURVEY

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Institute of Natural Resources Sustainability

Assessment & Potential of the 2007 USDA-NASS Cropland Data Layer for Statewide Annual Land Cover Applications

INTRODUCTION AND BACKGROUND

During 1999-2002, ISGS and INHS scientists conducted and successfully completed the Illinois Land Cover project[1], a multiple agency cooperative initiative with principal funding provided by the Illinois Department of Natural Resources (IDNR) and the Illinois Department of Agriculture (IDA). This was an update and revision of the preceding Illinois Land Cover project[2], funded by the IDNR Critical Trends Assessment Project. The 1999-2002 land cover project incorporated the USDA National Agricultural Statistics Service (NASS), Cropland Data Layer (CDL) for Illinois. This first-time NASS CDL data product considerably improved the characterization and accuracy for the classification of agricultural lands in Illinois, and this information was subsequently combined with the ISGS/INHS land cover classification of non-agricultural lands. However, the integration procedures were labor intensive, requiring an additional year of processing.

In 2006, NASS developed a new protocol for developing an enhanced CDL product, one which characterizes both agricultural *and* non-agricultural lands, producing a comprehensive, statewide land cover data product. NASS implemented this new protocol for producing the Illinois 2007 CDL data product, as well as for the Illinois 2008 CDL[3]. The description of this enhanced CDL data product is as follows:

> "The USDA, NASS Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land cover data layer with a ground resolution of 56 meters. The CDL is produced using satellite imagery from the Indian Remote Sensing RESOURCESAT-1 (IRS-P6) Advanced Wide Field Sensor (AWiFS) collected during the current growing season. Ancillary classification inputs include: the United States Geological Survey (USGS) National Elevation Dataset (NED), the USGS National Land Cover Dataset 2001 (NLCD 2001), and the National Aeronautics and Space

[1] Luman, D., T. Tweddale, B. Bahnsen, and P. Willis, 2004, Illinois Land Cover, 1:500,000-scale color map, 32.5" x 52", ISGS Illinois Map 12.

[2] Luman, D., Joselyn, M. and Suloway, L., 1996, Illinois Land Cover, 1:500,000-scale color map, Illinois Scientific Surveys Joint Report 3.

[3] NASS completed the Illinois 2008 CDL data product in September 2008, but it won't be publicly released until March, 2009 (personal communication, Patrick Willis, USDA-NASS, November 2008).

> Administration (NASA) Moderate Resolution Imaging Spectroradiometer (MODIS) 250 meter 16 day Normalized Difference Vegetation Index (NDVI) composites. Agricultural training and validation data are derived from the Farm Service Agency (FSA) Common Land Unit (CLU) Program[4]. The NLCD 2001 is used as non-agricultural training and validation data. The strength and emphasis of the CDL is agricultural land cover. Please note that no farmer reported data are derivable from the Cropland Data Layer. The purpose of the Cropland Data Layer Program is to use satellite imagery to (1) provide acreage estimates to the Agricultural Statistics Board for the state's major commodities and (2) produce digital, crop-specific, categorized geo-referenced output products."[5]

From 1999-2006, the NASS CDL Program used both the Landsat 5 TM and Landsat 7 ETM+ satellite sensor systems (launched March 1984 and April 1999, respectively), providing eight day repeat coverage and multispectral imagery at a 30 meter ground resolution. Landsat 5 is still operational but is now twenty-five years old. In 2003, Landsat 7 experienced an irreparable mechanical problem with its sensor system, resulting in the loss of one-quarter of the reflectance data in every image scene. The current forecast planning is for the Landsat Data Continuity Mission (LDCM) satellite to be launched in July, 2011, which will collect imagery data consistent with the preceding Landsat satellites. The stability and longevity of the Landsat program warranted investigation into alternative sensor systems by NASS, and beginning in 2007 NASS has been exclusively using the ResourceSat-1 AWiFS sensor, providing five-day repeat visit coverage and multispectral imagery at a 56 meter ground resolution. The AWiFS sensor temporal and spectral features are complementary, and in some aspects are improvements over the Landsat 5 and 7 sensor systems. However, the AWiFS reduced spatial resolution means that the effective mapping scale has also been reduced from about 1:125,000 (1 inch=2 miles) for the Landsat based maps to about 1:250,000 (1 inch=4 miles) for the 2007-2008 CDL data products. Appendix A shows the same geographic area at the 1:125,000 scale for Landsat TM/ETM+ and ResourceSat-1 AWiFS source imagery (pp. 1-2), and p.3 displays the 2007 CDL data at the more appropriate 1:250,000 mapping scale.

As stated in the above metadata documentation, "…*The strength and emphasis of the CDL is agricultural land cover*…" All agricultural land cover is statistically assessed for accuracy by NASS as a part of their agency's mandate. As also described above, land cover data from the 2001 USGS NLCD is used as training data to derive updated, non-agricultural land

[4] Additional information for NASS CDL source data: Indian Remote Sensing RESOURCESAT-1 <http://www.isro.org/pslvc5/>; National Elevation Dataset <http://ned.usgs.gov/>; National Land Cover Dataset <http://www.mrlc.gov/>; Moderate Resolution Imaging Spectroradiometer < http://modis.gsfc.nasa.gov/>; Common Land Unit (CLU) Program <www.fsa.usda.gov/Internet/FSA_File/clu_2007_infosheet.doc>.

[5] From USDA, National Agricultural Statistics Service, 2007 Illinois Cropland Data Layer Metadata documentation, <http://www.nass.usda.gov/research/Cropland/metadata/meta.htm>.

cover for the new, enhanced CDL data product. However, NASS conducts no statistical analysis of these data, and the relative accuracy of non-agricultural land cover in the Illinois 2007 CDL is not known. In order to assess the potential usefulness of NASS' annual land cover data product to Illinois agencies, ISGS and INHS scientists conducted a statistical evaluation of the non-agricultural land cover data in the 2007 CDL data product. ISGS and INHS worked in close collaboration with NASS to ensure consistency was maintained in the accuracy assessment procedures applied to both agricultural and non-agricultural land cover data.

2007 CDL LAND COVER CATEGORIES

The land cover categories contained within the Illinois 2007 CDL data product are listed in Table 1. Of the 42 individual categories included, just eleven cover types comprise one percent or more of the surface area and account for nearly all of the state's land cover (97.44 percent or 35,139,311 acres). Corn and soybean row crops alone account for slightly more than one-half of Illinois' total surface area (51.54 percent or 18,589,117 acres).

Category #	CDL Category	Acres	Percent
	Row Crops (1-20)		
1	Corn	11,703,821	32.45%
3	Rice	532	0.00%
4	Sorghum	26,352	0.07%
5	Soybeans	6,885,296	19.09%
6	Sunflowers	358	0.00%
11	Tobacco	26	0.00%
	Grains, Hay, Seeds (21-40)		
21	Barley	136	0.00%
23	Spring Wheat	142	0.00%
24	Winter Wheat	216,283	0.60%
25	Other Small Grains	3	0.00%
26	Winter Wheat/Soybeans (Double Cropped)	464,960	1.29%
27	Rye	355	0.00%
28	Oats	6,257	0.02%
29	Millet	59	0.00%
36	Alfalfa	51,051	0.14%
	Other Crops (41-60)		
42	Dry Beans	5,851	0.02%
43	Potatoes	3,163	0.01%
44	Other Crops	6,080	0.02%
47	Miscellaneous Vegetables and Fruits	37,935	0.11%
53	Peas	2,173	0.01%
58	Clover/Wildflowers	2,598	0.01%
	Open Non-Crop (61-65)		
61	Fallow/Idle Cropland	10,181	0.03%
62	Grassland/Pasture/Non-Ag	410,945	1.14%
63	Woodland	9,949	0.03%
	Tree Crops (66-80)		
67	Peaches	9	0.00%
68	Apples	9	0.00%
70	Christmas Trees	5	0.00%
	Other Non-Crop (81-99)		

87	Wetlands	43,836	0.12%
	NLCD Derived, Non-Crop (100-195)		
111	Open Water	621,874	1.72%
121	Developed/Open Space	2,426,046	6.73%
122	Developed/Low Intensity	1,695,232	4.70%
123	Developed/Medium Intensity	479,639	1.33%
124	Developed/High Intensity	178,577	0.50%
131	Barren	28,019	0.08%
141	Deciduous Forest	6,148,020	17.05%
142	Evergreen Forest	30,408	0.08%
143	Mixed Forest	85	0.00%
152	Shrub/Scrub	1,735	0.00%
171	Grassland Herbaceous	203,034	0.56%
181	Pasture/Hay	3,936,737	10.92%
190	Woody Wetlands	366,742	1.02%
195	Herbaceous Wetlands	61,450	0.17%
	STATE TOTAL	36,065,963	100.00%

Table 1. Illinois 2007 CDL individual land cover categories. Categories comprising at least one percent of the state's surface area are underlined.

ACCURACY ASSESSMENT ANALYSIS

NASS uses training data collected from agricultural and other sources to derive the land cover for CDL categories #1-87 (Table 1). These 28 cover categories are subsequently accuracy tested using independent validation samples generated from ground truth data derived from the USDA FSA Common Land Unit (CLU) Program. CDL categories #111-195 were derived using training data collected directly from the 2001 USGS NLCD and no independent sampling was conducted for the NASS-based accuracy assessment of these non-agricultural cover categories.

Explanation of Accuracy Measures

When a land cover classification is conducted using remote sensing imagery as the primary source data, it is important to provide an overall evaluation of the performance of the final land cover map product. Much research has been conducted on classification accuracy assessment techniques as they apply to map products derived from remote sensing data. While no universally accepted measures have been adopted, several standardized statistics are widely used and accepted which include the following: Producer's Accuracy, Omission Error, User's Accuracy, Commission Error, and Kappa. A brief summary of these statistics is presented below to better understand the results of the accuracy assessment analyses conducted by NASS and ISGS/INHS. [6]

Producer's Accuracy is calculated for each individual cover type and indicates the probability that an independently derived, ground reference sample will be correctly mapped. It

[6] By authors, and also adapted from USDA, National Agricultural Statistics Service, 2007 Illinois Cropland Data Layer Metadata documentation, <http://www.nass.usda.gov/research/Cropland/metadata/meta.htm>.

is so-called because the producers, or originators of the land cover classification, are principally interested in how accurately the ground reference samples can be classified using the primary remote sensing source imagery. Likewise, an *Omission Error* occurs when a reference sample is excluded from the land cover map category to which it belongs in the validation dataset. Otherwise stated, omission errors represent those samples in the land cover map pertaining to an actual class on the landscape which the computer classification has failed to recognize. Producer's accuracy (PA) and omission error (OE) are directly associated with one another in a simple, inverse manner as follows: PA%=100-OE% or OE%=100-PA%.

User's Accuracy indicates the probability that a sample pixel from the CDL land cover classification actually matches the ground reference data. It is an indication of the probability, or reliability that an independent sample drawn from the classification map actually represents that land cover category on the actual landscape. *Commission Error* occurs when a reference sample is included in an incorrect category according to the ground reference validation data. Stated another way, commission error refers to those samples from other landscape classes which the computer classification has incorrectly assigned as belonging to the particular landscape class of interest. Unfortunately for users of land cover information, authors have also directly associated commission error with user's accuracy, and while the two are traditionally shown together, these statistical measures are not similarly related in the same simple manner as producer's accuracy and omission error.

As a simple example, if the land cover classification assigns every image pixel to “Corn”, the producer's accuracy for the Corn category would be 100 percent with a 0 percent omission error. Conversely, a very high error of commission results, because all other crop types would be included in the incorrect category. These four statistics are calculated for all of the individual cover types and typically shown in the format of a table, or contingency matrix (discussed later).

Producer's accuracy, user's accuracy, omission error, and commission error each estimate error and assess accuracy utilizing only a portion of the entire contingency matrix, and the differing interpretations which result can be a source of unnecessary confusion to the user of land cover information. The *Kappa* statistic is attractive in that it effectively summarizes the entire contingency matrix to a single statistic. It is a measure of agreement based on the difference between the actual agreement and chance agreement involving the remotely sensed-based land cover classification map and the ground reference data. Stated in another manner, Kappa is a quantitative measure of the difference between the observed agreement between two images/maps and the agreement that may be contributed solely by the chance matching of two images/maps. Kappa effectively adjusts the overall percentage correct by subtracting the estimated contribution of chance agreement, which is to infer that the agreement between two images/maps cannot be attributed exclusively to the "success" of the computer classification. Kappa is stated as a proportion from 0.0-1.0, with 1.0 indicating perfect agreement. Lastly, the *Conditional Kappa* statistic is the agreement for an individual category within the entire contingency matrix, and is used to assess the accuracy of individual land cover categories in comparison to the ground reference data.

Crop Specific Accuracy Assessment

NASS conducted a detailed accuracy assessment of the crop-specific, agricultural categories contained within the Illinois 2007 CDL. Table 2 (bottom line) shows that the overall producer's accuracy is 97.6 percent, with a corresponding omission error of only 2.4 percent. The overall Kappa of 0.95 is very strong evidence that the success of the classification did not occur by chance. The individual producer's accuracy, user's accuracy, omission error, commission error, and Kappa statistics for the most predominant crop types – corn, soybeans, winter wheat, and winter wheat/soybeans (double cropping) are likewise high and is ample assurance that the crop-specific cover categories contained in NASS' CDL data product are well characterized.

CDL Crop-Specific Cover	Attribute Code	Correct Pixels	Producer's Accuracy	Omission Error	Kappa	User's Accuracy	Commission Error	Cond'l Kappa
Corn	1	2,772,986	98.7%	1.3%	0.97	97.6%	2.4%	0.95
Rice	3	77	88.5%	11.5%	0.89	40.3%	59.7%	0.40
Sorghum	4	3,934	78.8%	21.2%	0.79	40.3%	59.7%	0.40
Soybeans	5	1,322,321	96.9%	3.1%	0.96	96.7%	3.3%	0.95
Sunflowers	6	3	27.3%	72.7%	0.27	1.4%	98.6%	0.01
Barley	21	17	60.7%	39.3%	0.61	13.9%	86.1%	0.14
Spring Wheat	23	33	97.1%	2.9%	0.97	36.7%	63.3%	0.37
Winter Wheat	24	25,768	81.0%	19.0%	0.81	71.9%	28.1%	0.72
Win Wht/Soyb Dbl Crop	26	77,663	87.6%	12.4%	0.87	85.6%	14.4%	0.85
Rye	27	37	74.0%	26.0%	0.74	14.5%	85.5%	0.15
Oats	28	238	46.6%	53.4%	0.47	26.0%	74.0%	0.26
Alfalfa	36	3,998	64.4%	35.6%	0.64	28.7%	71.4%	0.29
DryBeans	42	634	87.0%	13.0%	0.87	58.6%	41.4%	0.59
Potatoes	43	145	65.9%	34.1%	0.66	38.5%	61.5%	0.38
Other Crops	44	209	52.5%	47.5%	0.53	29.0%	71.0%	0.29
Misc Veg & Fruits	47	6,076	89.4%	10.6%	0.89	61.0%	39.0%	0.61
Peas	53	67	69.1%	30.9%	0.69	46.5%	53.5%	0.47
Clover/Wildflowers	58	175	60.3%	39.7%	0.60	11.2%	88.8%	0.11
Fallow/Idle Cropland	61	225	41.4%	58.6%	0.41	8.0%	92.0%	0.08

	Correct Pixels	Producer's Accuracy	Omission Error	Kappa
OVERALL ACCURACY	4,214,606	97.6%	2.4%	0.95

Table 2. Statewide Agricultural Accuracy Report, USDA National Agricultural Statistics Service, 2007 Illinois Cropland Data Layer. The most predominant crops are underlined. Certain crop-specific categories shown in Table 1 are not included because they did not possess sufficient ground area to be statistically evaluated.

Accuracy Assessment of Non-Agricultural Categories

A primary focus of this project is to statistically assess the non-agricultural land cover categories contained in the Illinois 2007 CDL data product. The majority of these land cover data were originally derived from the Illinois portion of the 2001 USGS National Land Cover Dataset, sample points were then collected by NASS from each NLCD-derived cover category to be used as training data, and a supervised classification of the 2007 AWiFS imagery data was subsequently performed to generate updated, non-agricultural land cover (Table 1, CDL

categories #111-195). A statistical assessment of the resulting non-agricultural land cover was performed, but NASS clearly states, "…the *accuracy of the CDL non-agricultural land cover classes is entirely dependent upon the USGS, National Land Cover Dataset*..."[7] As previously mentioned, no independent ground reference validation data were collected by NASS to assess the performance of the NLCD-derived non-agricultural cover. Instead, sample pixels were collected directly from the original, source USGS NLCD land cover to calculate the accuracy measures, an approach which introduces considerable bias and renders the results suspect.

An independent and unbiased accuracy assessment of the non-agricultural categories is necessary in order for NASS CDL-based land cover information to be used in natural resource-based applications. Because of the difficulty in discriminating among certain cover types when using remote sensing imagery as the primary reference validation data, we decided to merge selected non-agricultural/non-crop specific CDL categories as follows:

- Categories #181-Pasture/Hay and #152-Shrub/Scrub are normally indistinguishable from #171-Grassland Herbaceous cover and were therefore combined. In addition, Shrub/Scrub cover has too little areal extent to be assessed as a separate cover category (Table 1)

- Categories #63-Woodland and #143-Mixed Forest were combined with #141-Deciduous Forest to create a single deciduous forest/woodland category. As with Shrub/Scrub, Mixed Forest cover occurs too infrequently to be evaluated separately.

- Lastly, three NASS-derived "Non-Crop" categories (Table 1) – #62-Grassland/Pasture/Non-Ag, #63-Woodland, and #87-Wetlands were individually combined with the Grassland Herbaceous, Deciduous Forest, and Emergent Herbaceous Wetlands categories, respectively.

The final arrangement and composition of the non-agricultural cover types is listed below in Table 3. The definitions of each of these eleven cover categories are provided in Appendix B. Based upon the experiences of several prior land cover projects, we developed a detailed protocol to perform the accuracy assessment analysis, and the step-by-step procedures are outlined in Appendix C.

[7] USDA, National Agricultural Statistics Service, 2007 Illinois Cropland Data Layer Metadata documentation, <http://www.nass.usda.gov/research/Cropland/metadata/meta.htm>.

CDL Category #	Non-Agricultural Category
111	Open Water
121	Developed/Open Space (includes rural roads)
122	Developed/Low Intensity
123	Developed/Medium Intensity
124	Developed/High Intensity
131	Barren
141	Deciduous Forest (also incorporates #143-Mixed Forest and #63-Woodland)
142	Evergreen Forest
171	Grassland Herbaceous (also incorporates #152-Shrub/Scrub, #181-Pasture/Hay, and #62-Grassland/Pasture/Non-Ag)
190	Woody Wetlands
195	Emergent Herbaceous Wetlands (also incorporates #87-Wetlands)

Table 3. Illinois 2007 NASS-CDL derived non-agricultural categories used for the accuracy assessment analysis.

Study Area Selection and Ground Reference Sources

Because NASS had conducted a thorough assessment of the 2007 CDL within the areas of the state dominated by agricultural land cover, we decided to limit our analysis to include only those counties which have 65 percent or more of their total surface area devoted to non-agricultural land cover types. This resulted in a sample set of 40 Illinois counties, from which 824 ground reference samples were selected using a random sampling approach, stratified by land cover category (Figure 1).

Since field verification of 824 ground reference samples distributed over a 40-county area would be cost-prohibitive, image interpretation of large-scale aerial photography in conjunction with other ancillary geospatial data sets was used to assign each ground reference sample to one of eleven non-agricultural cover categories. Temporal, spatial, spectral, and contextual factors are critical in the selection of appropriate reference imagery which will serve as the primary "ground truth" for an accuracy assessment analysis. The following examples on the following page demonstrate these factors:

- Temporal – The primary AWiFS satellite imagery data used for the CDL land cover classification were collected during the period of early April through early September, 2007. The reference aerial photography used for accuracy assessment should ideally be collected for a similar time period in order that ground features correlate with the satellite imagery. Therefore, Illinois 2007 USDA National Agriculture Imagery Program (NAIP) digital orthophotography, collected during the period from late June through early September, 2007 was acquired from the USDA Aerial Photography Field Office (APFO) and served as the primary source imagery for the interpretation.

  NAIP is summer "leaf-on", natural color format imagery. The interpretation of certain cover categories such as forested land require both leaf-on and leaf-off imagery in order to discriminate forest/woodland species types. Leaf-off USGS 2005 NAPP digital orthophotography was used in conjunction with the 2007 leaf-on NAIP to provide spring/summer multitemporal information (see Figure 2 for example).

- Spatial – In an accuracy assessment of land cover map products produced from satellite imagery, it is recommended that larger scale (more detail) aerial imagery be used in the identification of the ground reference samples. Figure 3 shows how the high spatial resolution of the 2007 NAIP imagery facilitates the accurate identification of this ground reference sample as a gas storage facility and assignment to the Developed/High Intensity cover category.

- Spectral – The 2007 Illinois NAIP was acquired in natural color format. In contrast, as a cost-share partner for the 2004 NAIP, IDNR requested the imagery be collected in color infrared format, providing enhanced interpretation of most surface features. Figure 4 shows the distinctive bright magenta (pink) color of emergent vegetation surrounding a shallow water wetland habitat.

- Contextual – Temporal, spatial, and spectral imagery factors may not provide the final and accurate interpretation of a ground reference sample area. Therefore, ancillary resources including digital elevation model data and 100-year flood zones were used as contextual information (Figure 5).

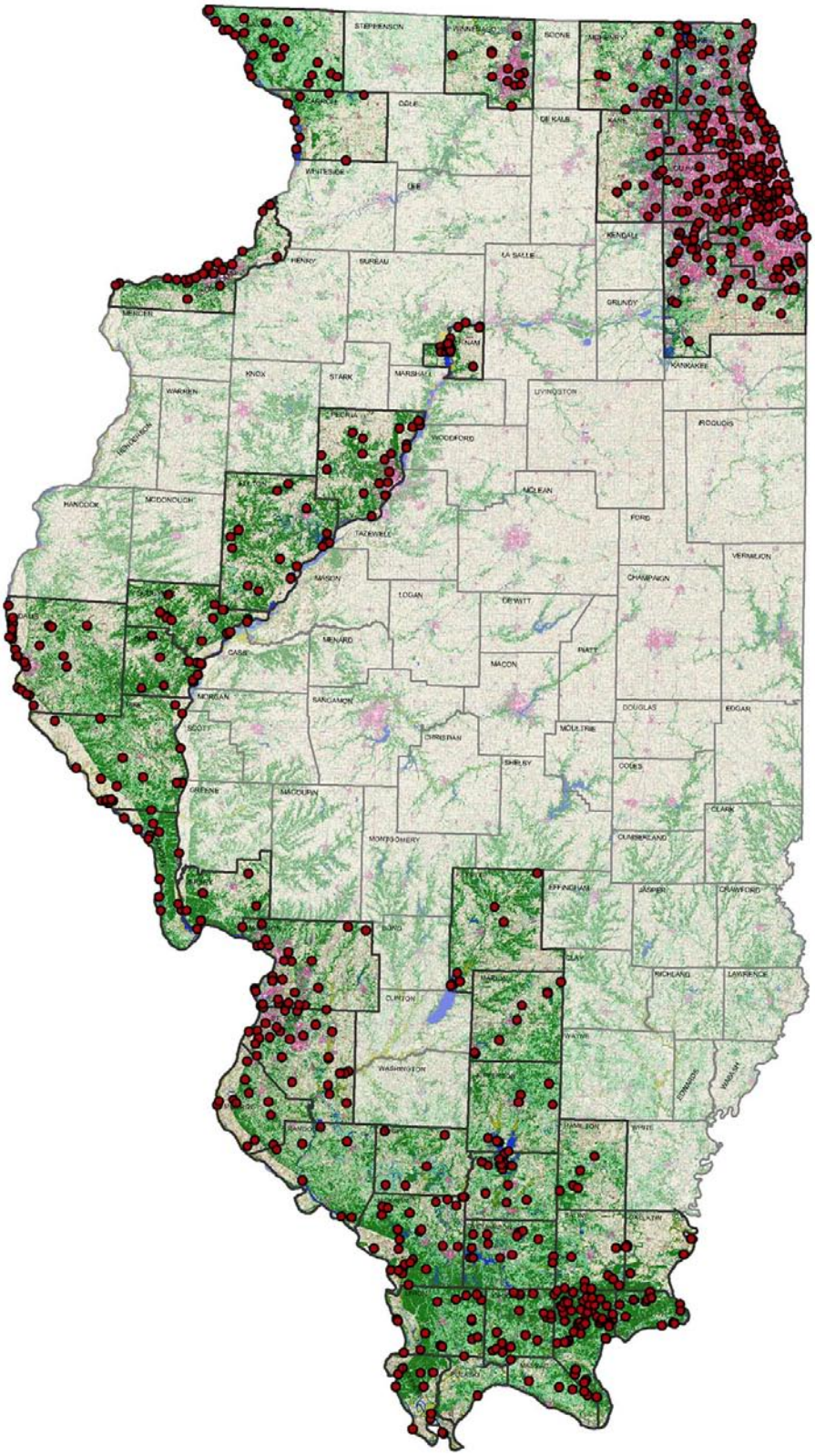


Figure 1. Illinois 2007 CDL land cover map. Highlighted counties contain 65 percent or more of their surface area as non-agricultural land cover. The geographic locations of the ground reference samples used for the non-agricultural accuracy assessment analysis are shown as red symbols.

Figure 2.a (top)-2.b (below). Discrimination of deciduous and evergreen forest cover using multitemporal imagery. The darker toned area of evergreens in Figure 2.b contrasts markedly with the surrounding deciduous trees in this early spring, 2005 USGS NAPP digital orthophoto. Other than a textural change, this species boundary is almost indistinguishable in the summer, natural color USDA NAIP digital orthophoto (Figure 2.a). Also shown are three ground reference point locations, with the 9x9 pixel sample area window used by the image interpreter to assign each sample area to a specific cover category.

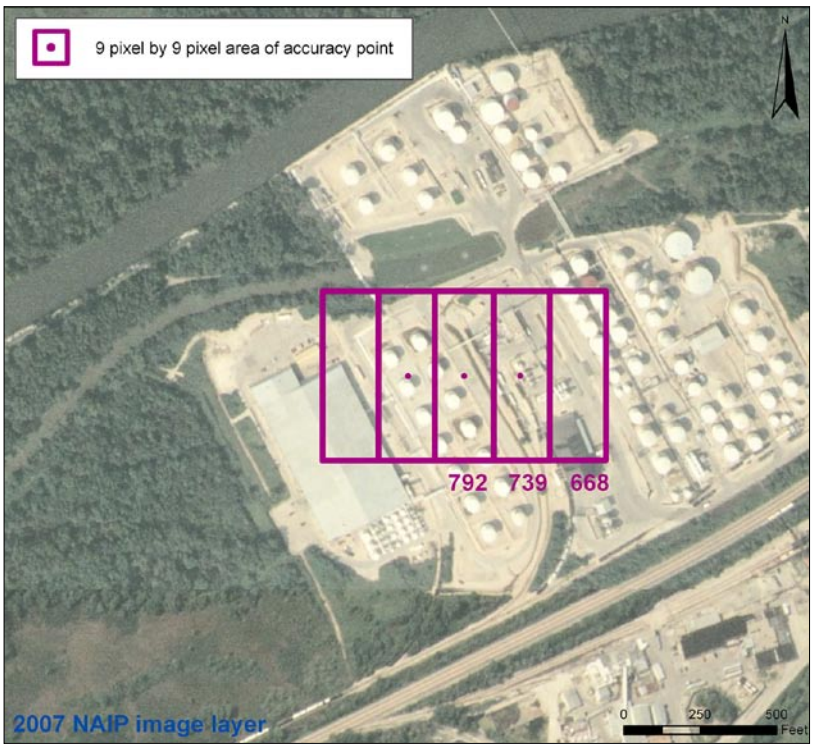


Figure 3. 2007 NAIP natural color image showing a large gas storage facility.

Figure 4. 2004 NAIP color infrared image of a wetland (center) adjacent to the Illinois River. The distinctive bright magenta (pink) color is emergent vegetation surrounding a shallow water environment. Deeper and more turbid surface water is shown at the upper right of the photo.

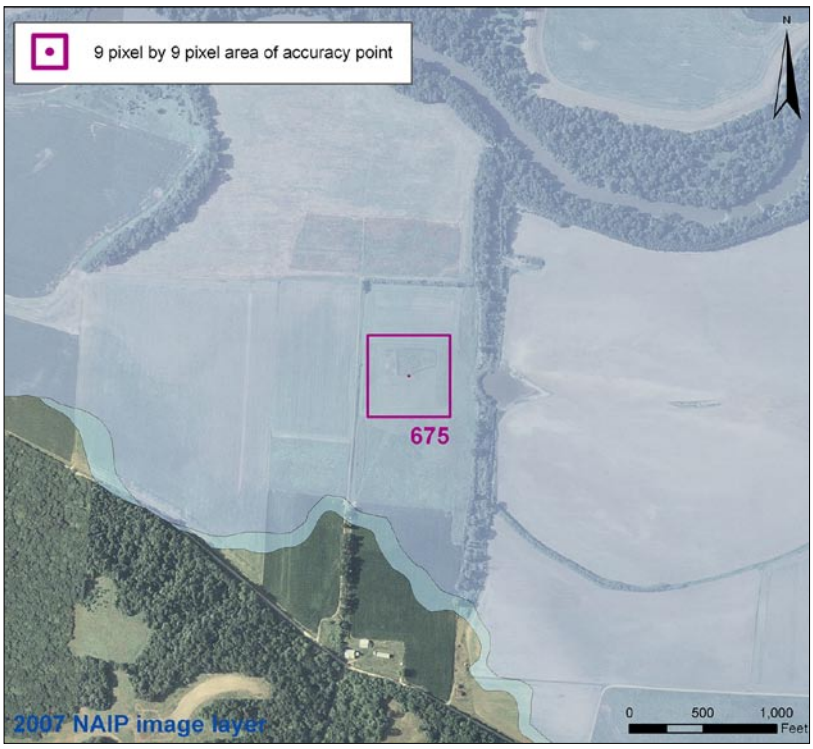


Figure 5. 2007 NAIP image of an agricultural area within the 100-year floodplain.

Accuracy Assessment Results

The results of the statistical analysis showed that the calculated overall accuracy of the 2007 Illinois CDL non-agricultural land cover categories is 88.6 percent with an overall Kappa statistic of 0.87. In the early 1970s at the USGS, James Anderson developed a land use and land cover classification system employing remote sensor data as the primary source information.[8] In the interim, very significant developments have occurred with aerial and satellite sensor systems – yet more than three decades later, the Anderson system remains well-established as a benchmark for land use and land cover analysis. One principle of the Anderson system is that the minimum level of interpretation accuracy in the identification and assignment of land use and land cover categories should be at least 85 percent, which is a frequently quoted *ad hoc* standard for studies involving land use and land cover classification. Therefore the overall accuracy level attained in this study easily exceeds this minimum criterion. Additionally, the calculated overall Kappa is very strong evidence that the agreement between the final classification map and the ground reference data did not occur by chance.

[8] JAMES R. ANDERSON, ERNEST E. HARDY, JOHN T. ROACH, and RICHARD E. WITMER, A Land Use And Land Cover Classification System For Use With Remote Sensor Data, Geological Survey Professional Paper 964 (a revision of the land use classification system as presented in U.S. Geological Survey Circular 671), United States Government Printing Office, Washington: 1976, 41p.

Explanation of the Contingency Matrix

Appendix D contains the detailed accuracy assessment contingency matrix for the non-agricultural land cover categories. Interpretation of the contingency matrix can become involved, but a brief summary is useful for demonstrating its information content. The individual categories form the primary column and row cells. Situated along the main diagonal (underlined and in bold type) show the number of correctly classified samples interpreted from the reference imagery data used for the accuracy assessment. The sum of the diagonal cell entries gives the total number of correctly classified samples (730); dividing this sum by the total number of samples (824) and multiplying by 100 determines the overall percentage correct statistic (88.6 percent).

The interpretation of the contingency matrix differs depending upon whether the reader examines the table on a column-by-column or row-by-row basis. Reading down each column of the primary contingency table, the cell entries show how the ground reference samples for an individual land cover category were assigned from the standpoint of the originators, or *producers* of the classification (USDA NASS). For example, 65 of the 73 "Open Water" reference samples were correctly assigned, resulting in a producer's accuracy of 89.0 percent. Conversely, reading across each row, the cell entries show how the ground reference samples for an individual land cover category were assigned from the perspective of the *user* of the classification. Using the same example as above, the number of correct samples remains the same (65), but the total number of reference samples that were classified as "Open Water" on the CDL land cover map is 72, resulting in a user's accuracy of 90.3 percent.

Examination of the off-diagonal cells shows the distribution of omission and commission errors, providing direct insight into the cover types assigned to misclassified samples. Note the pattern of misclassified samples associated with the Developed/Open Space category. From the producer's standpoint, 10 of the 78 reference samples were incorrectly assigned to Herbaceous Grassland and three to the Developed/Low Intensity category, not surprising given the similarities in surface cover and category definitions (Appendix B). From the standpoint of a user of the land cover map, 11 reference samples actually belonging to the Developed/Low Intensity category were incorrectly assigned to six other cover types, a mixed pattern that is more puzzling to explain. It is no wonder that the contingency matrix is also commonly referred to as a "confusion matrix".

Non-Agricultural Land Cover Summary Table

The detailed contingency matrix is often presented in the form of a summary table to simplify discussion of the accuracy assessment analysis. Table 4 shows that the user's accuracy statistic for the individual non-agricultural cover categories ranges from a minimum of 77.4 percent for Herbaceous Grassland to 100 percent for Evergreen Forest. The corresponding

Non-Agricultural Land Cover Category	Category #	Producer's Accuracy	Omission Error	Cond'l Kappa	User's Accuracy	Commission Error	Cond'l Kappa
Open Water	111	89.0%	11.0%	0.88	90.3%	9.7%	0.89
Developed/Open Space	121	83.3%	16.7%	0.82	85.5%	14.5%	0.84
Developed/Low Intensity	122	93.5%	6.5%	0.93	94.7%	5.3%	0.94
Developed/Medium Intensity	123	90.9%	9.1%	0.90	97.2%	2.8%	0.97
Developed/High Intensity	124	83.5%	16.5%	0.82	93.0%	7.0%	0.92
Barren	131	86.2%	13.8%	0.85	80.0%	20.0%	0.78
Deciduous Forest	141 (63,143)	97.5%	2.5%	0.97	84.8%	15.2%	0.83
Evergreen Forest	142	98.6%	1.4%	0.98	100.0%	0.0%	1.00
Grassland Herbaceous	171 (62,152,181)	86.7%	13.3%	0.85	77.4%	22.6%	0.75
Woody Wetlands	190	82.5%	17.5%	0.81	93.0%	7.0%	0.92
Herbaceous Wetlands	195 (87)	82.6%	17.4%	0.81	81.4%	18.6%	0.80

Table 4. Summary accuracy assessment report of the 2007 Illinois CDL non-agricultural land cover categories.

Kappa statistic for each individual cover category (conditional Kappa) ranges from 0.75 to 1.0. Conversely, the producer's accuracy statistic ranges from 82.5 percent (Woody Wetlands) to 98.6 percent (Evergreen Forest), with the associated conditional Kappa statistic values of .81 and .98, respectively. Another tenet of the Anderson classification system is, "…*The accuracy of interpretation for the several categories should be about equal...*" [9] A cursory inspection of the producer's and user's accuracy levels in Table 4 shows this principle was maintained for the non-agricultural land cover categories.

Combined Agricultural and Non-Agricultural Accuracy Summary Table

The final step in the accuracy assessment was to combine the ground reference sample data from the ISGS-INHS non-agricultural analysis with the NASS-derived assessment data for the remaining CDL land cover categories. The resulting contingency matrix, containing the detailed analysis of 32 individual land cover categories, is included in this report as Appendix E. The summary statistics for each land cover category are presented in Table 5.a. Note that for cover categories having small areal extent within the state, *e.g.* Tobacco, Other Grains, Oats (see Table 1), the accuracy values will be much lower, or not reported at all (too few reference samples) and therefore correspondingly high omission and commission errors occur.

As mentioned previously, two crop-specific categories, corn and soybeans, account for over one-half of the total surface area of the state. The producer's accuracy and user's accuracy values are above 95 percent for these two dominant categories, and the Kappa statistic values are .95 or higher. Land cover accounting for at least one percent of the state's surface area are underlined in Table 5.a. Combining the statistical data across for all 32 land cover categories, the statewide overall accuracy is 97.6 percent with an overall Kappa statistic of 0.95 (Table 5.b).

[9] JAMES R. ANDERSON, p.9

Individual Land Cover Category	Value	Producer's Accuracy	Omission Error	Cond'l Kappa	User's Accuracy	Commission Error	Cond'l Kappa
Corn	1	98.7%	1.3%	0.96	99.2%	0.8%	0.98
Rice	3	88.5%	11.5%	0.89	41.4%	58.6%	0.41
Sorghum	4	78.8%	21.2%	0.79	46.8%	53.2%	0.47
Soybeans	5	96.9%	3.1%	0.95	98.7%	1.3%	0.98
Sunflowers	6	27.3%	72.7%	0.27	2.6%	97.4%	0.03
Tobacco	11	0.0%	100.0%	0.00	0.0%	100.0%	0.00
Barley	21	60.7%	39.3%	0.61	21.3%	78.8%	0.21
Spring Wheat	23	97.1%	2.9%	0.97	50.8%	49.2%	0.51
Winter Wheat	24	81.0%	19.0%	0.81	79.6%	20.4%	0.79
Other Grains	25	0.0%	100.0%	0.00	0.0%	100.0%	0.00
Winter Wheat / Soybeans	26	87.6%	12.4%	0.87	89.8%	10.2%	0.90
Rye	27	74.0%	26.0%	0.74	30.3%	69.7%	0.30
Oats	28	46.6%	53.4%	0.47	52.4%	47.6%	0.52
Alfalfa	36	64.4%	35.6%	0.64	84.6%	15.4%	0.85
Dry Beans	42	87.0%	13.0%	0.87	62.2%	37.8%	0.62
Potatoes	43	65.9%	34.1%	0.66	50.3%	49.7%	0.50
Other Crops	44	52.5%	47.5%	0.53	62.2%	37.8%	0.62
Misc. Vegetables	47	89.4%	10.6%	0.89	64.5%	35.5%	0.64
Peas	53	69.1%	30.9%	0.69	50.4%	49.6%	0.50
Clover / Wildflowers	58	60.3%	39.7%	0.60	62.5%	37.5%	0.62
Idle / Fallow	61	41.4%	58.6%	0.41	29.8%	70.2%	0.30
Open Water	111	89.0%	11.0%	0.89	20.8%	79.2%	0.21
Developed/Open Space	121	83.3%	16.7%	0.83	1.7%	98.3%	0.02
Developed/Low Intensity	122	93.5%	6.5%	0.94	24.9%	75.1%	0.25
Developed/Medium Intensity	123	90.9%	9.1%	0.91	81.4%	18.6%	0.81
Developed/High Intensity	124	83.5%	16.5%	0.84	93.0%	7.0%	0.93
Barren	131	86.2%	13.8%	0.86	48.7%	51.3%	0.49
Deciduous Forest	141 (63,143)	97.5%	2.5%	0.97	6.3%	93.7%	0.06
Evergreen Forest	142	98.6%	1.4%	0.99	100.0%	0.0%	1.00
Grassland Herbaceous	171 (62,152,181)	86.7%	13.3%	0.87	0.2%	99.8%	0.00
Woody Wetlands	190	82.5%	17.5%	0.82	61.1%	38.9%	0.61
Herbaceous Wetlands	195 (87)	82.6%	17.4%	0.83	4.1%	95.9%	0.04

Table 5.a. Combined summary accuracy assessment report of the 2007 Illinois CDL agricultural and non-agricultural land cover categories. Land cover categories accounting for one percent or more of the state's surface area are underlined.

Table 5.b. Illinois 2007 NASS CDL Overall Accuracy Assessment	
Number of Correct Samples	4,215,336
Number of Total Samples	4,317,189
Overall Percent Correct	97.6%
Overall Kappa Statistic	.95

CONCLUSION

The results of this accuracy assessment analysis clearly demonstrate that the enhanced U.S. Department of Agriculture NASS CDL product, produced for the first time in 2007 for Illinois, is sufficiently accurate to be used by natural resource agencies and other organizations needing regional land cover information on a recurring basis. Prior to 2007, detailed non-agricultural land cover was not well characterized by NASS, and the 1999-2002 multiagency-funded project to combine USDA NASS' agricultural land cover with non-agricultural cover separately produced by state personnel was determined to be costly and required an additional year of effort for integration of the two classification maps.

Accurate and updated land cover data recurring on an annual basis will have multiple benefits for resource applications throughout Illinois including ecosystem assessment and planning, wildlife management, relating green and gray infrastructure, disaster mitigation (Figure 6), change analysis, to mention a few. Updated land cover information better describes the current condition of Illinois' cultural and physical landscape, and increases the efficacy of management decisions. Furthermore, because the enhanced CDL data product is also produced annually by NASS for states adjacent to Illinois (except Kentucky), cross-border analyses will benefit by having consistent land cover information. Because the NASS CDL data product is publicly available and is "GIS ready", it is easily accessible and immediately useable in a variety of mapping software. While CDL data products continue to be available for free download at the USDA Geospatial Data Gateway <http://datagateway.nrcs.usda.gov/>, as a direct benefit of this project, these and other land cover data products are now locally available within the state at the Illinois Natural Resources Geospatial Data Clearinghouse <http://www.isgs.uiuc.edu/nsdihome/webdocs/landcover/>. This new Internet resource will ensure the continuity, widespread public access, viewing, and continued free download of Illinois land cover data and associated natural resource information.

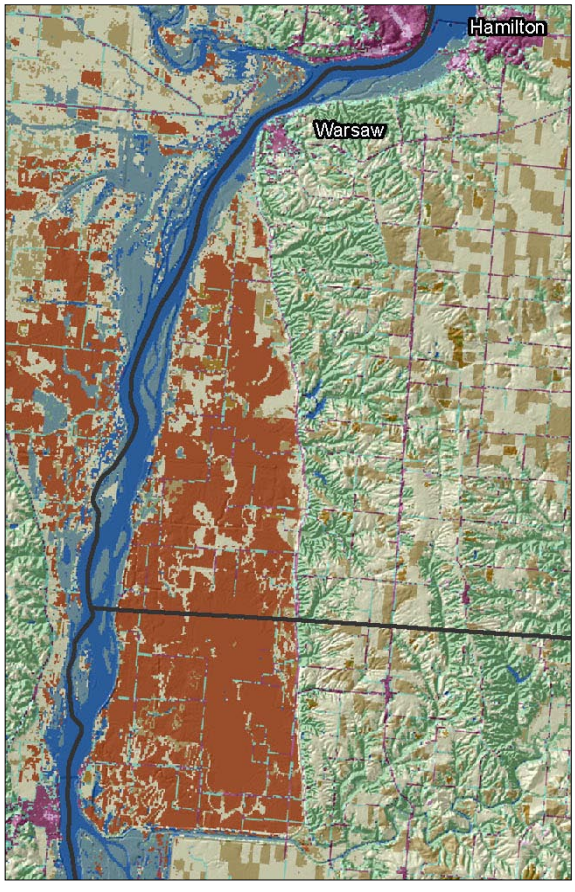
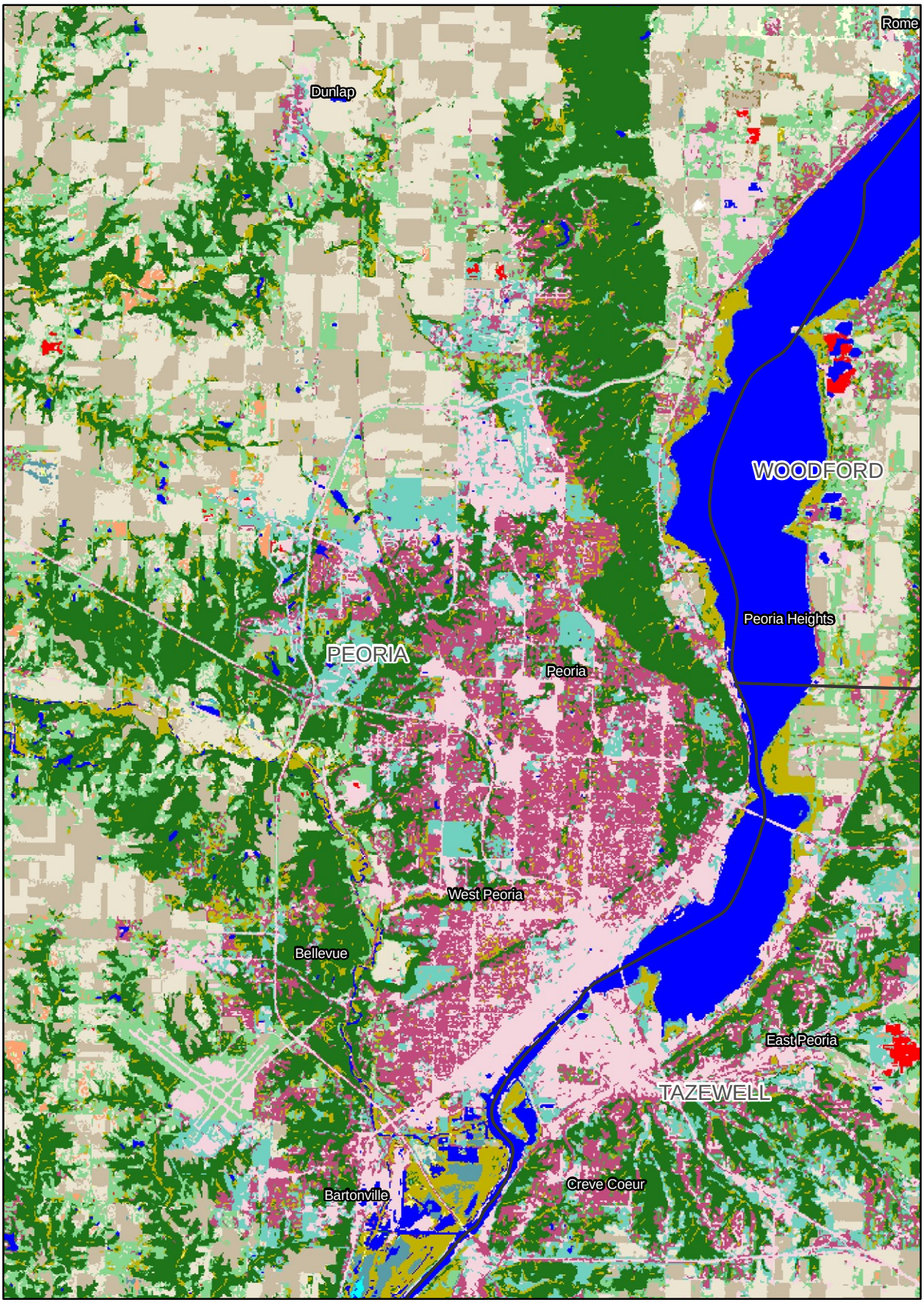


Figure 5. Portion of the Illinois 2008 NASS CDL data product showing a 50 mi² area of damaged and/or destroyed cropland (dark red-orange) resulting from the 2008 summer Midwest Flood event. Geographic location is the Mississippi River flood plain in south-western Hancock County and northwestern Adams County, IL.

Appendix A

Landsat-ResoureSat Comparison

1999-2000 Illinois Interagency Land Cover Project - Peoria and Environs
30-meter Landsat 5 TM and Landsat 7 ETM+ Source Imagery

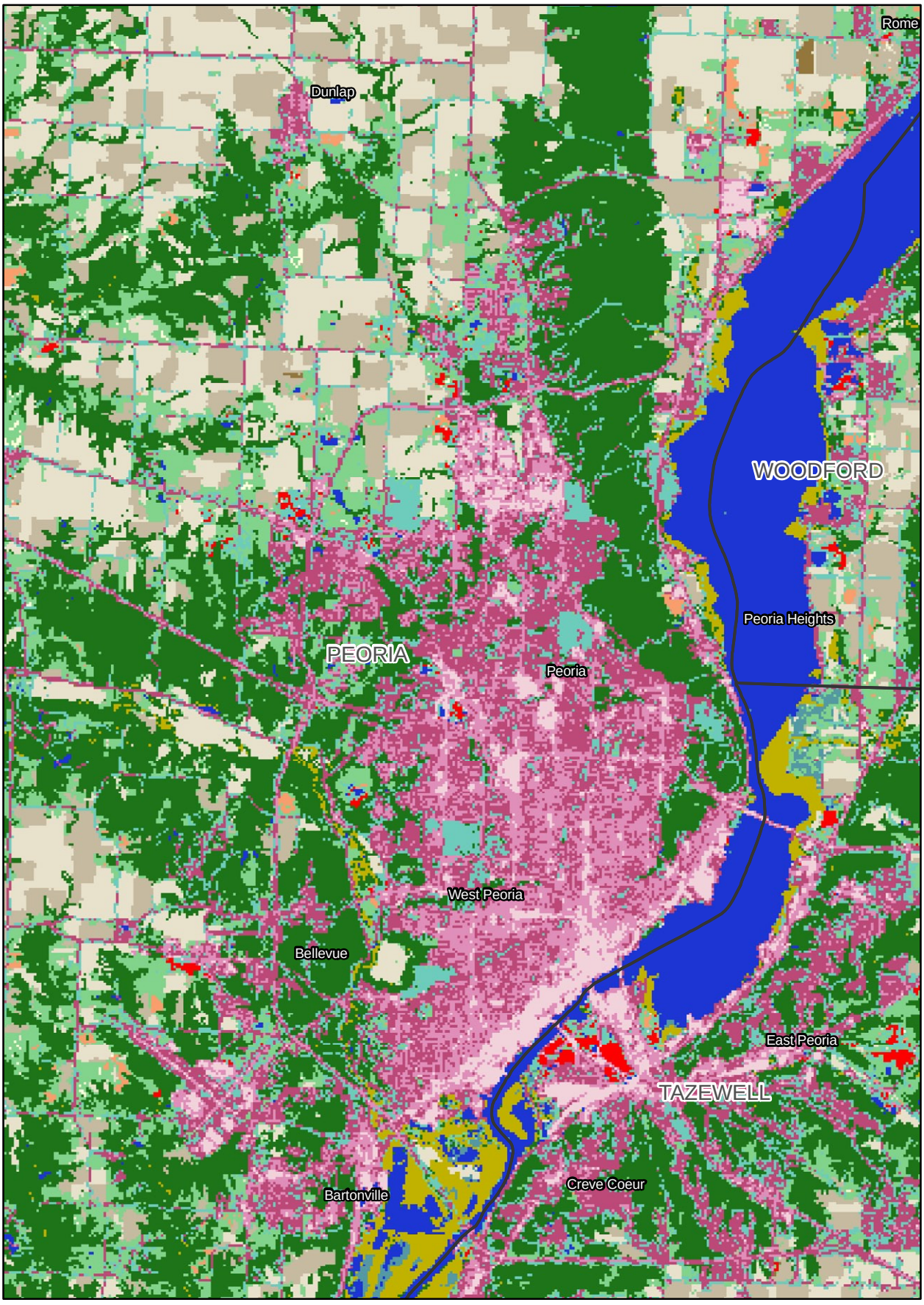


1999-2000 Illinois Land Cover
Class_names

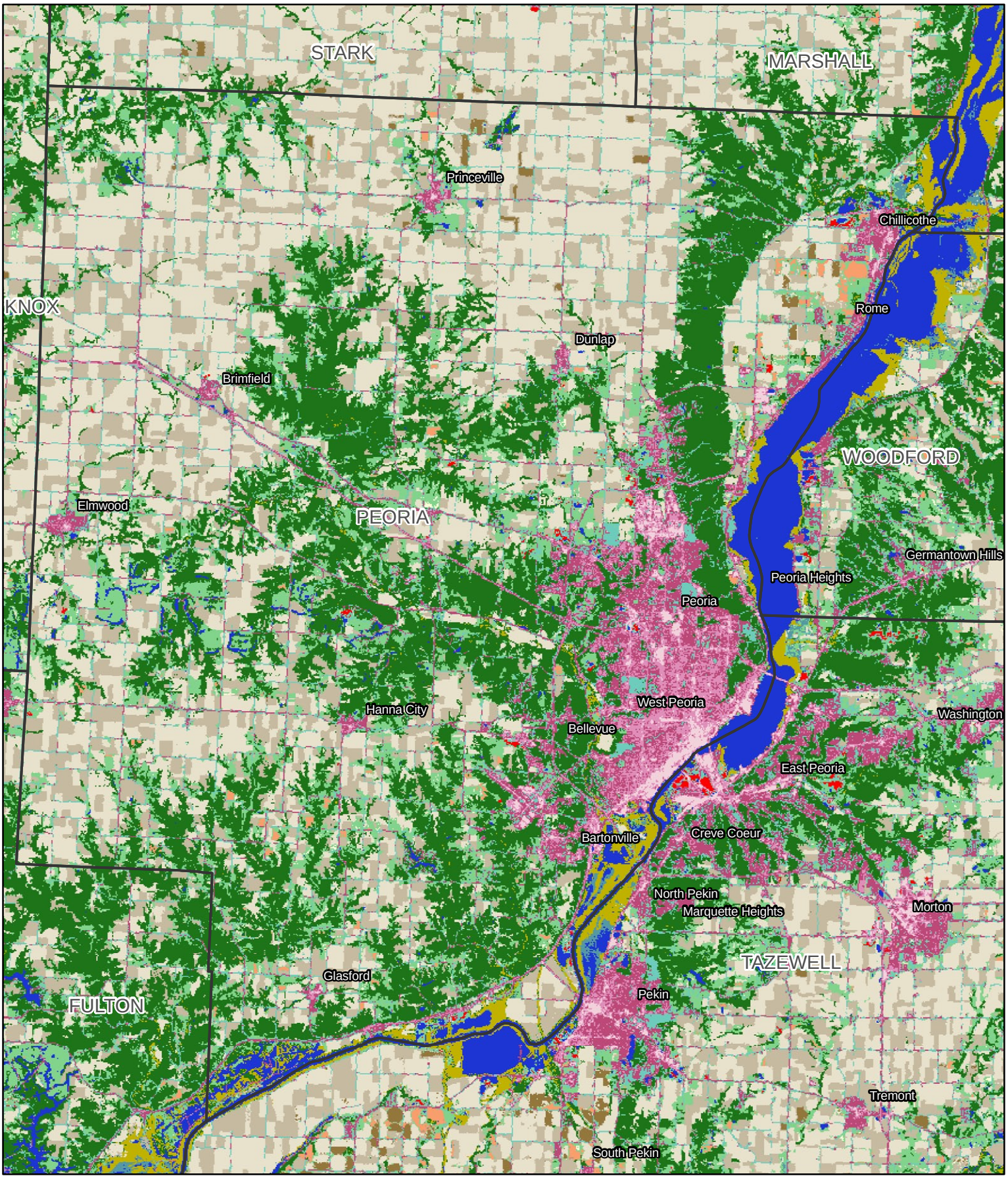
- Barren & Exposed Land
- Coniferous
- Corn
- Deep Marsh
- Floodplain Forest
- High Density
- Low/Medium Density
- Other Agriculture
- Other Small Grains & Hay
- Partial Canopy/Savannah Upland
- Rural Grassland
- Seasonally/Temporarily Flooded
- Shallow Marsh/Wet Meadow
- Shallow Water
- Soybeans
- Surface Water
- Swamp
- Upland
- Urban Open Space
- Winter Wheat
- Winter Wheat/Soybeans

Miles
0 0.5 1 2 3 4

Illinois 2007 USDA NASS Cropland Data Layer - Peoria and Environs
56-meter ResourceSat-1 AWiFS Source Imagery



Illinois 2007 USDA NASS Cropland Data Layer - Peoria and Environs
56-meter ResourceSat-1 AWiFS Source Imagery



Appendix B

NLCD 2001 Definitions

NLCD 2001 Land Cover Definitions

Illinois 2007 NASS CDL Non-Agricultural Land Cover

Code:	Category:
111	**Open Water**
	All areas of open water, generally with less than 25% cover of vegetation or soil.
121	**Developed/Open Space**
	Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20 percent of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.
122	**Developed/Low Intensity**
	Areas characterized by a high percentage (30 percent or greater) of constructed materials (e.g. asphalt, concrete, buildings, etc). Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20-49 percent of total cover. These areas most commonly include single-family housing units.
123	**Developed/Medium Intensity**
	Areas characterized by a high percentage (30 percent or greater) of constructed materials (e.g. asphalt, concrete, buildings, etc). Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50-79 percent of the total cover. These areas most commonly include single-family housing units.
124	**Developed/High Intensity**
	Areas characterized by a high percentage (30 percent or greater) of constructed materials (e.g. asphalt, concrete, buildings, etc). Includes highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80 to100 percent of the total cover.
131	**Barren**
	Areas characterized by bare rock, gravel, sand, silt, clay, or other earthen material, with little or no "green" vegetation present regardless of its inherent ability to support life. Vegetation, if present, is more widely spaced and scrubby than that in the "green" vegetated categories; lichen cover may be extensive. Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover.

141	**Deciduous Forest**
	Areas dominated by trees (natural or semi-natural woody vegetation) generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species shed foliage simultaneously in response to seasonal change.
142	**Evergreen Forest**
	Areas dominated by trees (natural or semi-natural woody vegetation) generally greater than 5 meters tall, and greater than 20% of total vegetation cover. More than 75 percent of the tree species maintain their leaves all year. Canopy is never without green foliage.
143	**Mixed Forest**
	Areas dominated by trees (natural or semi-natural woody vegetation) generally greater than 5 meters tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75 percent of total tree cover.
152	**Shrubland**
	Areas characterized by natural or semi-natural woody vegetation with aerial stems, generally less than 6 meters tall, with individuals or clumps not touching to interlocking. Both evergreen and deciduous species of true shrubs, young trees, and trees or shrubs that are small or stunted because of environmental conditions are included.
171	**Grassland Herbaceous**
	Upland areas characterized by natural or semi-natural herbaceous vegetation. Areas dominated by grammanoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.
181	**Pasture/Hay**
	Upland areas characterized by natural or semi-natural herbaceous vegetation. Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 percent of total vegetation.
190	**Woody Wetlands**
	Areas where the soil or substrate is periodically saturated with or covered with water. Areas where forest or shrubland vegetation accounts for greater than 20 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.
195	**Emergent Herbaceous Wetlands**
	Areas where the soil or substrate is periodically saturated with or covered with water. Areas where perennial herbaceous vegetation accounts for 75-100 percent of the cover and the soil or substrate is periodically saturated with or covered with water.

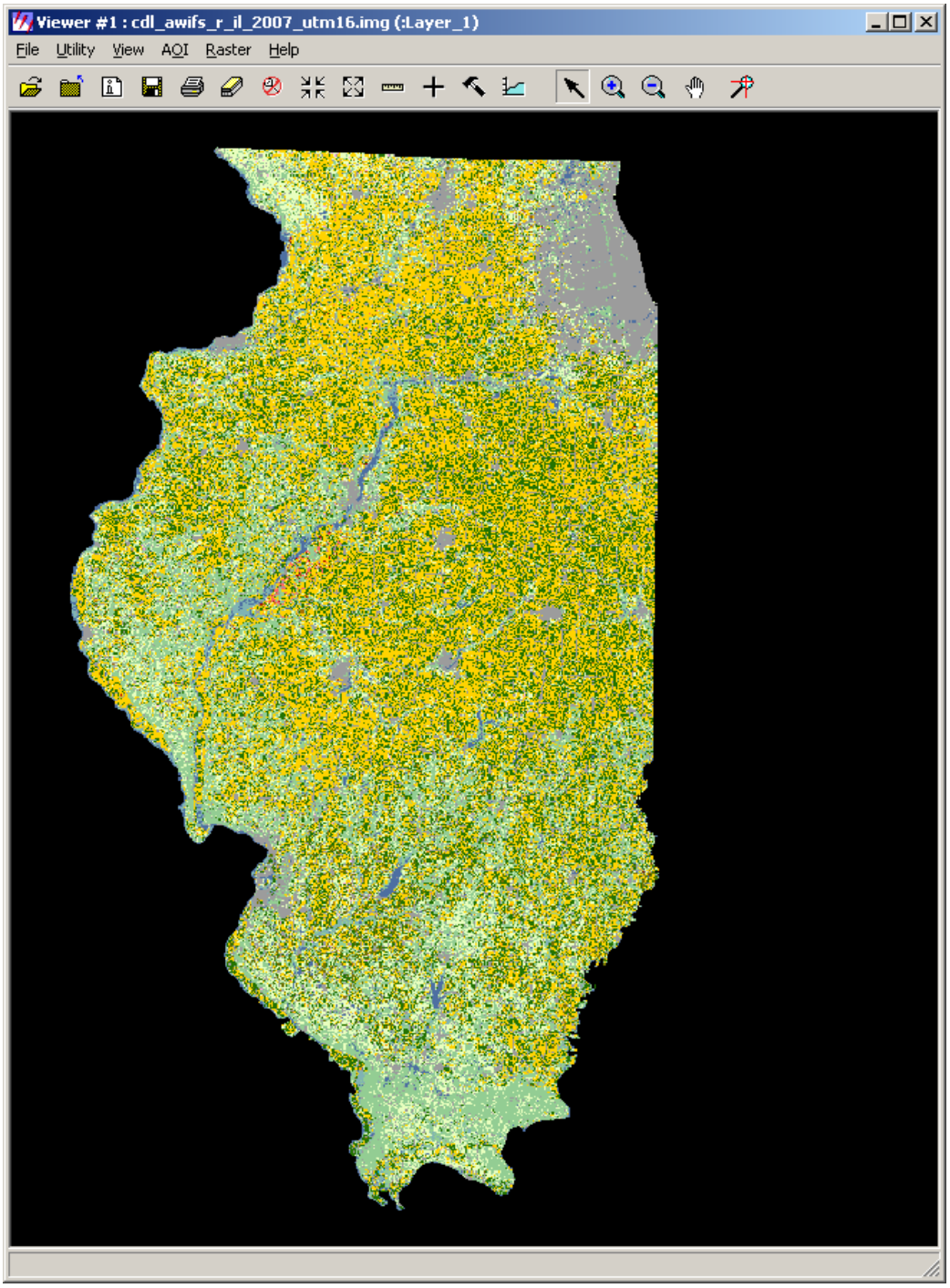
Appendix C

Accuracy Assessment Procedures

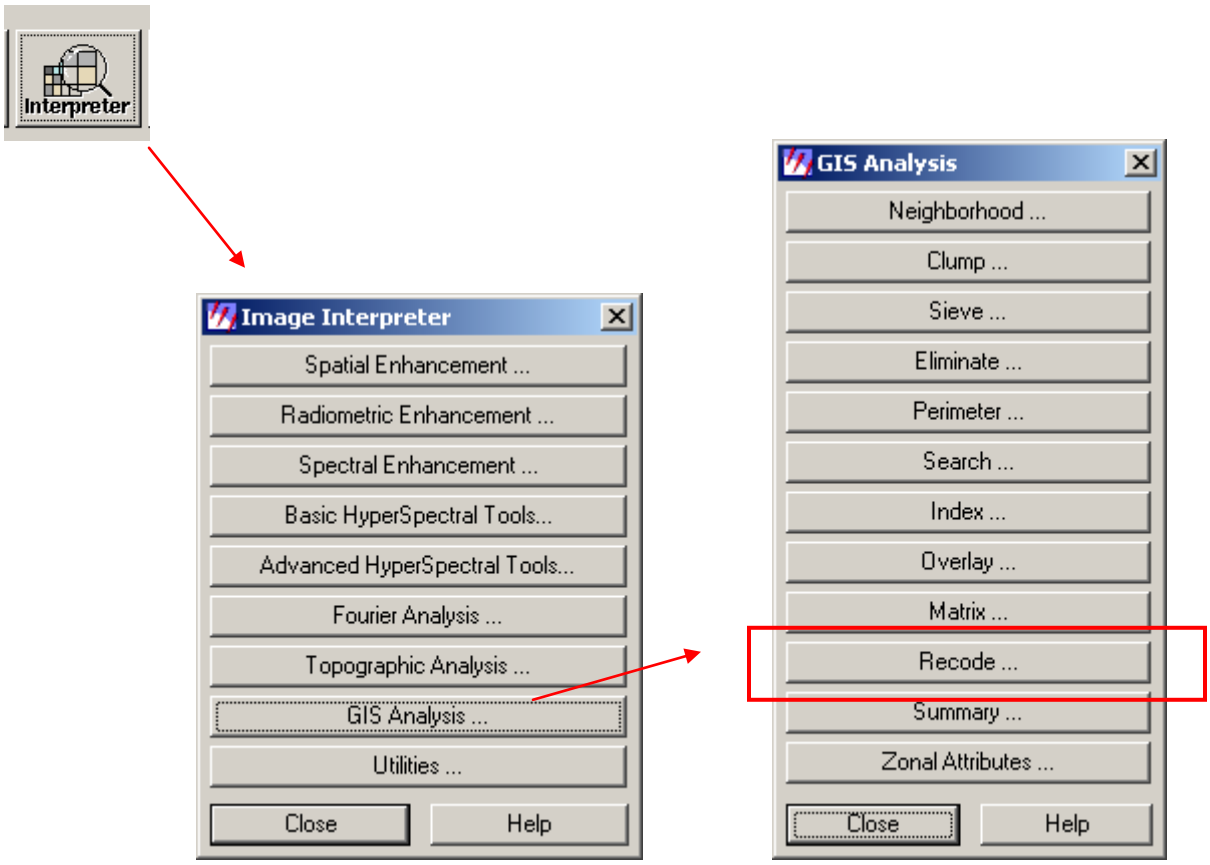
Process Steps and Procedures for 2007 Illinois NASS-CDL Accuracy Assessment of Non-Agriculture Categories

(1) Recode the 2007 Illinois NASS-CDL to the 11 categories used to conduct the accuracy assessment

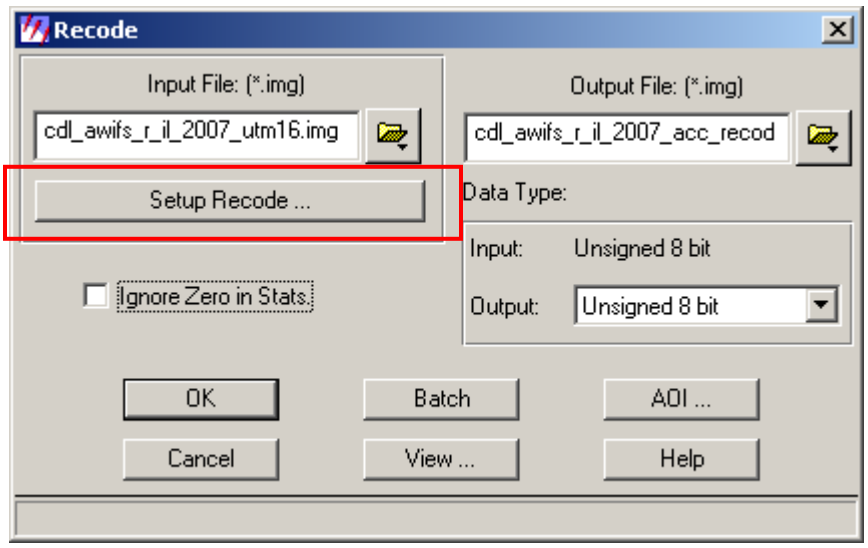
In Imagine, open **cdl_awifs_r_il_2007_utm16.img** in a Viewer window



Click on **Interpreter – GIS Analysis - Recode**



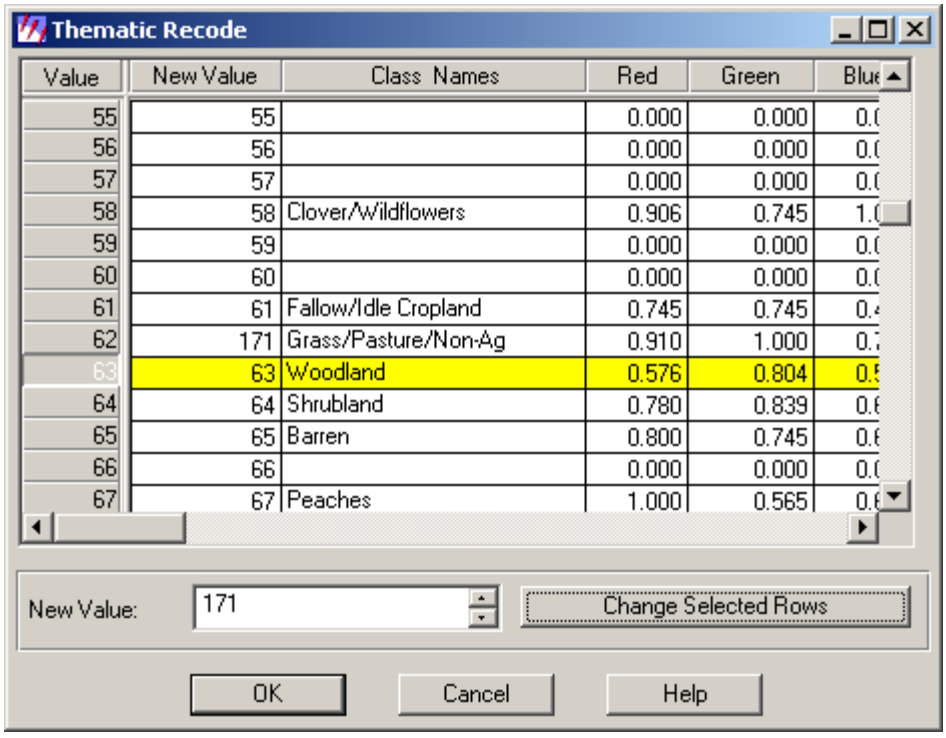
Input file is **cdl_awifs_r_il_2007_utm16.img**
Output file is **cdl_awifs_r_il_2007_acc_recode.img**



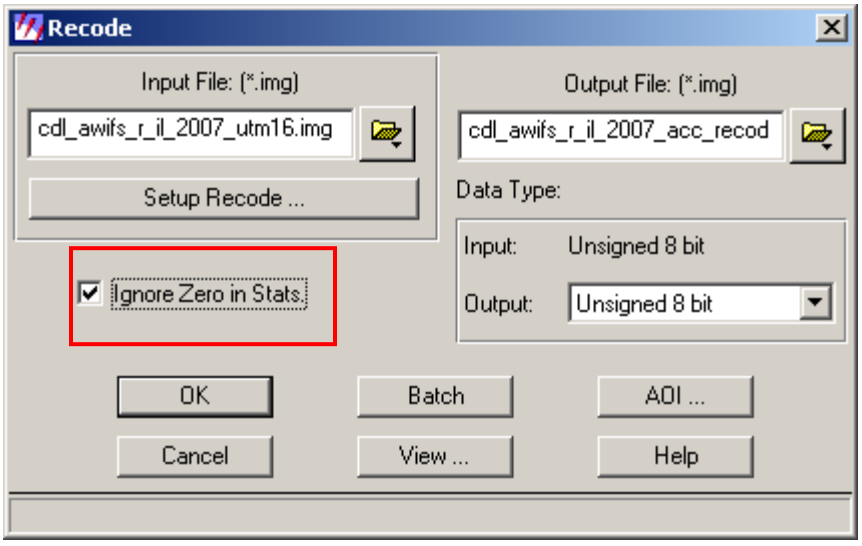
Click on **Setup Recode**. This opens a new window.

Select a category to recode (62 – Grass/Pasture/Non-Ag). Type in the new Value (181) and click on **change selected rows**. Repeat for each category that needs to be recoded (see above).

Then click OK



Make sure to check the Ignore Zero in Stats box
Then click OK

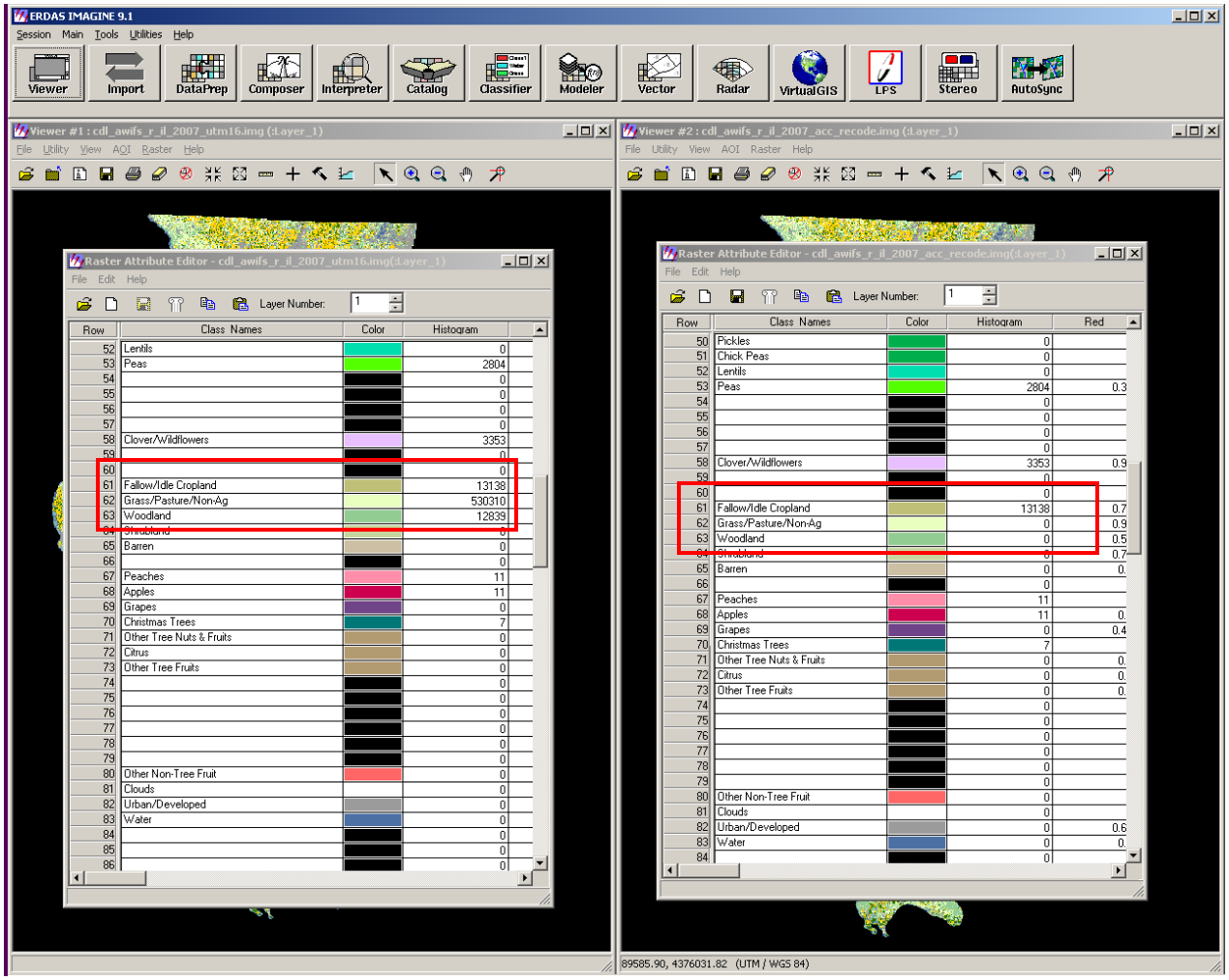


Open the recoded image in a new window (**cdl_awifs_r_il_2007_acc_recode.img**) and double check histogram to make sure image was recoded correctly.

Need to add the Class Name column back into the attribute table as it has been removed after running the recode function.

Click on **Raster – Attributes** to view the attribute table.
Click on **Edit – Add Class Names**. Then Copy and Paste the Class Names column from the original image.

Next, double checks the histogram to make sure values were recoded correctly.



(2) Create a shapefile of the 40 counties that will be used in the accuracy assessment

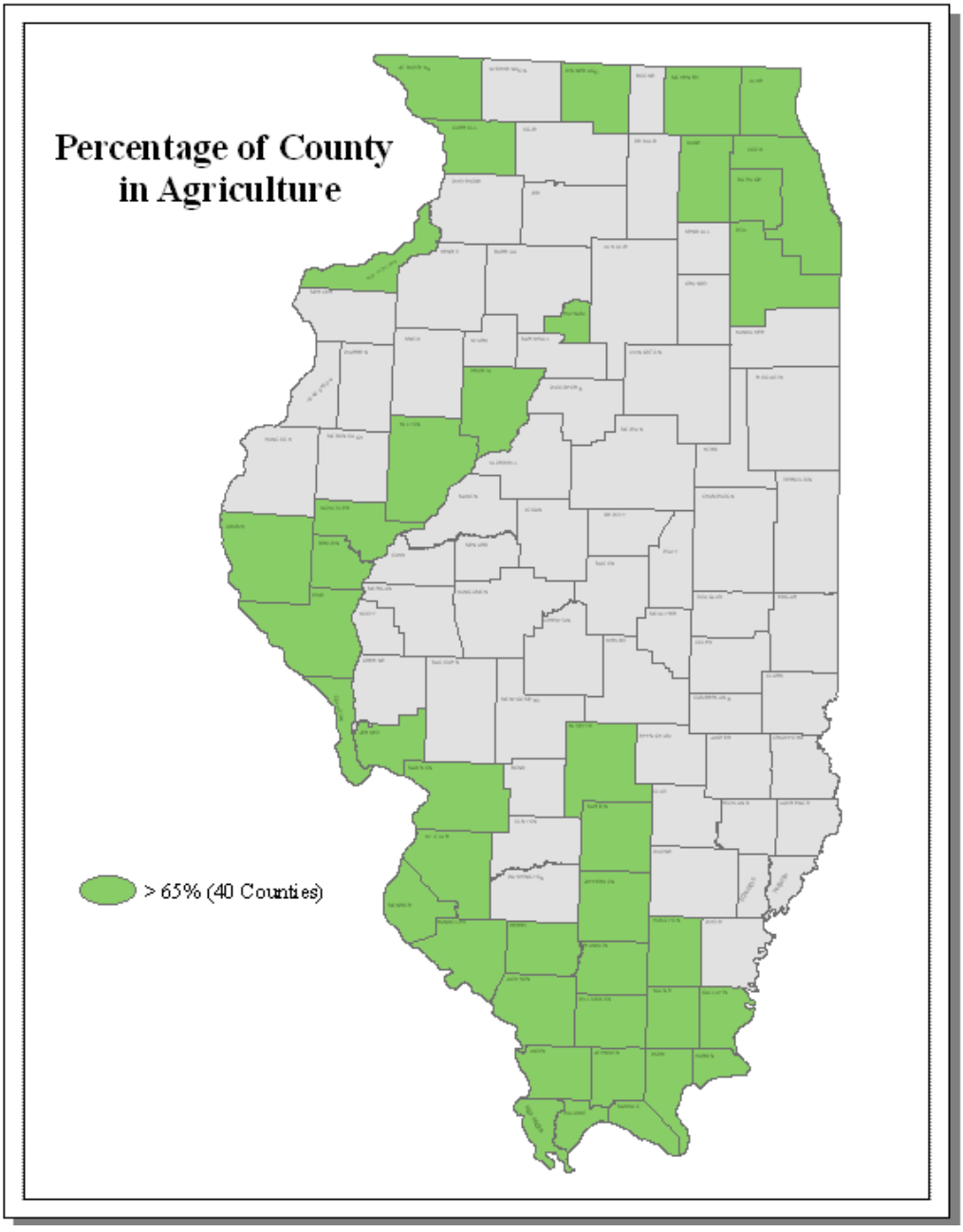
In ArcMap, load the counties feature class (IL_BNDY_County_Py)
Select the 40 counties from **lc99_00_ag_rank_w_incnty_2008.xls**

Right Click on the feature class (IL_BNDY_County_Py)in the Table of Contents and select Export – Data

Save file as lc_40acc_counties.shp

In ArcCatalog (Data Management Tools – Projections and Transformations – Feature – Project) reproject lc_40acc_coutnies.shp from GCS – Lat, Long to UTM, Zone 16 – Datum WGS 1984.

Save file as **lc_40acc_counties_utm16_wgs84.shp**



(3) Subset (or clip) the recoded NASS-CDL layer to include only the counties that will be used in the accuracy assessment

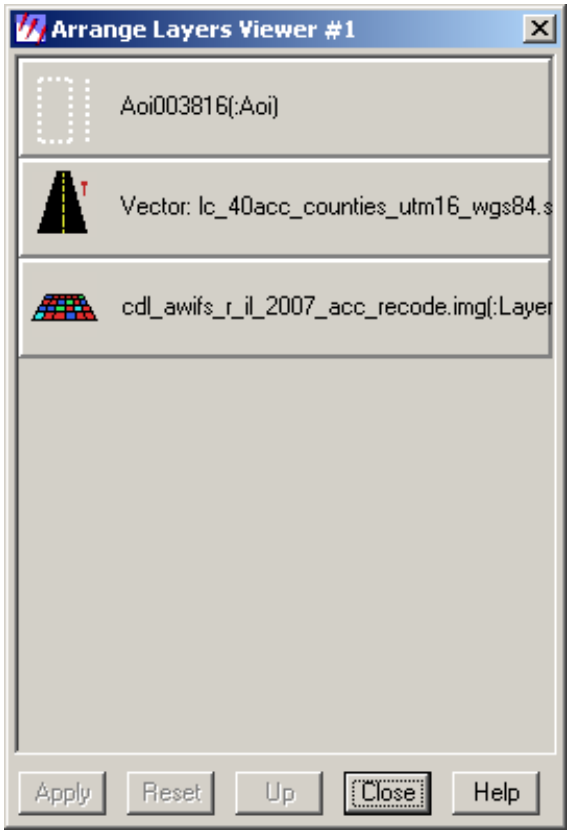
In Imagine, open **cdl_awifs_r_il_2007_acc_recode.img** in a Viewer window

Open the vector layer **lc_40acc_counties_utm16_wgs84.shp** on top of re-coded NASS-CDL Land Cover layer

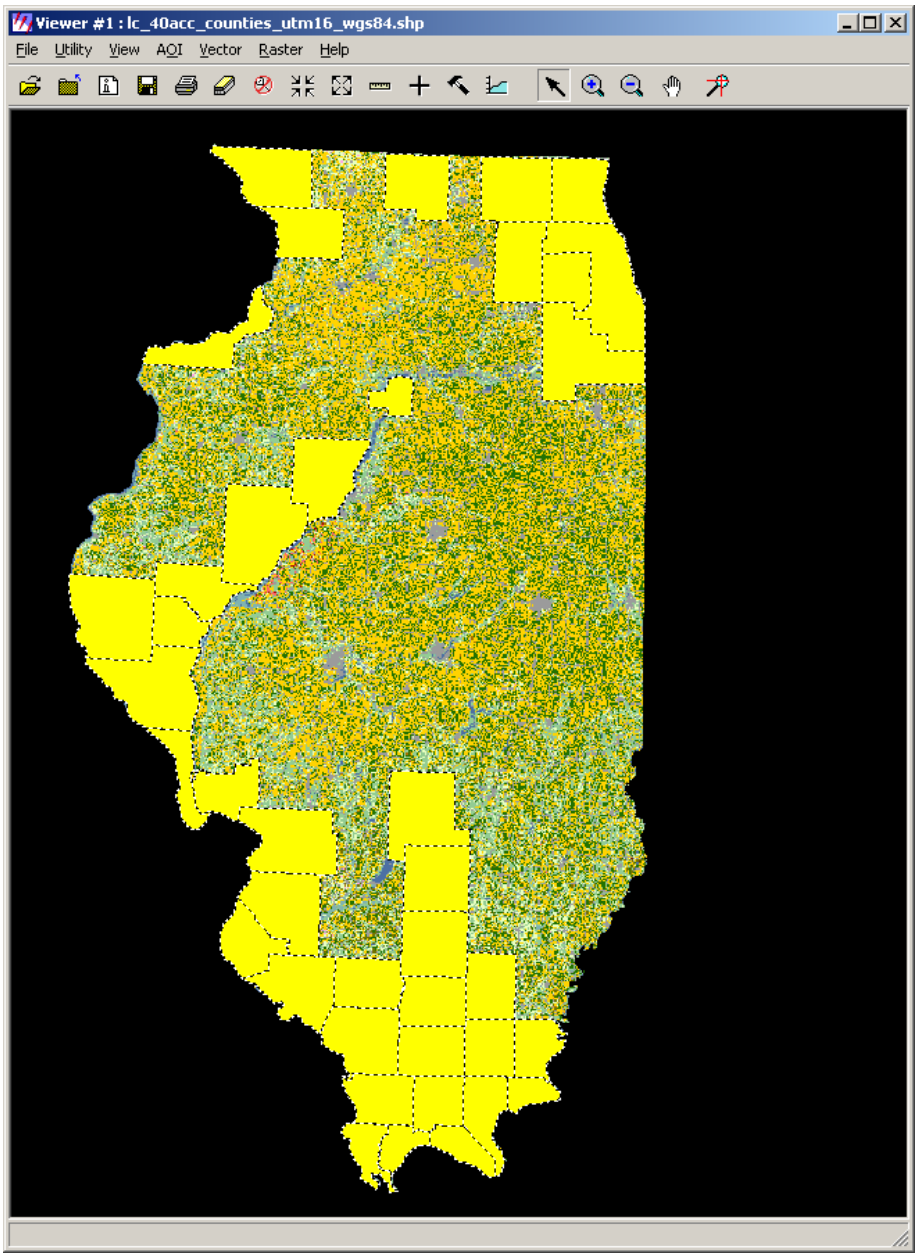
Open the attribute table for the counties by selecting Vector – Attributes from the menu. Right click on the records and select all.

Select **File --- New --- AOI Layer**

To make sure you have created a new AOI layer select View Arrange layers to view all layers that are currently opened in the viewer.

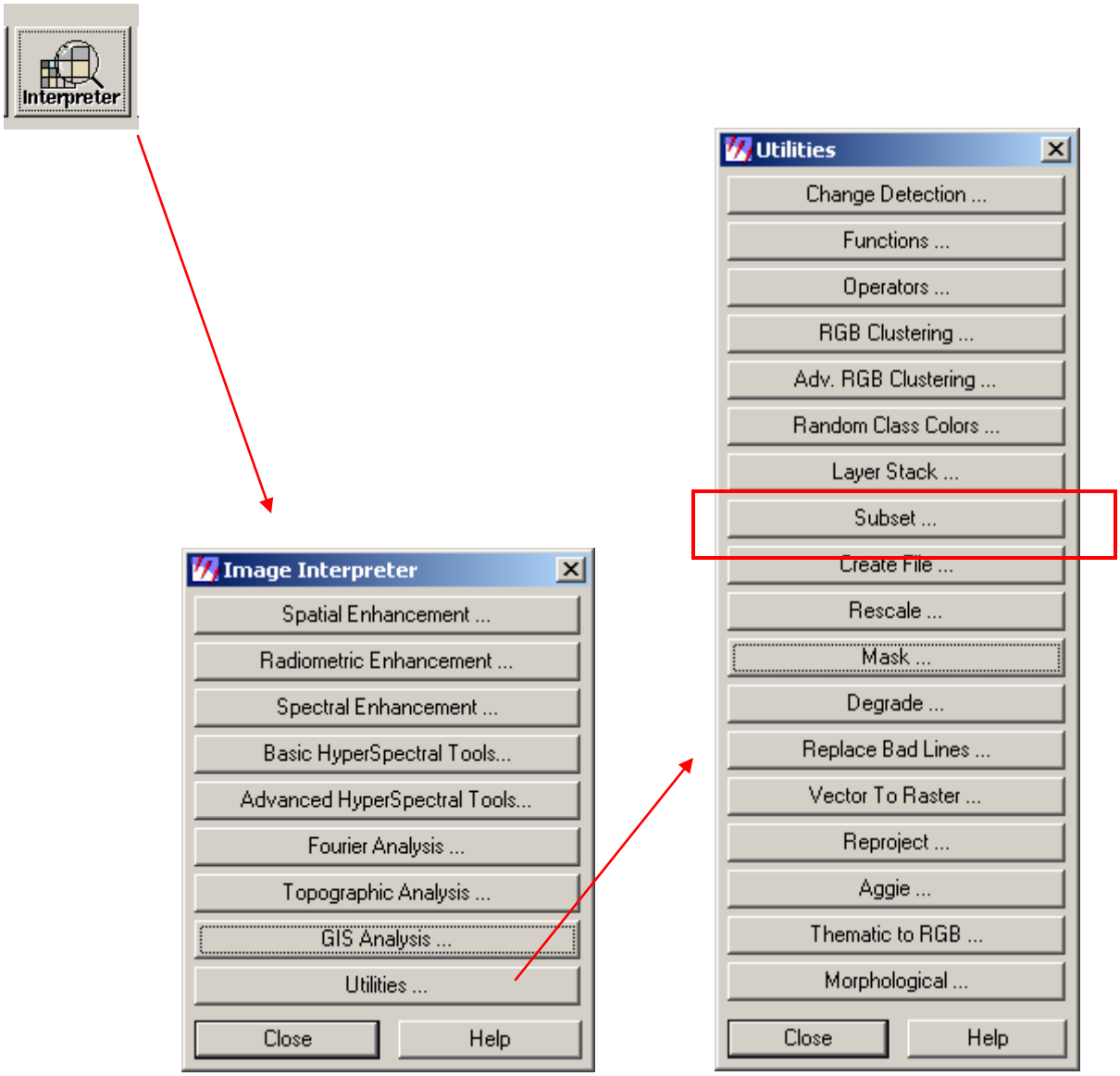


Select **AOI --- Copy Selection to AOI….** You should see dashed lines around all of the selected counties.

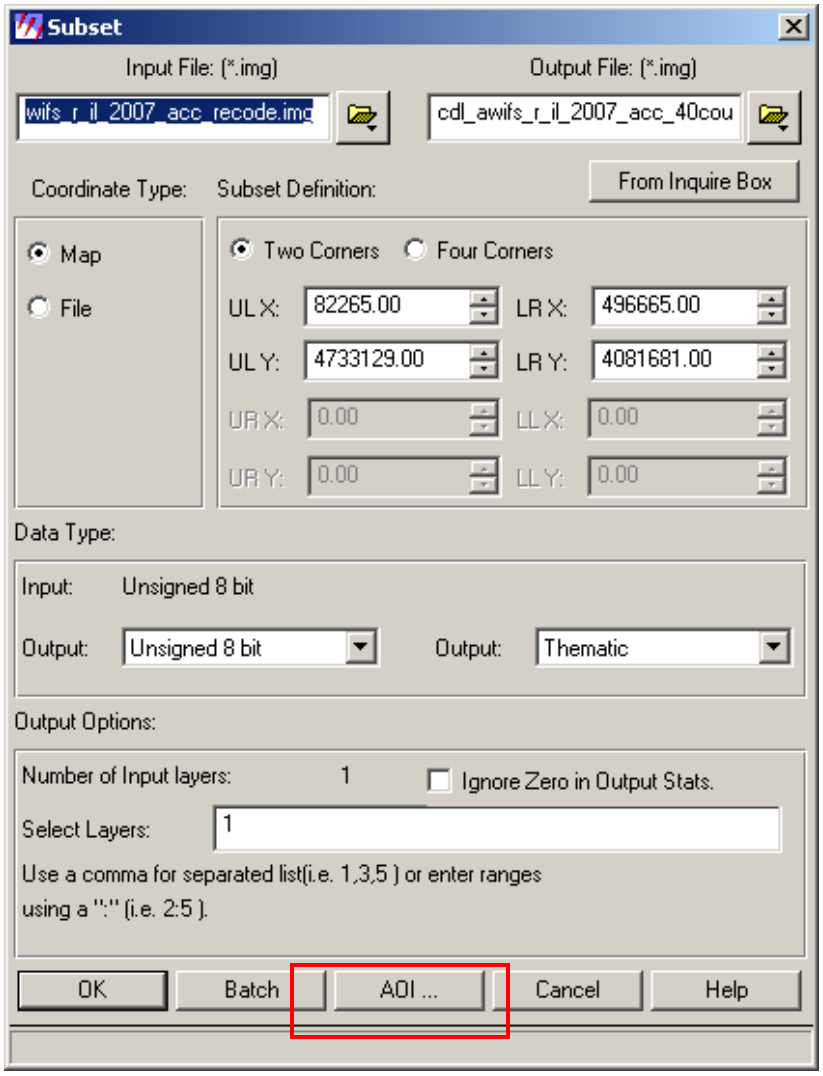


Then select **File --- Save AOI As… (lc_40acc_counties.aoi)**

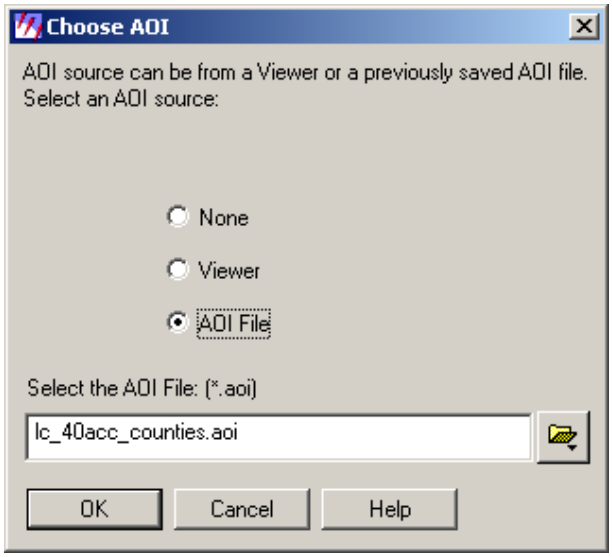
Click on **Interpreter – Utilities – Subset**



Input file is **cdl_awifs_r_il_2007_acc_recode.img**
Output file is **cdl_awifs_r_il_2007_acc_40counties.img**

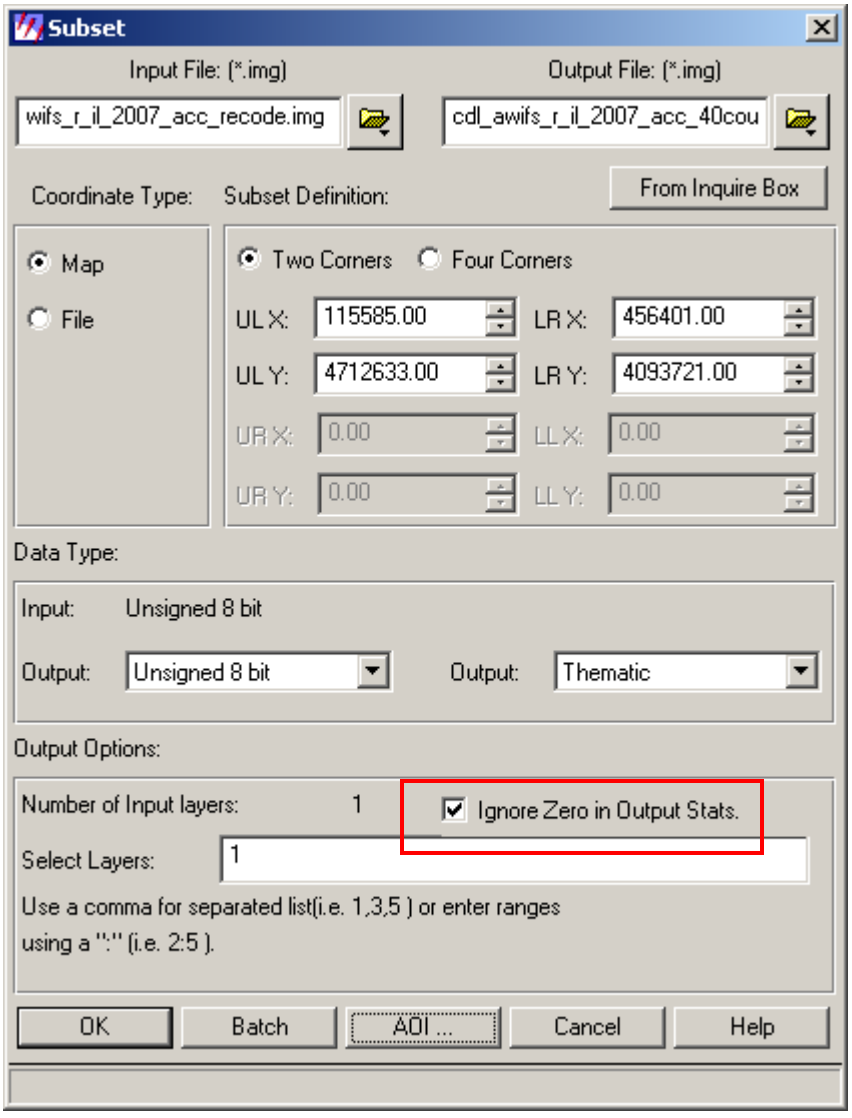


Click on **AOI** and select the **AOI File (lc_40acc_counties.aoi)**



Click OK

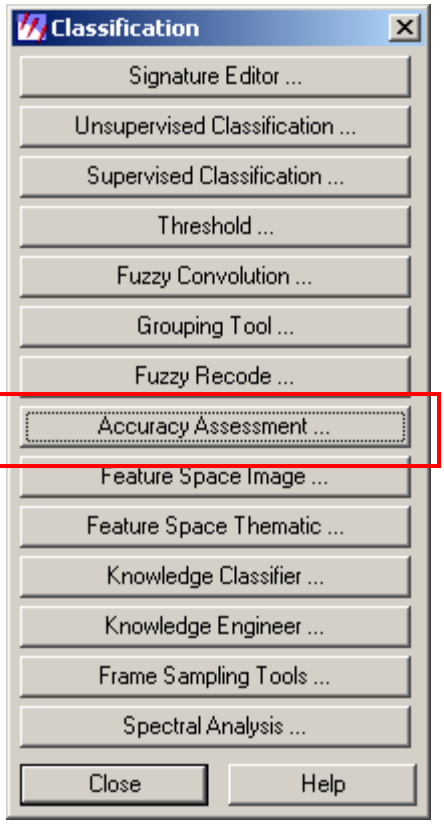
Make sure to check the **Ignore Zero** in Stats box
Then click OK



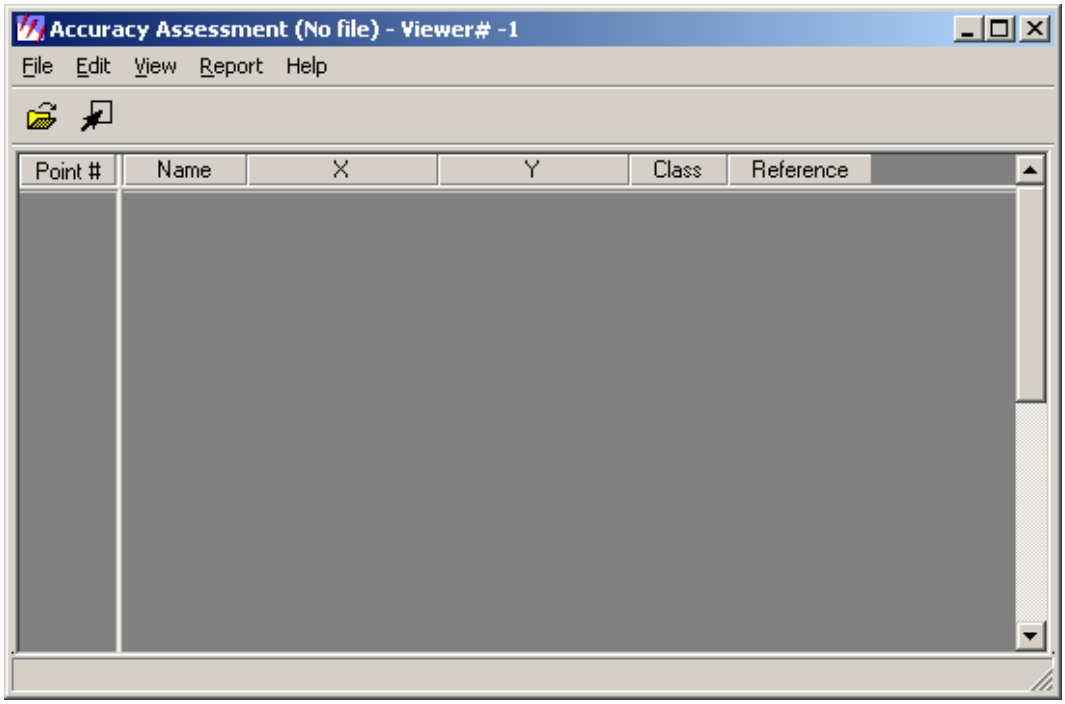
Next, double check the subset image to make sure it was clipped properly.

(4) Generate a specified number of stratified random sampling points to be used for the accuracy assessment. *We determined we would create a minimum of 70 points per category. (11 categories x 70 points per category = 770 points. Creating 825 will allow us extra 55 extra points in can we need to throw out any duplicate points or points that occur may occur in agricultural areas)*

In Imagine, click on **Classifier --- Accuracy Assessment**

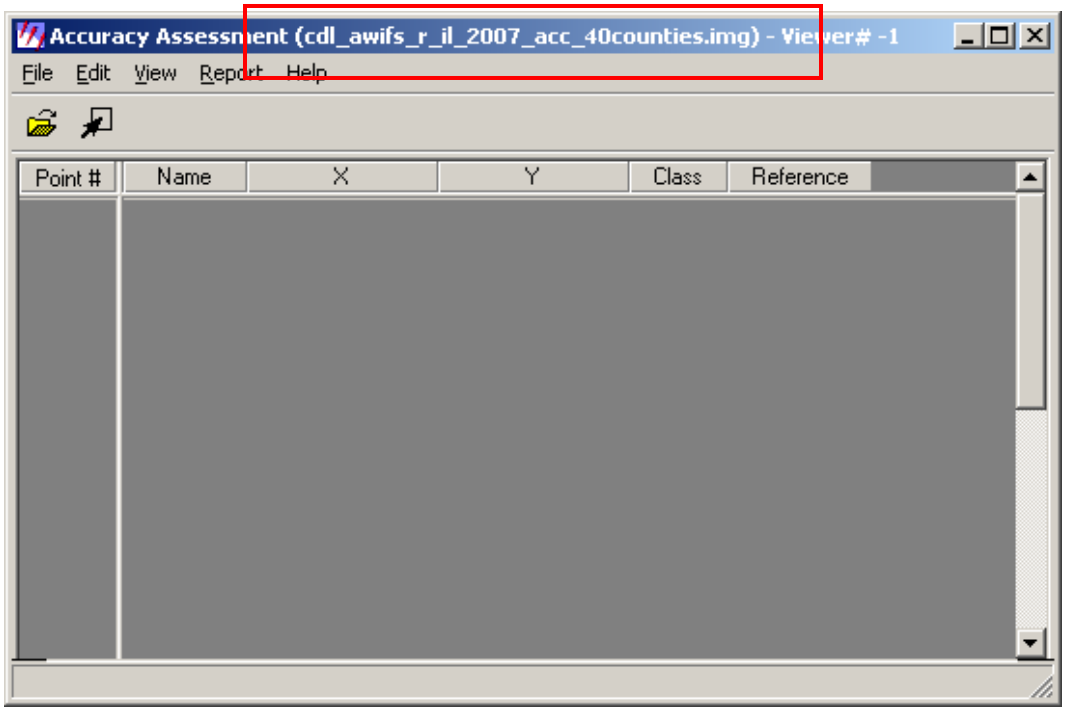


This brings up a blank “cell array” window.



In the Accuracy Assessment Window click on **File --- Open**

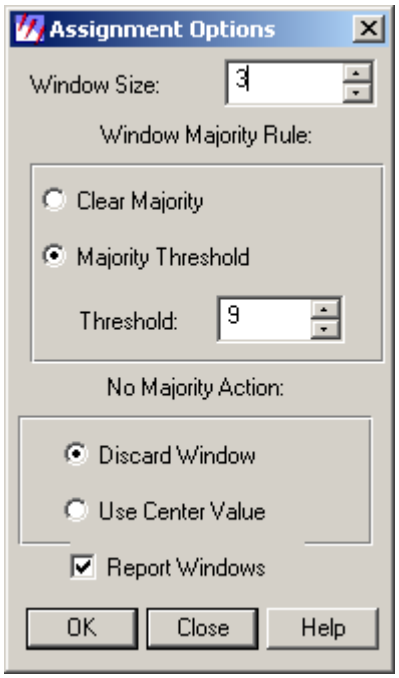
This will allow the user to open a *.img file. The file you want to open is your classified image. NOTE: The points are saved within the classified image you specify and if you have already saved accuracy assessment points to this same file, then those saved point will open up in the Accuracy Assessment viewer.



Then click on **Edit --- Class Value Assignment Options**

The following criteria will be used

- window size = 3 x 3
- window majority rule =majority threshold
- majority threshold = 9
- no majority action = discard window



Click OK

Then click on **Edit --- Create/Add Random Points**

This dialog enables you to randomly select points from the classified image to be used in the accuracy assessment process.

A search window is used to derive the class value for a selected pixel. The center pixel of the search window is the selected pixel if the pixel values in the window meet the criteria set by you. For example, you may specify that a majority of pixel values must exist in the search window in order for it to be used. If a majority does not exist, then the center pixel for the search window is not selected and the window is discarded. If a majority does exist, then the center pixel is selected and it is assigned the majority class value.

After the points are generated, you must enter the estimated class values for the points into the Accuracy Assessment CellArray. These reference values are compared to the values that were assigned when the pixels were selected.

Select **Edit -> Create/Add Random Points** from the Accuracy Assessment dialog menu bar to open this dialog.

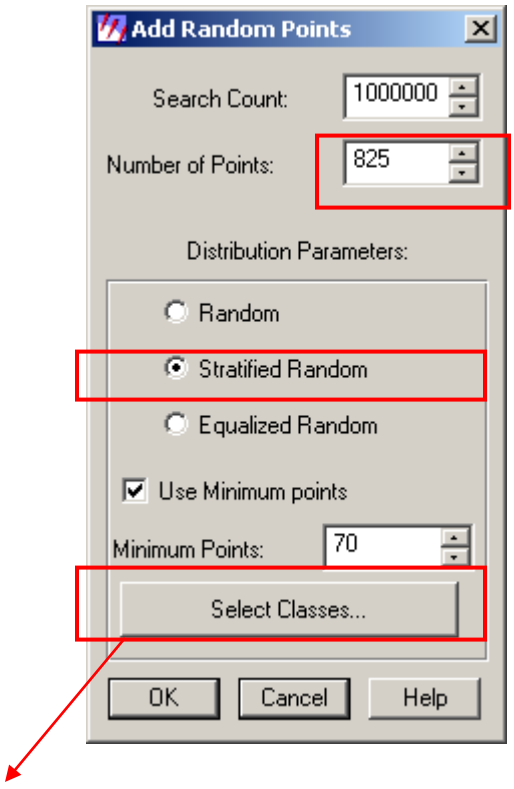
Search Count This is the number of search windows that will be used for gathering the random points per run. For example, if you enter 100, then a maximum of 100 search windows will be used throughout the image for gathering the points. (**1,000,000**)

NOTE: If you are having trouble (that is, the application exhausts the search count quickly or gets only a few points per run) increase this number to a very large one so you do not need to interact with the application often.

Number of Points: Enter the total number of random points to be generated. (**825**)

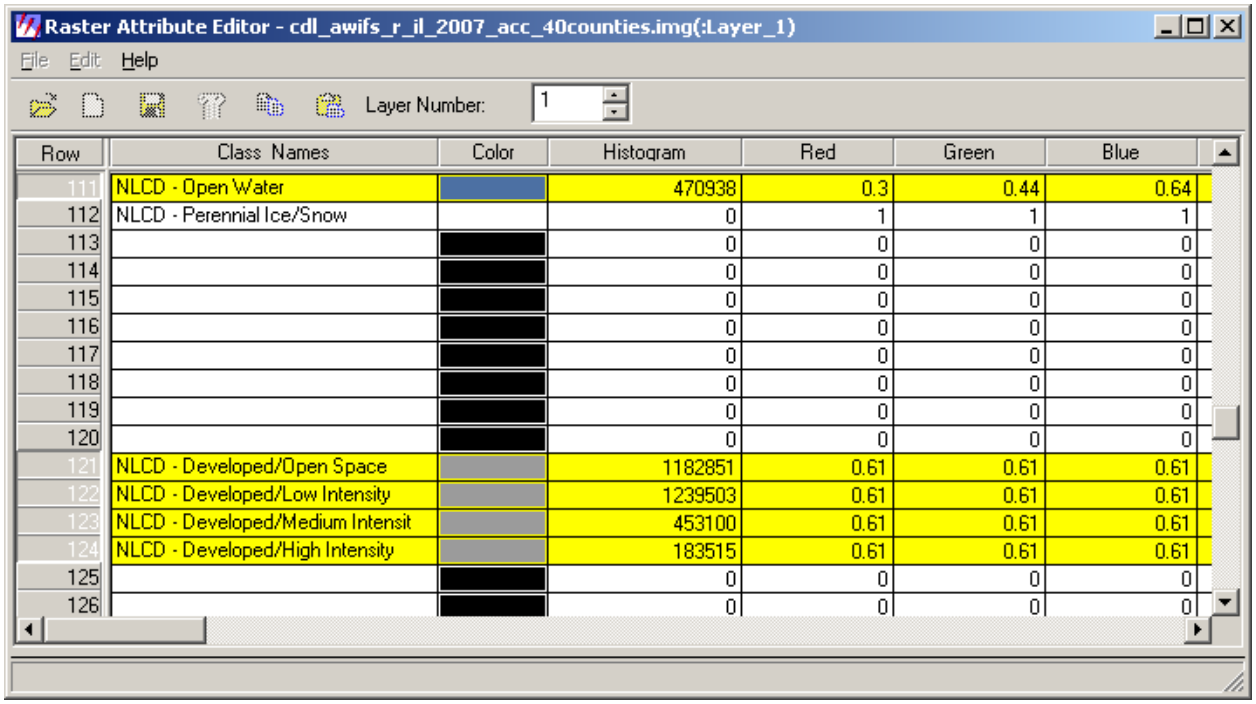
Select **Stratifed Random** from the **Distribution Parameters:**

NOTE: If you select Stratified Random above, you can select the minimum number of points for each class. This ensures that the smallest class has enough points for a true measure of accuracy assessment. (**70**)



Then Click the **Select Classes** box to specify the classes in the thematic layer to be used in selecting the random points. The Raster Attribute Editor opens. Select the classes in the Raster Attribute Editor CellArray that you want to be used in selecting random points. By default, all of the classes are selected. Choose the following:

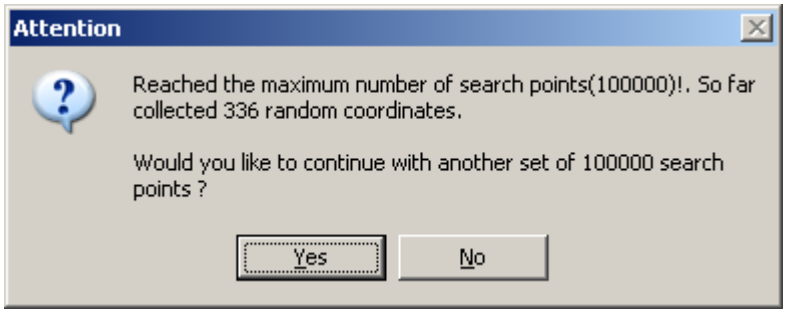
"111" NLCD - Open Water
"121" NLCD - Developed/Open Space
"122" NLCD - Developed/Low Intensity
"123" NLCD - Developed/Medium Intensity
"124" NLCD - Developed/High Intensity
"131" NLCD - Barren
"141" NLCD - Deciduous Forest
"142" NLCD - Evergreen Forest
"171" NLCD - Grassland Herbaceous
"190" NLCD - Woody Wetlands
"195" NLCD - Herbaceous Wetlands



Row	Class Names	Color	Histogram	Red	Green	Blue
111	NLCD - Open Water		470938	0.3	0.44	0.64
112	NLCD - Perennial Ice/Snow		0	1	1	1
113			0	0	0	0
114			0	0	0	0
115			0	0	0	0
116			0	0	0	0
117			0	0	0	0
118			0	0	0	0
119			0	0	0	0
120			0	0	0	0
121	NLCD - Developed/Open Space		1182851	0.61	0.61	0.61
122	NLCD - Developed/Low Intensity		1239503	0.61	0.61	0.61
123	NLCD - Developed/Medium Intensit		453100	0.61	0.61	0.61
124	NLCD - Developed/High Intensity		183515	0.61	0.61	0.61
125			0	0	0	0
126			0	0	0	0

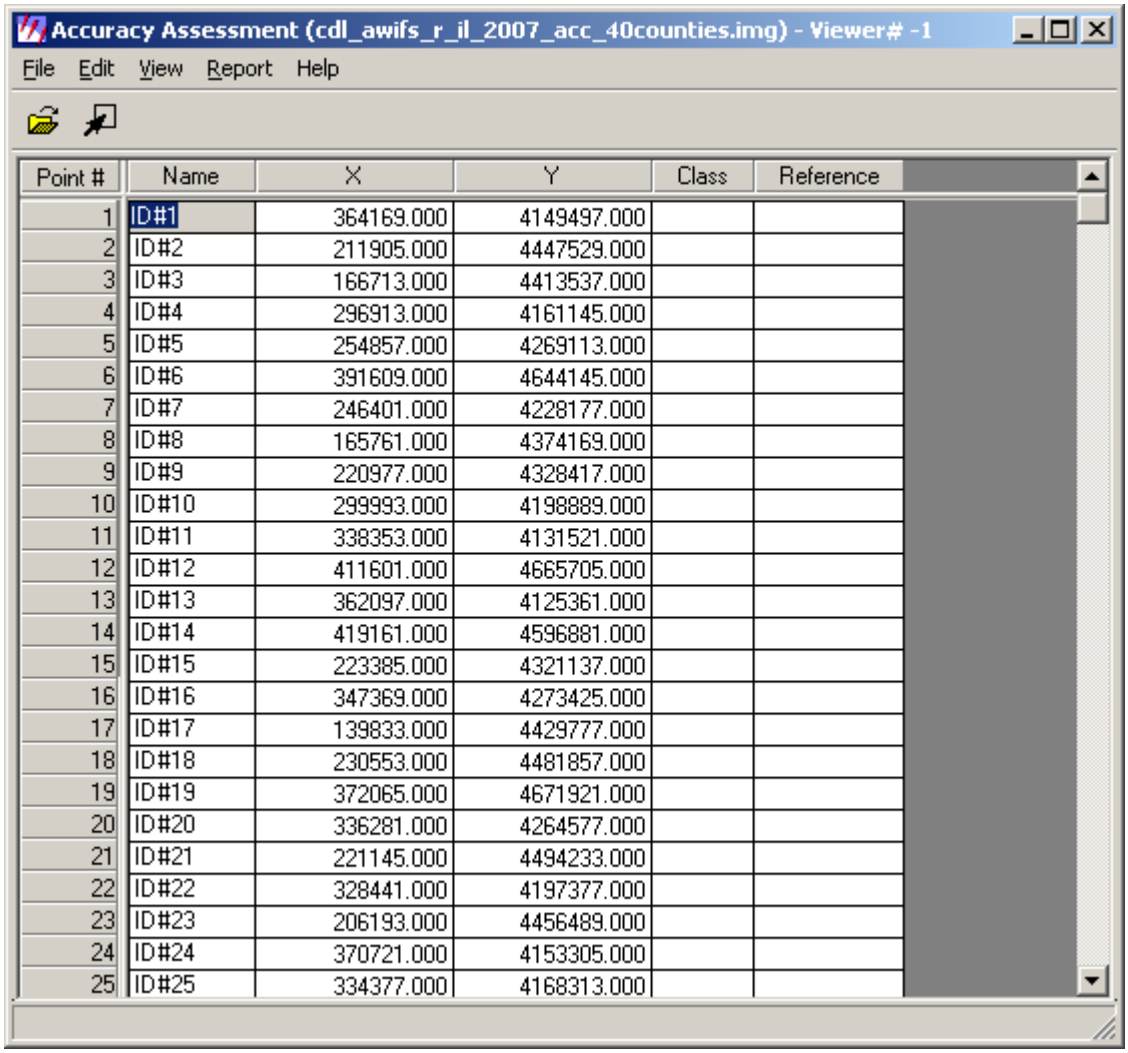
DO not close the Raster Attribute Editor, it closes automatically.

Click **OK** to run this process with the options selected and close this dialog.



If the above box appears click **Yes** until the program reaches the number of random point requested. NOTE: This may take clicking OK many times.

Once you reach the desired number you will see the points appear in the Accuracy Assessment Window



Point #	Name	X	Y	Class	Reference
1	ID#1	364169.000	4149497.000		
2	ID#2	211905.000	4447529.000		
3	ID#3	166713.000	4413537.000		
4	ID#4	296913.000	4161145.000		
5	ID#5	254857.000	4269113.000		
6	ID#6	391609.000	4644145.000		
7	ID#7	246401.000	4228177.000		
8	ID#8	165761.000	4374169.000		
9	ID#9	220977.000	4328417.000		
10	ID#10	299993.000	4198889.000		
11	ID#11	338353.000	4131521.000		
12	ID#12	411601.000	4665705.000		
13	ID#13	362097.000	4125361.000		
14	ID#14	419161.000	4596881.000		
15	ID#15	223385.000	4321137.000		
16	ID#16	347369.000	4273425.000		
17	ID#17	139833.000	4429777.000		
18	ID#18	230553.000	4481857.000		
19	ID#19	372065.000	4671921.000		
20	ID#20	336281.000	4264577.000		
21	ID#21	221145.000	4494233.000		
22	ID#22	328441.000	4197377.000		
23	ID#23	206193.000	4456489.000		
24	ID#24	370721.000	4153305.000		
25	ID#25	334377.000	4168313.000		

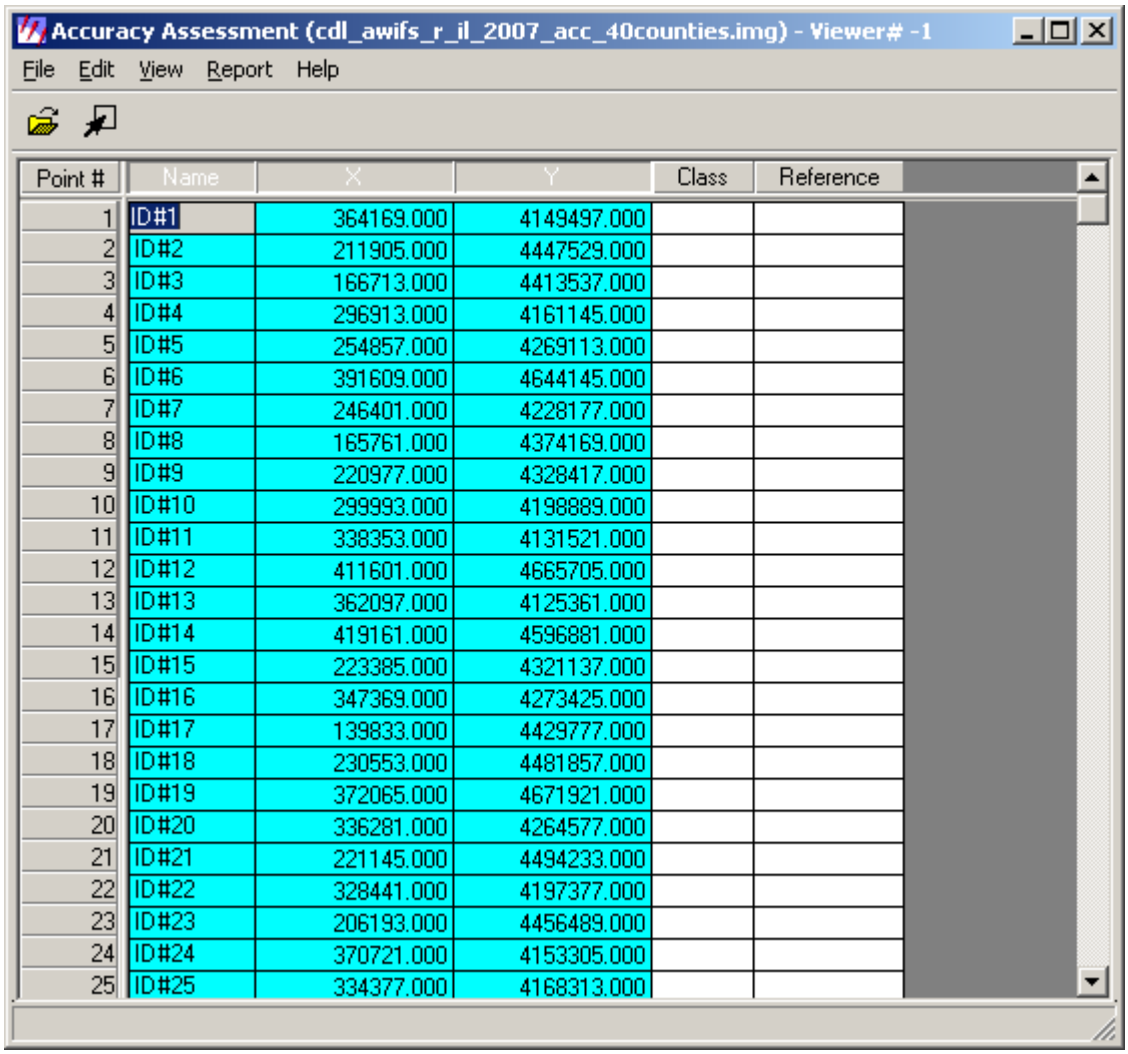
Select **File --- Save Table**. This will save the points to the image.

To check to make sure the class value assignments were used you can create a cell report for each of the points.

Select **Report --- Cell Report** from the menu

Then select File --- Save As to save as an ASCII text file.
(**40counties_825pts_3x3_cellreport.txt**)

Select all the columns you want to use and then right click and select Export to a *.dat file or copy and paste into excel and save as a both and *.xls and *.dbf file.

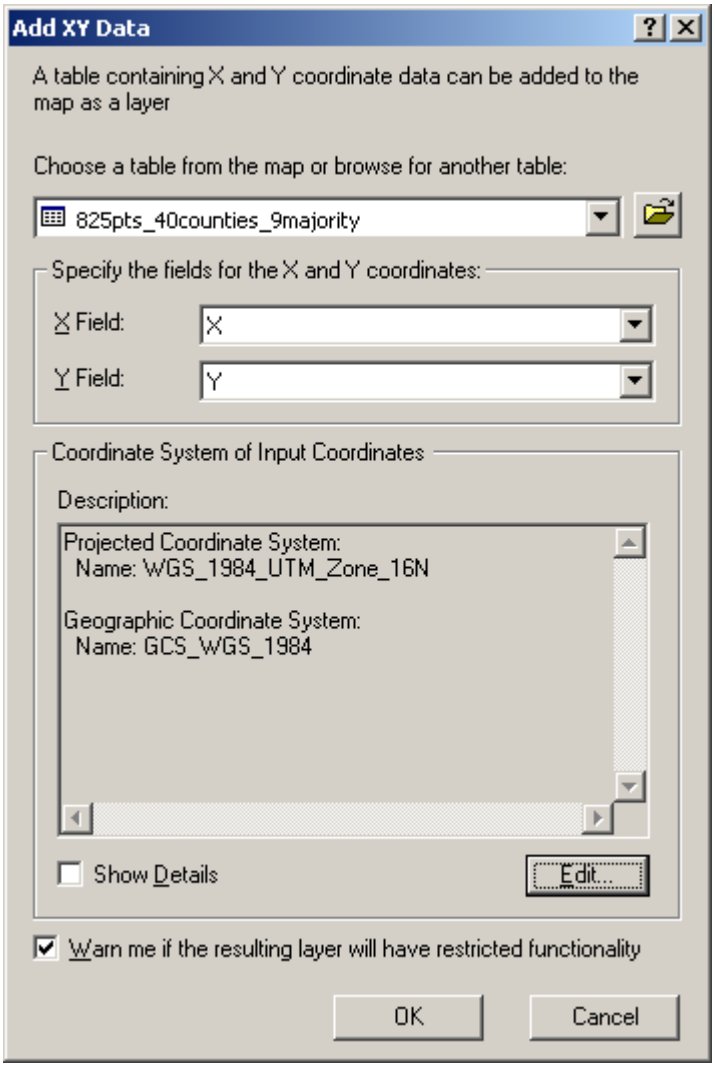


Point #	Name	X	Y	Class	Reference
1	ID#1	364169.000	4149497.000		
2	ID#2	211905.000	4447529.000		
3	ID#3	166713.000	4413537.000		
4	ID#4	296913.000	4161145.000		
5	ID#5	254857.000	4269113.000		
6	ID#6	391609.000	4644145.000		
7	ID#7	246401.000	4228177.000		
8	ID#8	165761.000	4374169.000		
9	ID#9	220977.000	4328417.000		
10	ID#10	299993.000	4198889.000		
11	ID#11	338353.000	4131521.000		
12	ID#12	411601.000	4665705.000		
13	ID#13	362097.000	4125361.000		
14	ID#14	419161.000	4596881.000		
15	ID#15	223385.000	4321137.000		
16	ID#16	347369.000	4273425.000		
17	ID#17	139833.000	4429777.000		
18	ID#18	230553.000	4481857.000		
19	ID#19	372065.000	4671921.000		
20	ID#20	336281.000	4264577.000		
21	ID#21	221145.000	4494233.000		
22	ID#22	328441.000	4197377.000		
23	ID#23	206193.000	4456489.000		
24	ID#24	370721.000	4153305.000		
25	ID#25	334377.000	4168313.000		

Then in ArcMap, make a shapefile out of the saved *.dbf file using the Add XY data feature under the Tools menu.

Add the saved *.dbf file (**825pt_40counties_3x3_majthreshold9.dbf**)

Select **Tools --- Add XY Data**



Then Export your 825_points events feature to a shapefile by right clicking on the feature then selecting **Data --- Export data**. (825pts_40counties_maj9.shp)

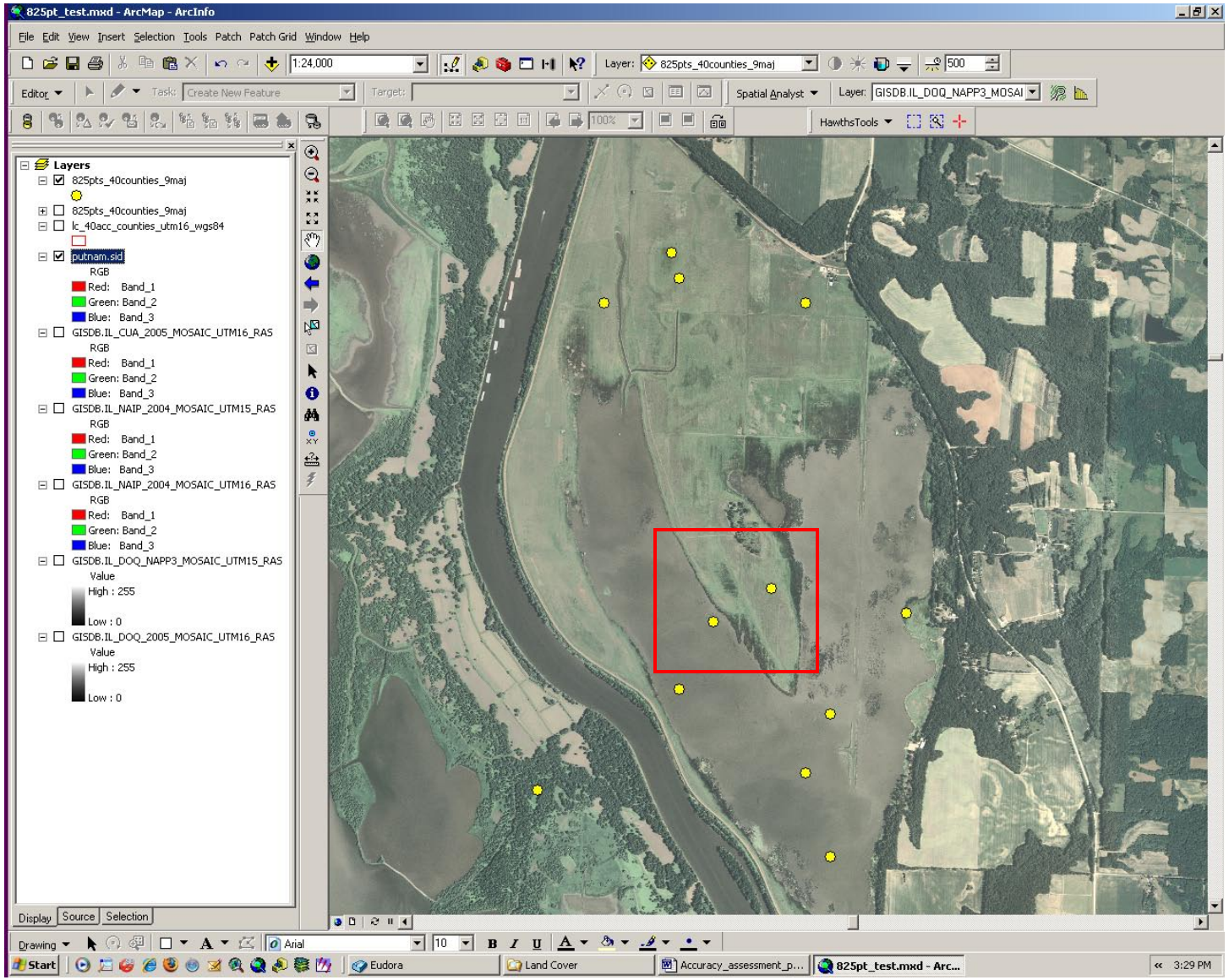
Then using the 2007 NAIP aerial photography and other ancillary data sources, fill in the reference column for each point with the appropriate land cover type code.

Attributes of 825pts_40counties_9maj

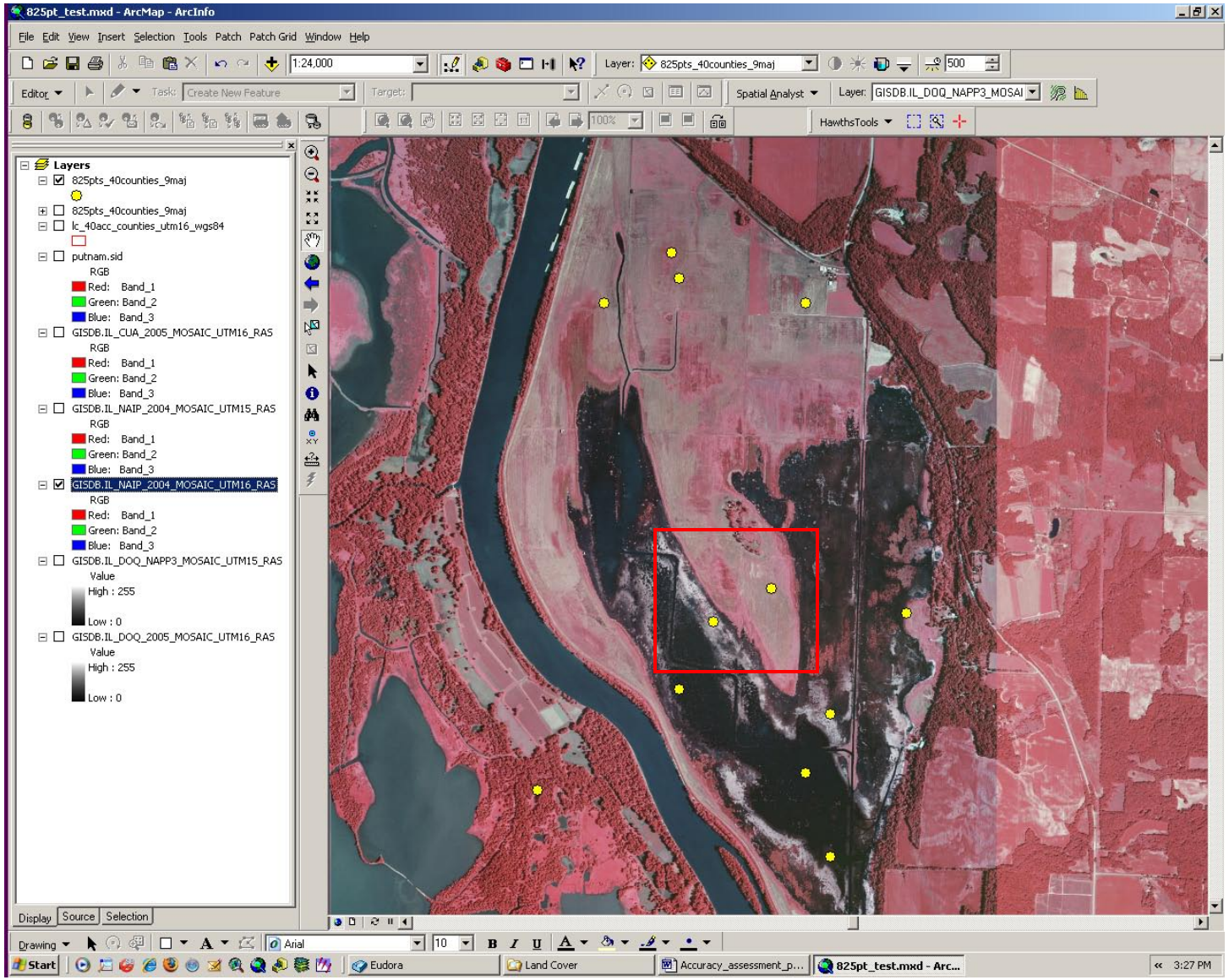
FID	Shape	ID	X	Y	REFERENCE	COMMENTS
0	Point	ID#1	345857	4150617		
1	Point	ID#2	279217	4170833		
2	Point	ID#3	310689	4133929		
3	Point	ID#4	356945	4207345		
4	Point	ID#5	376601	4155937		
5	Point	ID#6	346025	4290001		
6	Point	ID#7	392113	4626001		
7	Point	ID#8	212465	4243969		

Record: 0 Show: All Selected Records (0 out of 825 Selected) Options

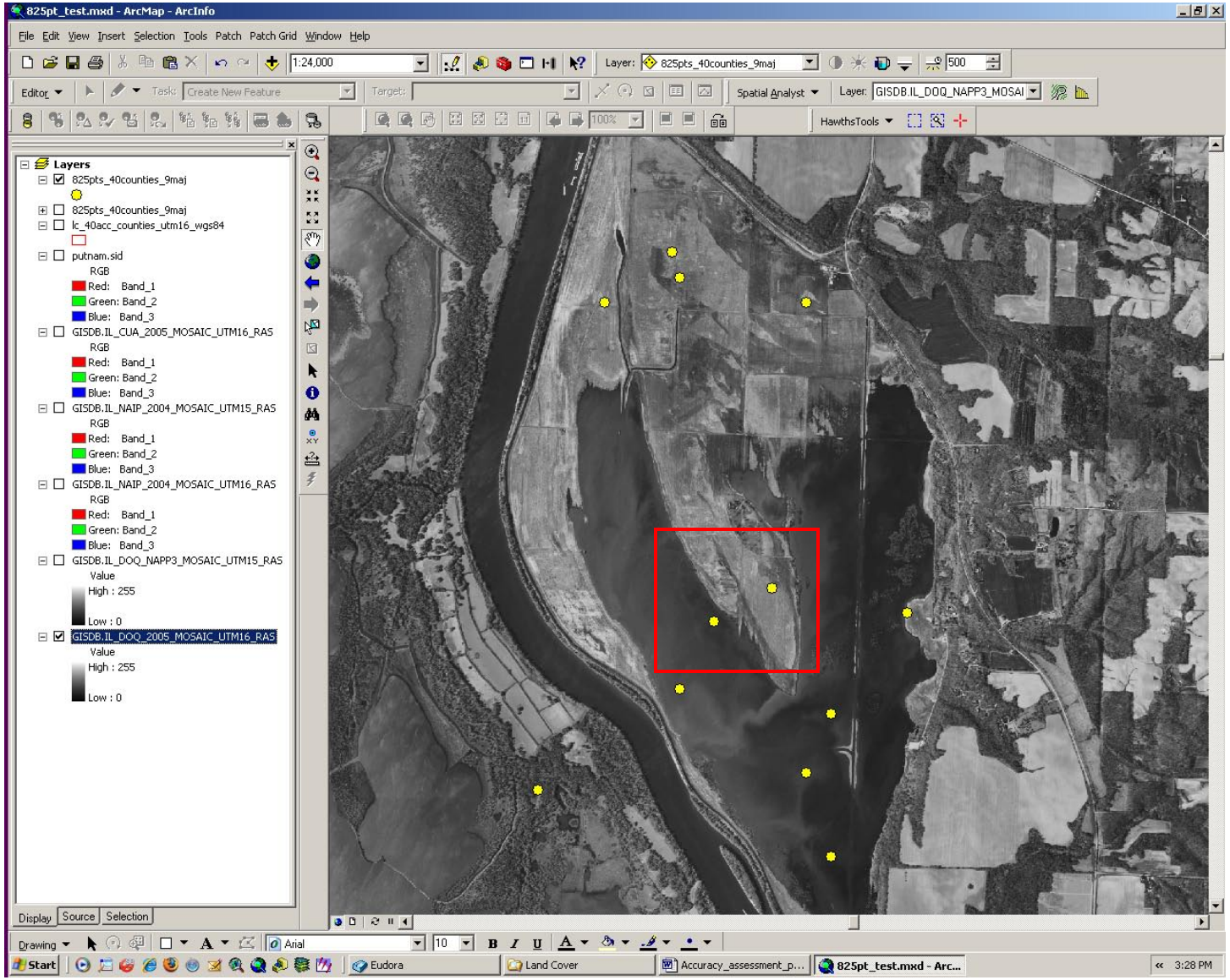
The 1 meter 2007 NAIP true color imagery is used as the primary source data layer when determining the reference categories



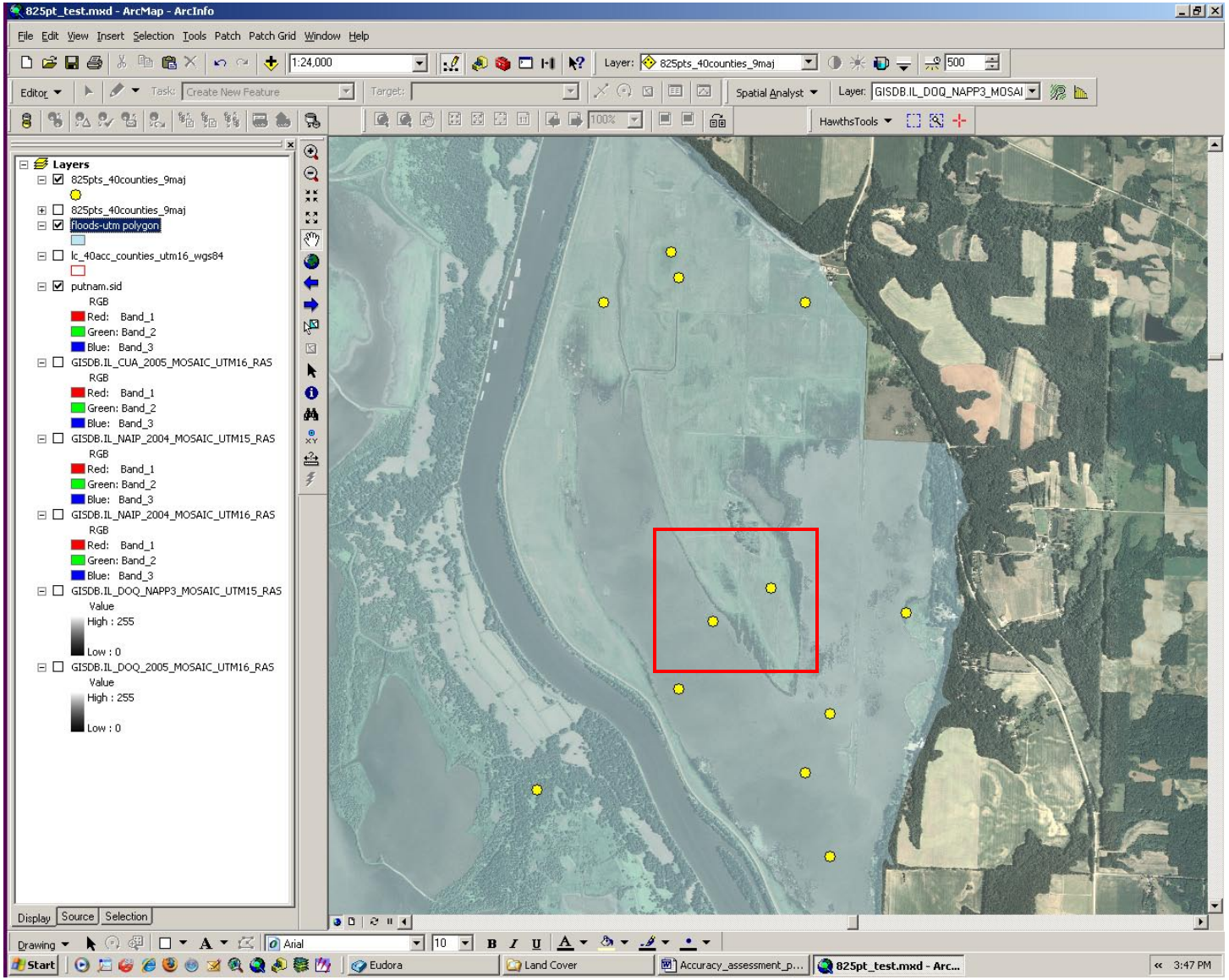
The 2 meter 2004 NAIP color infrared imagery will be as an ancillary data source to help identify the reference categories



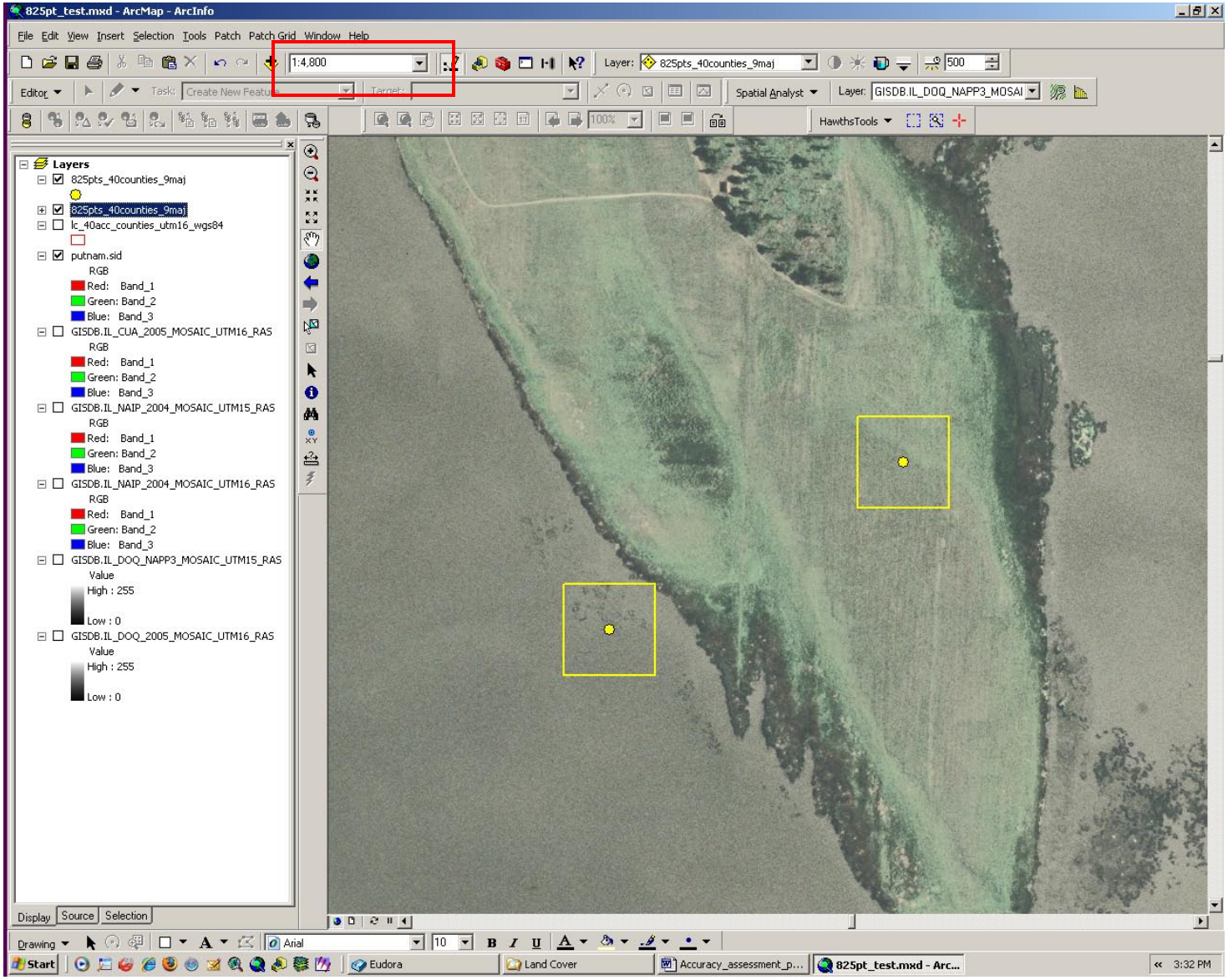
The 0.5 meter 2005 NAPP black and white imagery will be used as another ancillary data source to help identify the reference categories



The 100-year floodzones will be another ancillary data source to help identify the reference categories, especially herbaceous and wooded wetlands



A scale of 1:4,800 will be used to identify the final reference categories. A box representing a 3x3 cluster of Awifs pixels is shown below.



Appendix D

Non-Agricultural Land Contingency Matrix

Appendix D. Contingency Matrix for 2007 Illinois NASS CDL Non-Agricultural Land Cover

		Reference Data														
Classified Data	Non-Agricultural Land Cover Category		Open Water	Developed/ Open Space	Developed/ Low Intensity	Developed/ Medium Intensity	Developed/ High Intensity	Barren	Deciduous Forest	Evergreen Forest	Grassland Herbaceous	Woody Wetlands	Herbaceous Wetlands	Classified (Row) Total	User's Accuracy	Commission Error	Cond'l Kappa
		Category #	111	121	122	123	124	131	141 (63,143)	142	171 (62,152,181)	190	195 (87)				
	Open Water	111	**65**	0	0	0	0	1	0	0	1	0	5	72	90.3%	9.7%	0.89
	Developed/Open Space	121	0	**65**	2	1	1	3	0	0	1	3	0	76	85.5%	14.5%	0.84
	Developed/Low Intensity	122	0	3	**72**	1	0	0	0	0	0	0	0	76	94.7%	5.3%	0.94
	Developed/Medium Intensity	123	0	0	1	**70**	1	0	0	0	0	0	0	72	97.2%	2.8%	0.97
	Developed/High Intensity	124	0	0	0	2	**66**	3	0	0	0	0	0	71	93.0%	7.0%	0.92
	Barren	131	1	0	0	2	11	**56**	0	0	0	0	0	70	80.0%	20.0%	0.78
	Deciduous Forest	141 (63,143)	0	0	0	0	0	1	**78**	1	7	5	0	92	84.8%	15.2%	0.83
	Evergreen Forest	142	0	0	0	0	0	0	0	**70**	0	0	0	70	100.0%	0.0%	1.00
	Grassland Herbaceous	171 (62,152,181)	0	10	2	1	0	1	0	0	**65**	0	5	84	77.4%	22.6%	0.75
	Woody Wetlands	190	0	0	0	0	0	0	2	0	1	**66**	2	71	93.0%	7.0%	0.92
	Herbaceous Wetlands	195 (87)	7	0	0	0	0	0	0	0	0	6	**57**	70	81.4%	18.6%	0.80
	Reference (Column) Total		73	78	77	77	79	65	80	71	75	80	69	824			
	Producer's Accuracy		89.0%	83.3%	93.5%	90.9%	83.5%	86.2%	97.5%	98.6%	86.7%	82.5%	82.6%		730		
	Omission Error		11.0%	16.7%	6.5%	9.1%	16.5%	13.8%	2.5%	1.4%	13.3%	17.5%	17.4%			88.6%	
	Conditional Kappa		0.88	0.82	0.93	0.90	0.82	0.85	0.97	0.98	0.85	0.81	0.81				0.87

Non-Agricultural Land Cover Categories
Overall Accuracy Assessment

Number of Correct Reference Samples	730
Number of Total Reference Samples	824
Overall Percent Correct	88.6%
Overall Kappa Statistic	0.87

Appendix E

Agriculture and Non-Agricultural Land Contingency Matrix

FSA + NLCD Land Cover Categories
Overall Accuracy Assessment

Number of Correct Samples	4215336
Number of Total Samples	4317189
Overal Percent Correct	97.64%
Overall Kappa Statistic	0.9509

Reference Data

Classified Data: Individual Land Cover Category	Value	Corn 1	Rice 3	Sorghum 4	Soybeans 5	Sunflowers 6	Tobbaco 11	Barley 21	Spring Wheat 23	Winter Wheat 24	Other Grains 25	WW/ Soybeans 26	Rye 27	Oats 28	Alfalfa 36	Dry Beans 42	Potatoes 43	Other Crops 44	Misc. Vegetables 47	Peas 53	Clover/ Wildflowers 58	Idle/ Fallow 61	NLCD - Open Water 111	NLCD - Developed/ Open Space 121	NLCD - Developed/ Low Intensity 122	NLCD - Developed/ Medium Intensity 123	NLCD - Developed/ High Intensity 124	NLCD - Barren 131	NLCD - Deciduous Forest 141 (63,143)	NLCD - Evergreen Forest 142	NLCD - Grassland Herbaceous 171 (62,152,181)	NLCD - Woody Wetlands 190	NLCD - Herbaceous Wetlands 195	Classified (Row) Total	User's Accuracy	Comission Error	Conditional Kappa
Sunflowers	6	41	0	6	60	**3**	0	0	0	0	0	1	2	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	115	2.61%	97.39%	0.0261
Tobbaco	11	4	0	0	75	0	**0**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	79	0.00%	100.00%	0.0000
Barley	21	4	0	0	10	0	0	**17**	0	35	0	1	0	1	8	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	80	21.25%	78.75%	0.2125
Spring Wheat	23	0	0	0	0	0	0	0	**33**	6	0	0	0	25	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	65	50.77%	49.23%	0.5077
Winter Wheat	24	244	0	0	300	0	0	0	0	**25768**	0	5916	5	25	50	3	0	13	13	2	7	7	0	0	0	0	0	0	0	0	0	0	32353	79.65%	20.35%	0.7950
Other Grains	25	0	0	0	0	0	0	0	0	3	**0**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0.00%	100.00%	0.0000
WW / Soybeans	26	163	0	10	5310	0	0	0	0	3369	0	**77663**	0	1	8	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	86526	89.76%	10.24%	0.8954
Rye	27	0	0	0	0	0	0	0	0	83	0	1	**37**	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	122	30.33%	69.67%	0.3033
Oats	28	85	0	0	29	0	0	1	0	68	0	10	0	**238**	12	0	1	0	4	0	0	6	0	0	0	0	0	0	0	0	0	0	454	52.42%	47.58%	0.5242
Alfalfa	36	261	0	6	319	0	0	0	0	52	0	31	0	3	**3998**	1	1	9	13	9	20	2	0	0	0	0	0	0	0	0	0	0	4725	84.61%	15.39%	0.8459
Dry Beans	42	127	0	0	157	0	0	0	0	49	0	1	0	0	4	**634**	30	2	4	0	2	10	0	0	0	0	0	0	0	0	0	0	1020	62.16%	37.84%	0.6215
Potatoes	43	100	0	0	16	0	0	0	0	8	0	0	0	0	4	0	**145**	5	10	0	0	0	0	0	0	0	0	0	0	0	0	0	288	50.35%	49.65%	0.5034
Other Crops	44	14	0	0	32	0	0	0	0	65	0	2	0	2	1	0	0	**209**	5	0	0	6	0	0	0	0	0	0	0	0	0	0	336	62.20%	37.80%	0.6220
Misc. Vegetables	47	2808	0	19	433	1	0	0	0	20	0	24	0	6	6	6	9	0	**6076**	10	0	2	0	0	0	0	0	0	0	0	0	0	9420	64.50%	35.50%	0.6445
Peas	53	7	0	0	6	0	0	0	0	4	0	0	0	33	4	0	0	0	12	**67**	0	0	0	0	0	0	0	0	0	0	0	0	133	50.38%	49.62%	0.5037
Clover / Wildflowers	58	11	0	0	11	0	0	0	0	6	0	6	0	0	71	0	0	0	0	0	**175**	0	0	0	0	0	0	0	0	0	0	0	280	62.50%	37.50%	0.6250
Idle / Fallow	61	131	0	2	330	0	0	0	1	25	0	4	0	22	2	0	0	11	1	0	0	**225**	0	0	0	0	0	0	0	0	0	0	754	29.84%	70.16%	0.2983
NLCD - Open Water	111	126	2	1	98	0	0	0	0	4	0	4	0	0	0	0	0	0	0	0	0	5	**65**	0	0	0	0	0	1	0	0	5	312	20.83%	79.17%	0.2083
NLCD - Developed/Open Space	121	2099	0	4	1312	0	0	0	0	104	0	142	0	2	21	0	2	12	15	1	0	13	0	**65**	2	1	1	3	0	0	1	3	0	3803	1.71%	98.29%	0.0171
NLCD - Developed/Low Intensity	122	127	0	0	71	0	0	0	0	6	0	6	0	0	1	0	0	0	0	0	0	2	0	3	**72**	1	0	0	0	0	0	0	0	289	24.91%	75.09%	0.2491
NLCD - Developed/Medium Intensity	123	6	0	0	4	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	**70**	1	0	0	0	0	0	0	86	81.40%	18.60%	0.8140
NLCD - Developed/High Intensity	124	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	**66**	3	0	0	0	0	0	71	92.96%	7.04%	0.9296
NLCD - Barren	131	12	0	3	20	0	0	0	0	4	0	1	0	0	1	1	0	0	0	0	1	2	1	0	0	2	11	**56**	0	0	0	0	0	115	48.70%	51.30%	0.4869
NLCD - Deciduous Forest	141 (63,143)	542	0	3	456	0	0	0	0	34	1	58	0	0	12	0	0	12	2	0	0	27	0	0	0	0	0	1	**78**	1	7	5	0	1239	6.30%	93.70%	0.0629
NLCD - Evergreen Forest	142	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**70**	0	0	0	70	100.00%	0.00%	1.0000
NLCD - Grassland Herbaceous	171 (62,152,181)	13446	1	170	9928	6	1	5	0	1273	0	3014	4	69	1755	61	10	88	124	3	78	176	0	10	2	1	0	1	0	0	**65**	0	5	30296	0.21%	99.79%	0.0021
NLCD - Woody Wetlands	190	28	0	0	7	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	1	**66**	2	108	61.11%	38.89%	0.6111
NLCD - Herbaceous Wetlands	195 (87)	852	1	2	386	0	0	5	0	37	0	17	0	7	5	1	0	8	1	0	0	5	7	0	0	0	0	0	0	0	0	6	**57**	1397	4.08%	95.92%	0.0408
Reference (Column) Total		2809711	87	4991	1365234	11	4	28	34	31795	1	88625	50	511	6211	729	220	398	6795	97	290	543	73	78	77	77	79	65	80	71	75	80	69	**4317189**			
Producer's Accuracy		98.69%	88.51%	78.82%	96.86%	27.27%	0.00%	60.71%	97.06%	81.04%	0.00%	87.63%	74.00%	46.58%	64.37%	86.97%	65.91%	52.51%	89.42%	69.07%	60.34%	41.44%	89.04%	83.33%	93.51%	90.91%	83.54%	86.15%	97.50%	98.59%	86.67%	82.50%	82.61%		**4215336**		
Omission Error		1.31%	11.49%	21.18%	3.14%	72.73%	100.00%	39.29%	2.94%	18.96%	100.00%	12.37%	26.00%	53.42%	35.63%	13.03%	34.09%	47.49%	10.58%	30.93%	39.66%	58.56%	10.96%	16.67%	6.49%	9.09%	16.46%	13.85%	2.50%	1.41%	13.33%	17.50%	17.39%			**97.64%**	
Conditional Kappa		0.9629	0.8851	0.7878	0.9544	0.2727	0.0000	0.6071	0.9706	0.8090	0.0000	0.8738	0.7400	0.4657	0.6433	0.8697	0.6591	0.5251	0.8940	0.6907	0.6034	0.4143	0.8904	0.8332	0.9351	0.9091	0.8354	0.8615	0.9750	0.9859	0.8657	0.8250	0.8260				**0.9509**