FLORIDA 2004
CROPLAND DATA LAYER

Claire Boryan, Remote Sensing Analyst
USDA, National Agricultural Statistics Service
3251 Old Lee Highway, Suite 305, Fairfax, Virginia 22030
Claire_Boryan@nass.usda.gov

ABSTRACT

This paper provides an overview of the methodology and data used to produce the Florida 2004 Cropland Data Layer (CDL), a crop-specific classification that identifies major agricultural crops in the state. The National Agricultural Statistics Service (NASS) of the USDA has used satellite imagery since the 1970's to aid in the acreage estimation of large area crops in major producing states (Craig, 2001; Hanuschak *et al.* 1982). The Florida 2004 CDL served as an opportunity to evaluate the efficacy of new procedures not previously employed in the production of a CDL. Spring and fall image mosaics were created using TM and ETM+ SLC-off gap filled data. Gap filled images were used to supplement the TM data due to extensive cloud cover. Farm Service Agency (FSA) Common Land Unit (CLU) data and Florida Agricultural Statistics Service (FASS) GIS Citrus Grove Layer Data served as ground truth information. The classification was performed using a decision tree analysis methodology and See5 software[1]. Ancillary data including agriculture, date, and cloud masks were used in the classification process. The National Land Cover Data set (NLCD) was used to identify land cover types that were not agriculture. An error matrix was created by comparing the final classification with the FSA CLU and the FASS GIS Citrus Grove Layer data. The error matrix provided an assessment of the accuracy of specific crop categories or groups of categories. User's and producer's accuracies were derived. The overall agricultural accuracy was calculated as 85.6% with a Kappa of 0.746.

INTRODUCTION

NASS has used multispectral satellite imagery since the 1970's to aid in the acreage estimation of large area crops in major producing states. Early research and later operational NASS remote sensing programs used imagery from the Landsat multi-spectral scanner (MSS) instruments until the late 1980's, research on Landsat 5 TM started in 1986 and it was adopted for operational acreage estimation use in 1991. By April of 1999, Landsat 5 TM and Landsat 7 ETM+ data were used in combination to produce major crop estimates at the state and county levels plus digital Cropland Data Layer (CDL) GIS products for six major crop producing states (Allen *et al.* 2002; Craig, 2001; Hanusachak, *et al.* 1982; Mueller, 2000)

NASS' CDL contains crop-specific digital data layers, suitable for use in GIS applications. At the present time, this program annually produces CDL's of the following States: Arkansas, Illinois, Indiana, Iowa, Mississippi, Missouri Bootheel, Nebraska, North Dakota and Wisconsin, with the goal of producing digital categorized geo-referenced output products. This program represents a cooperative venture between two USDA agencies, NASS and the Production Estimates and Crop Assessment Division (PECAD) of the Foreign Agriculture Service (FAS). PECAD houses an image archive from which NASS remote sensing analysts select optimal imagery for their CDL classifications. Additionally, ventures between NASS field offices and their respective state government partners and universities have been forged to produce CDL products.

The Florida 2004 CDL was a one time project providing a crop-specific classification of the major crops in Florida. No crop acreage estimates were derived from the Florida 2004 CDL classification. The project served as an opportunity to test new procedures. Historically, June Agricultural Survey (JAS) data were used as ground truth information for the production of all previous CDLs but the JAS data for the State of Florida were too limited. Florida has 100 JAS segments (much less than major agriculture producing states) providing approximately 260 square miles of training data. However, only 25 square miles, approximately 10%, were made up of high intensity

[1] Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

agriculture and 62 square miles were specialty crops. The majority of the JAS training data were made up of pasture/rangeland & low intensity agriculture rather than specialty crops. To perform a quality assessment of Florida agriculture, other training data had to be investigated. The FSA CLU data provided more complete coverage of agriculture crops with the only drawback being a concentration on row crops. Consequently, the FSA CLU data in combination with the FASS GIS Citrus Grove Layer data (citrus only) were used as ground truth information.

Landsat 7 ETM+ SLC-off gap filled data were used to supplement the available Landsat 5 TM data. See5 decision tree software was used rather than NASS' in-house maximum likelihood software, Peditor. The See5 software allowed for use of ancillary information, including agriculture, cloud and date masks, as well as, imagery mosaics which had never been used in the production of a CDL. In addition, the NLCD produced by the Multi Resolution Land Characteristics Consortium (MRLC) was employed, post processing, for the identification of non-agricultural cover types.

METHODOLOGY

Imagery

The Landsat 5 TM Sensor was launched in 1984, and is still producing images. Landsat TM was the primary source for imagery in the production of the Florida 2004 CDL. On November 26, 2005, the back-up solar array drive on Landsat 5 began exhibiting unusual behavior. The solar array drive maintains the proper pointing angle between the solar array (solar panels) and the sun. After a month-long investigation in December 2005 and testing in January 2006, new operating procedures were developed that will allow Landsat 5 to continue operations.

On May 31, 2003 an instrument malfunction occurred onboard the Landsat 7 ETM+ sensor. The Scan Line Corrector (SLC), which compensates for the forward motion of the satellite, had failed. The Landsat 7 ETM+ sensor is still capable of acquiring useful image data with the SLC turned off, particularly within the central portion of any given scene. NASS' Spatial Analysis Research Section (SARS) stopped using Landsat 7 data operationally, after crop year 2003. However, for this one time project, Landsat ETM+ SLC-off gap filled data provided by the USGS EROS Data Center (EDC) was utilized to supplement the Landsat TM data due to the extensive cloud cover that was prevalent in the region during the growing season. The SLC-off gap filled data were created using a segmentation model approach in which the missing image pixels in a SLC-off image were replaced with data values derived from coincident spectral information. The technique uses a multi-scale segment model to determine "boundaries" of like landscape units identified in circa 2000 SLC-on imagery. The "boundaries" steer the interpolation of spectral data across the missing image pixels of SLC-off imagery (Maxwell S. et al., 2007). Landsat gap filled data acquired on 4/04/2004 were used for Path/Rows15/41, 15/42/ and 15/43. Gap filled data acquired on 05/04/2004 were used for Path/Rows 17/39 and 17/40. Gap filled data acquired on 04/16/2004 and 7/21/2004 were used for Path/Rows 19/39 and 19/40.

Imagery mosaics for spring and fall dates were created from the Landsat TM and SLC-off gap filled data. All bands of TM and ETM+ data were used as inputs other than the TM and ETM+ thermal (band 6) due to its coarse spatial resolution. The imagery was analyzed in 8 bit color depth. Optimal imagery was selected, taking into consideration the best possible dates for crop separation and lowest percentage of cloud cover. Imagery from the Landsat TM sensor is acquired on a 16 day repeat cycle. Imagery acquisition every 16 days is not optimal for separation of crop types under the best of conditions (low percentage cloud cover). Consequently, the SLC-off gap filled images were used in combination with the TM data. The selection of imagery with optimal dates for crop separation is always a goal but not always a reality. For the production of the Florida 2004 CDL, no data acquired during the month of August was suitable for use in the classification and an early May 5th image had to be added to the fall mosaic because a later date could not be found with a sufficiently low percentage of cloud cover.

This classification was performed in two parts. One classification was run of the northern panhandle region where a majority of the row crops, such as corn, cotton, soybeans and peanuts, are grown. A second classification was performed of the southern peninsula where the citrus and sugarcane, as well as, specialty crops are grown. This division of the state follows the mapping zone delineations determined by the MRLC for the creation of the NLCD. The MRLC derived these mapping zones based on patterns of climate, soils, vegetation, landform and spectral reflectance in order to improve the accuracy of the NLCD (Homer *et al*. 2004).

The Florida Panhandle was classified using seven dates of TM imagery which were merged to create spring and fall image mosaics. Imagery preparation was performed using Leica ERDAS Imagine 9.0. Image dates include April 3, 2004; April 15, 2004; April 16, 2004; April 17, 2004; and May 4, 2004 for the spring mosaic and May 5, 2004, July 21, 2004, July 22, 2004 and September 1, 2004 for the fall mosaic. The Landsat 5 TM imagery were manually registered prior to being merged into a mosaic. The Landsat 7 ETM+ gap filled data were orthorectified. See Figures 1 and 2.

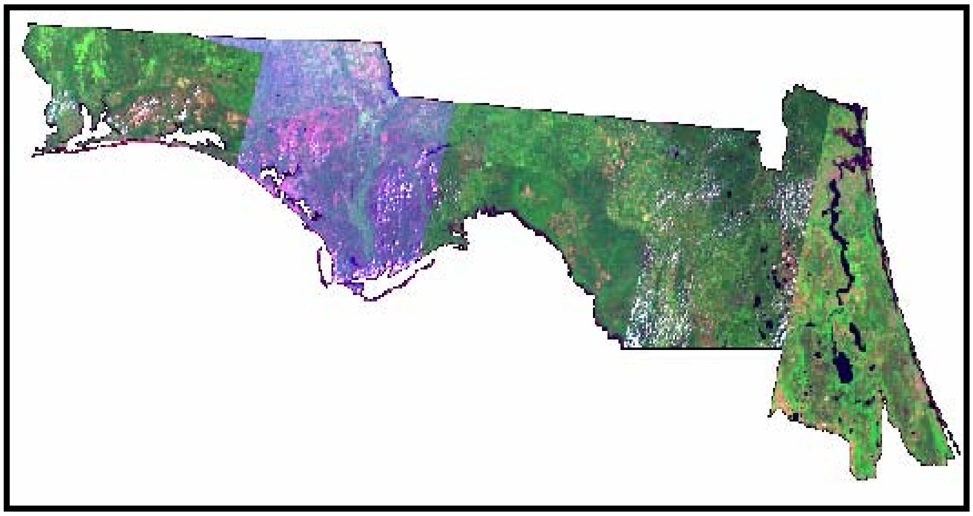


Figure 1. Spring Mosaic.

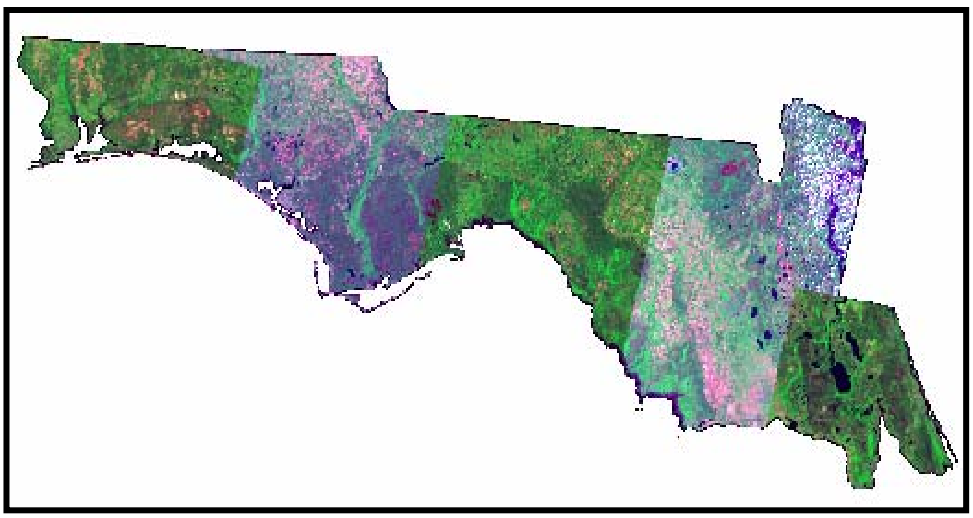
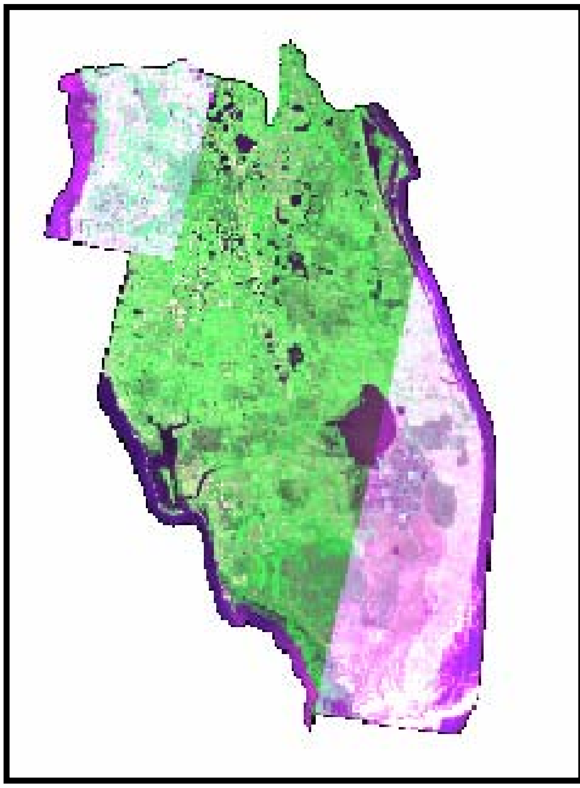


Figure 2. Fall Mosaic.

The Southern Peninsula region was classified using similar spring and fall imagery mosaics. Certain portions of the land cover were missing due to extreme amounts of cloud cover in the available imagery. It was determined that using single date imagery was preferable to including imagery with a very high percentage of cloud cover. Three dates of TM and Gap Filled imagery were merged to create the single spring image mosaic (Figure 3). These dates include April 3, 2004; April 4, 2004; and May 4, 2004. Four dates of TM imagery were merged together to create a single fall image mosaic (Figure 4). These fall dates include April 28, 2004; May 30, 2004; September 1, 2004; and November 29, 2004. The April and May dates were only used due to the high percentage of cloud cover that existed in later imagery acquisitions of this area.



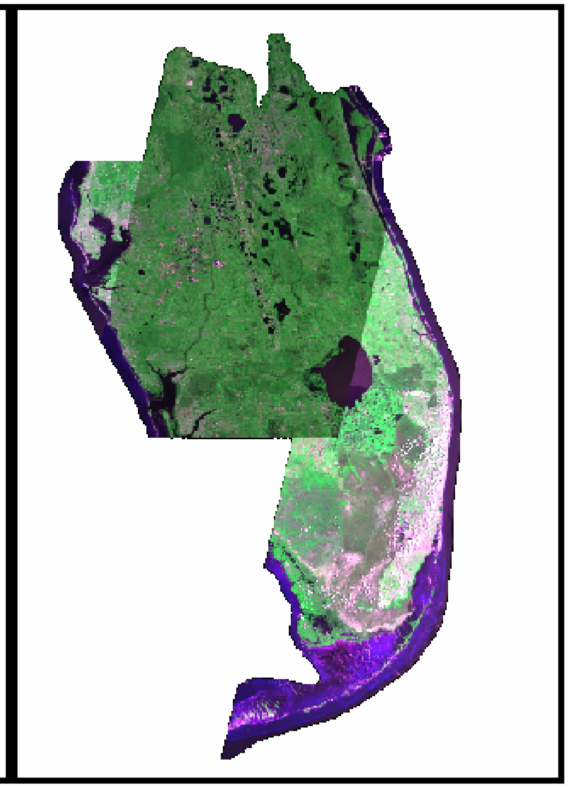


Figure 3. Spring Mosaic. **Figure 4.** Fall Mosaic.

Analysis District and Scene Dates Maps

Figures 5 and 6 illustrate the analysis districts and scene dates that were used in the Florida CDL. The Analysis District 1 (shades of green) is the Panhandle Region. Analysis District 2 (shades of orange) is the Peninsula. The partitioning of the state into two sections follows the mapping zone delineations determined by the MRLC for the creation of the NLCD 2001 (Homer, *et al.* 2004). The acquisition dates (unitemporal or multitemporal) of the Landsat imagery were included in the legend.

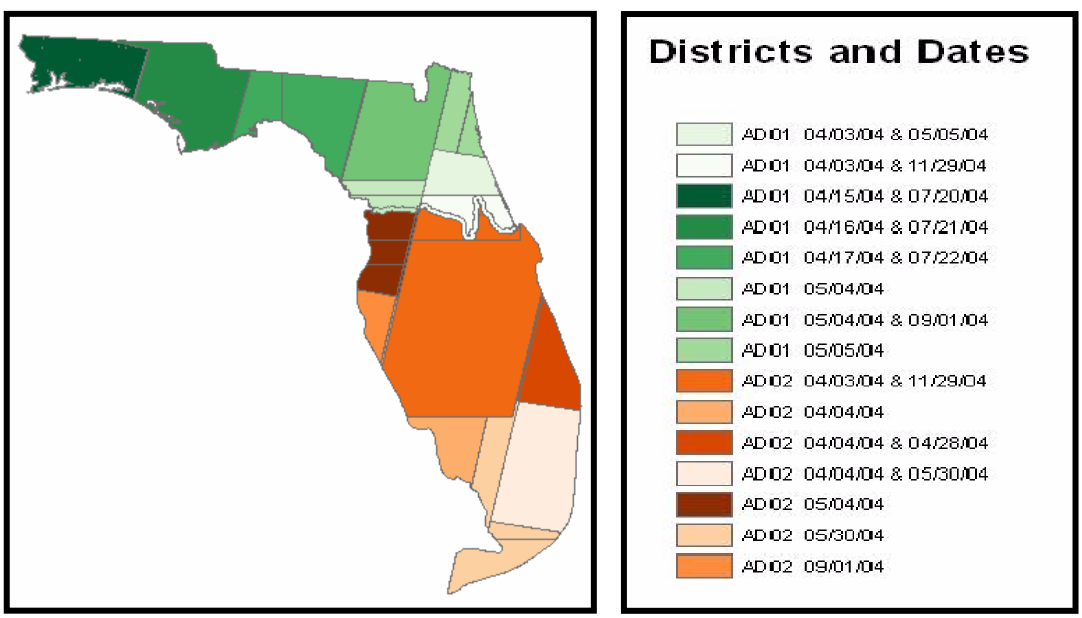


Figure 5 and Figure 6. Analysis Districts and Scene Dates.

Ground Truth Information

The USDA's Farm Service Agency (FSA) Common Land Unit (CLU) system was established to create a standardized GIS data layer of the nation's farms and fields. The CLU program supports farm commodity and conservation programs and disaster response. CLU data were updated every growing season when producers reported crops type and crop acreage for their fields to the FSA county office. The CLU system created digitized polygon boundaries of semi-permanent 'fields' in ESRI shape file format. Attribute information was maintained in a separate data base format known as FSA 578 Administrative Data. There was one CLU shape file per county. (Anderson, *et al*. 2005; Hearld, J. 2002)

To prepare the ground truth for the use in the production of the Florida 2004 CDL, the individual county CLU shape files were merged together to form a single state wide CLU shape file. One problem with the FSA data being organized by county was that a farmer could report, in a single county office, for all crop land he operated in the state, irrespective of the county where the fields were physically located. Consequently, when the CLU data were merged with the FSA data it was by "County of Administration" rather than county of geography. The CLU boundaries were merged with the FSA 578 data, using ESRI's ArcGIS 9.1, to acquire the crop type information. Only CLU polygons with one crop type were selected out for use as ground truth. Many CLUs were made up of multiple crop types (mixed fields) but these CLUs were not appropriate for training. Additionally the FSA CLU data did not provide information on all crop types. Only farm service program crop information was collected. This tends to exclude specialty crops. Crops for which there was an abundance of CLU data included Grass and Conservation Reserve Program (CRP). Some areas such as urban, water bodies and large non-agricultural zones were not digitized, and existed only as holes in the CLU shape file(s). After a full state CLU shape file was created and linked to the 578 data, each field was buffered in one pixel and rasterized (30 meter cell) using ESRI's ArcMap 9.1 for use with the See5 software.

Separate ground truth information was acquired specifically for the citrus crop. The FASS Citrus GIS Grove Layer was developed through a collaborative effort between SARS and FASS. The FASS GIS Citrus Grove layer was a major undertaking in which all of the citrus groves in the State of Florida were digitized and linked to attribute information. A personal geodatabase containing a consistent set of spatial and tabular datasets was established for every citrus producing county in Florida. Previously, this information was maintained on paper using cartographic techniques in which over 1200 maps were manually updated every other year. (Johnson, 2006).

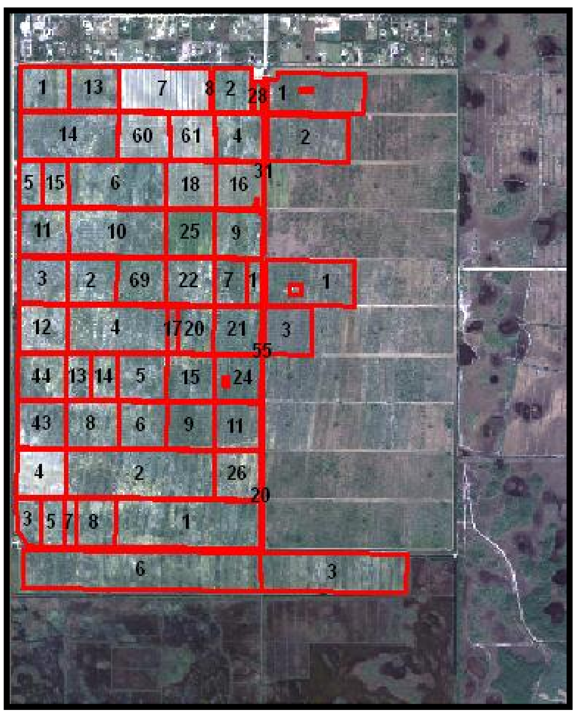


Figure 7. Florida GIS Citrus Grove Layer Boundaries
To prepare this information for used as ground truth, all county GIS Citrus Grove Layer shape files were merged into an individual state wide shape file. This shape file was then buffered (30 meters/1 pixel) in one pixel to avoid edge pixels, rasterized to a 30 meter cell size and merged with the CLU data. The GIS Citrus Grove Layer data were used only for the classification of the Peninsula. The CLU data were used alone for the classification of the Panhandle. It should be noted that the FSA CLU and Citrus Grove Layer data are only available to cooperating agencies and not available to the public.

See5 Decision Tree Software

A decision tree analysis methodology was used to produce the Florida 2004 Cropland Data Layer. This was the first opportunity to produce a full state CDL using a commercial decision tree software known as See5. Historically an in-house maximum likelihood classification software, Peditor, was used to produce NASS' state wide CDLs (Ozga, and Craig, 1995; Day, 2002). Decision tree or classification tree methodologies have outperformed maximum likelihood classifiers in the literature (Friedl and Brodley, 1997; Hansen et al., 1996; Lawrence et al., 2004). Decision trees have become increasingly popular for a variety of reasons. Decision trees are nonparametric and make no assumption regarding the distribution of the input data. They can accommodate large data volumes, are repeatable, efficient, and are tolerant of outliers. Additionally a wide assortment of data sources can be used with the decision tree software including raw imagery, derived spectral information including vegetation indices, digital elevation information, slope information, and a variety of GIS data layers (Friedl and Brodley, 1997; Lawrence et al., 2004). Additionally the See5 Boosting option set to 10 was employed. Boosting has been found to improve classification accuracies by using classification errors to help refine the decision tree (Lawrence *et al*., 2004; Quinlan, 1996). A perceived drawback of the software was the inability to edit and improve the classification. Other than locating additional training for a crop that was classified poorly, little could be done to improve the classification. Classification results were post processed using a minimum mapping unit (MMU) filter to eliminate individual misclassified pixels. The Imagine NLCD "Smart Eliminate" tool was applied with a MMU of 20 acres (90 TM and ETM+ pixels).

Figure 8. Agriculture Mask.

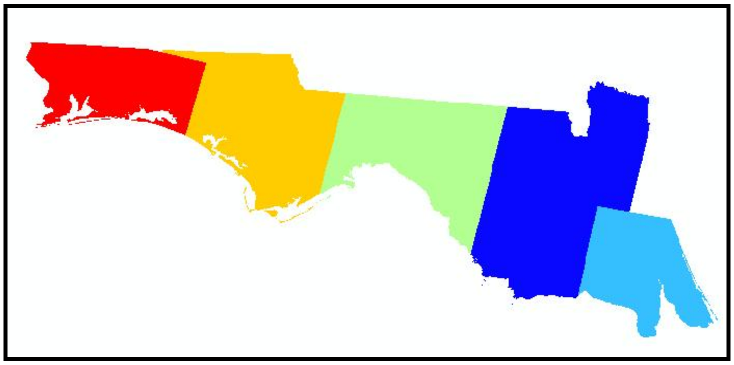


Figure 9. Date Mask.

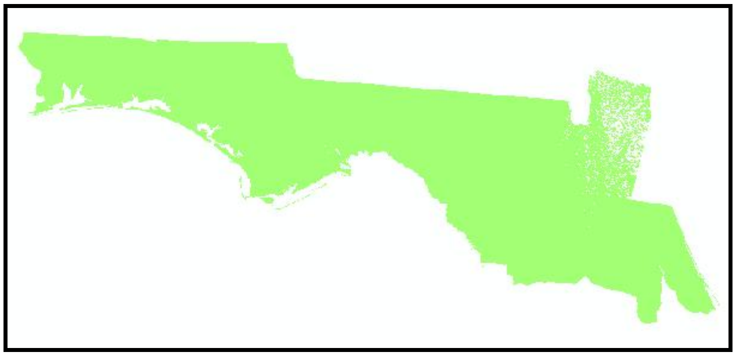


Figure 10. Cloud Mask.

Ancillary Data

Numerous types of ancillary data were used in the classification process. Preparation of ancillary data was conducted using Leica's ERDAS Imagine 9.0 software. Agricultural masks were created to focus the area to be classified and to reduce the opportunity for mixing with non-agricultural categories. The agriculture masks were created using the NLCD "Cultivated Crop" and Pasture/Hay" Categories (Homer *et al.* 2004). Date masks were created for use with multi scene mosaics. Individual date masks were created to match with the dates of imagery selected to build the spring and fall mosaics. Cloud masks were created to eliminate clouded areas from consideration in the classification. Figures 8-11 illustrate the types of ancillary data used in the production of the Florida 2004 CDL. The National Land Cover Data set (NLCD) 2001 was used to identify land cover types that were not agriculture (Homer *et al.* 2004). The NLCD was not used as part of the classification process, except for preparation of the agricultural mask. Post classification, it was merged with the agricultural classification to provide information regarding non agricultural categories.

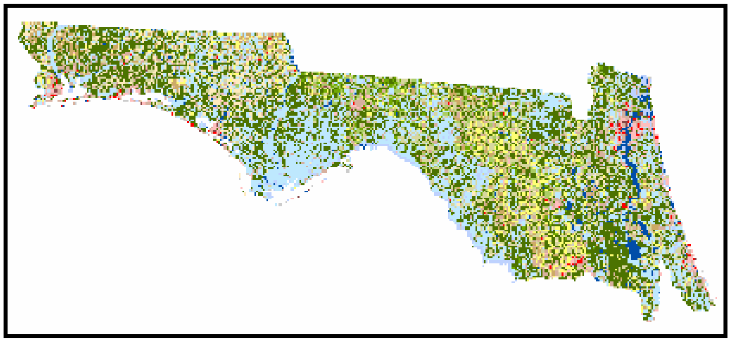
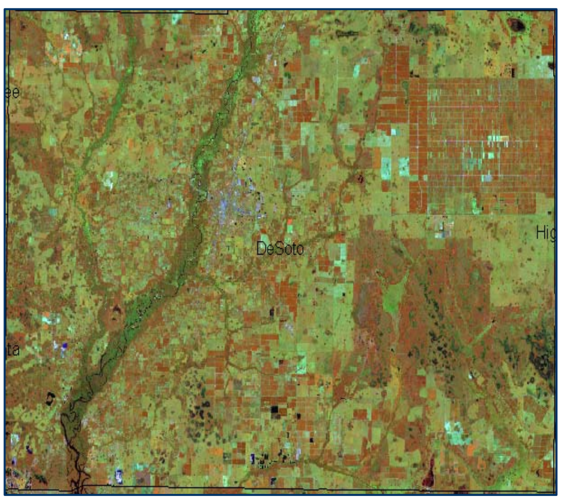


Figure 11. National Land Cover Data (NLCD) Set.

Accuracy Assessments

After the classification was performed, the accuracy of the CDL was assessed both visually and quantitatively. Figures 12 and 13 provide examples, at the county level, of visual comparisons between the raw imagery and the Florida CDL. The quantitative accuracy assessment of the Florida CDL was provided using an error matrix which was created using Imagine and Excel software. The error matrix was created by intersecting the final full state classification with the FSA CLU and the FASS GIS Grove Layer Data. The error matrix provides an assessment of the accuracy of specific crop categories or groups of categories



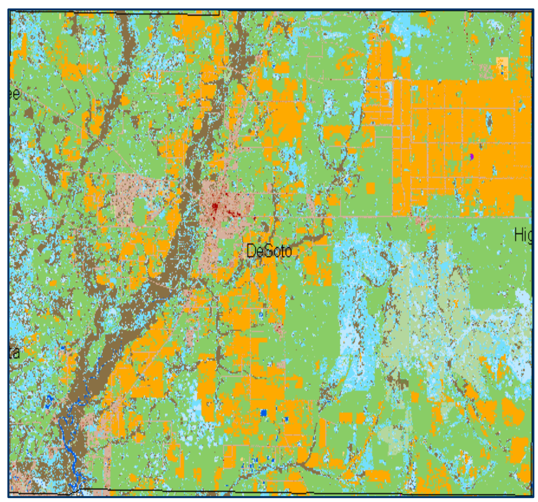


Figure 12 and Figure 13. Zooms of the raw Landsat TM imagery (Bands 4, 5, 3) and the Florida 2004 CDL of DeSoto county.

Table 1 identifies the user's accuracies and Table 2 identifies the producer's accuracies for the Florida 2004 CDL. The user's accuracy indicates the probability that a pixel from the CDL classification actually matches the ground truth data and measures the errors of commission. The producer's accuracy relates to the probability that a ground truth pixel (FSA CLU data or Grove Layer data) will be correctly mapped and measures the errors of omission (Congalton and Green, 1999). Table 3 is the error matrix. User's accuracies are identified along the far right column. Producer's accuracies were identified along the bottom row. Pixels in the Florida CDL that are classified correctly when compared with the FSA CLU and the FASS GIS Grove Layer ground truth information were identified along the diagonal and highlighted in yellow.

The user's and producer's accuracies for a majority of crops were very strong. The category of "Other Crop" did not perform well because it was a mixed category. The user's accuracy for "Sorghum" was much higher (98.74%) than was the producer's accuracy (45.58%) indicating that a number of fields were missed but those that were classified as "Sorghum" were generally correct. A good percentage of the Sorghum in the ground truth was classified to "Other Crop" in the CDL. The "Grass/Pasture" category had a lower user's accuracy (60.50%) than producer's accuracy (97.45%) indicating the existence of a large number of errors of commission in this category probably because the "Grass" was the dominant crop type in the CLU data. The training data for "Grass" seems to have overwhelmed several of the other crops. The "Citrus" category had a strong user's accuracy (99.97%) indicating that an area identified as citrus in the CDL was almost 100% of the time citrus on the ground but had a weaker producer's accuracy (80.89%) indicated that about 20% of the citrus crop which actually exists on the ground was missed. After review of the error matrix it is evident that a good portion of "Citrus" in the ground truth was classified as "Grass/Pasture" or "Other Crop". It is expected that citrus groves that were omitted from the classification include those that were relatively newly planted (sand or soil would dominate the signature) or abandoned (reflectance might be different than a grove in production).

The overall agricultural accuracy of the Florida 2004 CDL was calculated as 85.6%. The Kappa statistic: was calculated as 0.746. Accuracies of the NLCD data are not included, but are available on the MRLC web site. Figure 14 is the final Florida 2004 CDL classification with National Land Cover Data (NLCD) non-agricultural categories included.

Table 1. User's Accuracies

Corn	**83.42%**
Cotton (Upland)	**94.12%**
Sorghum	**98.74%**
Soybeans	**98.29%**
Peanuts:	**96.43%**
WinterWheat/Rye/Oats/Millet	**93.69%**
Other Crop	**11.89%**
Watermelon/Other Fruit	**92.86%**
Grass/Pasture	**60.50%**
Sugarcane:	**90.74%**
Citrus	**99.97%**

Table 2. Producer's Accuracies

Corn	**93.99%**
Cotton (Upland)	**97.98%**
Sorghum	**45.58%**
Soybeans	**97.62%**
Peanuts:	**97.99%**
WinterWheat/Rye/Oats/Millet	**91.47%**
Other Crop	**22.27%**
Watermelon/Other Fruit	**75.24%**
Grass/Pasture	**97.45%**
Sugarcane:	**99.24%**
Citrus	**80.89%**

Table 3. Error Matrix

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	Reference													
	Corn	Cotton	Sorghum	Soybeans	Peanuts	WW/Rye/	Other Cı	Sugarcan	Waterm	Grass/Pa	Citrus	Totals	User Accuracies	
Corn	41386	21	0	1	14	71	6	0	98	7727	275	49599	83.441	
Cotton(Upla	73	96224	0	97	453	598	47	0	13	4733	0	102238	94.118	
Sorghum	12	0	2592	0	0	0	0	0	0	21	0	2625	98.743	
Soybeans	33	35	0	17624	39	50	6	0	0	155	0	17942	98.228	
Peanuts	155	966	1	27	148665	201	135	0	357	3541	123	154171	96.429	
WW/Rye/Oa	69	150	9	9	135	69652	3277	0	0	1052	0	74342	93.691	
Other Crop_	1473	0	2618	0	0	0	4313	1	0	1306	26575	36286	11.886	
Sugarcane	134	0	0	0	0	0	1474	56013	0	0	4108	61729	90.74	
Watermelor	0	3	0	0	46	50	382	0	6812	43	0	7336	92.857	
Grass/Pastu	696	810	116	296	2356	5528	9713	428	1771	729996	454562	1E+06	60.517	
Citrus	0	0	0	0	0	0	12	0	3	525	2055068	2E+06	99.974	
Sum Classif	44031	98209	5336	18054	151708	76150	19365	56442	9054	749099	2540711	4E+06		
Accuracies	93.993	97.979	48.5757	97.618256	97.9942	91.46684	22.2721	99.23993	0.75237	97.4499	80.8855		0.8567	
	Producer Accuracies											Agriculture Accuracy		
												(NLCD) not included		

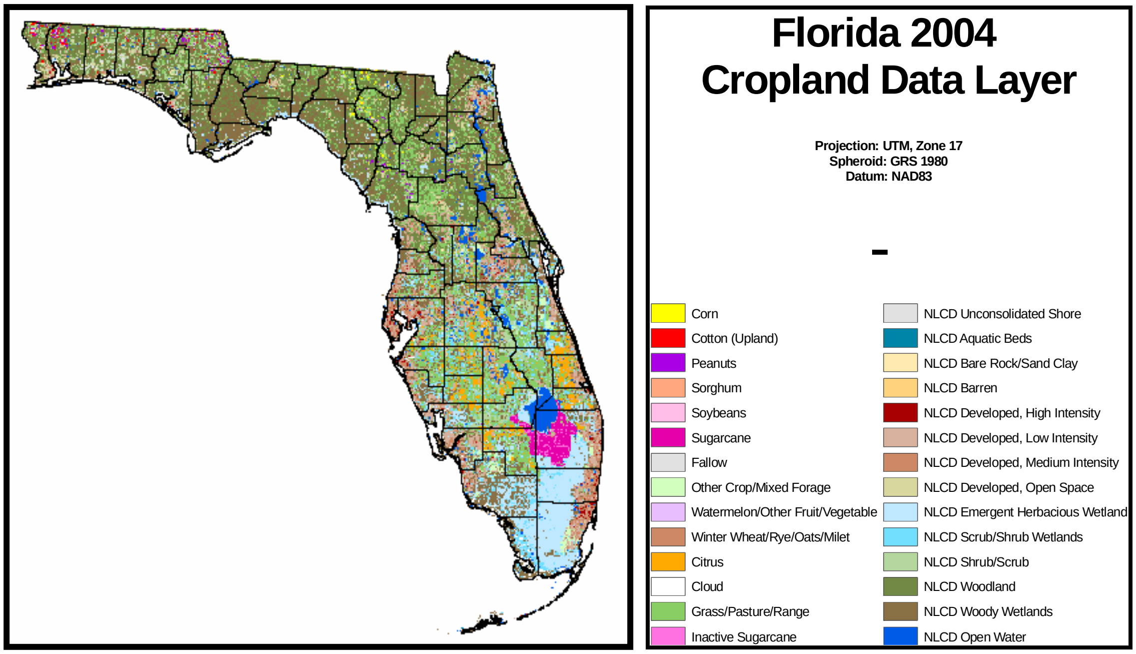


Figure 14. Final Florida 2004 CDL classification with National Land Cover Data (NLCD) non-agricultural categories included.

CONCLUSION

The production of the Florida 2004 CDL was an excellent opportunity to test new procedures and new attribute data including; Landsat ETM+ SLC-off gap filled data, agriculture masks, cloud masks, date masks, multi-date image mosaics, FSA CLU data and FASS GIS Grove Layer data for ground truth information; and See5 Decision Tree Software. Many of these procedures and data hold promise for use in future state-wide CDLs. However, some of the techniques attempted were not satisfactory and will not be utilized in the future.

Florida was a difficult state to classify crops using multispectral imagery, due to the high percentage of cloud cover that was prevalent during the growing season. Sufficient cloud free TM data were not available during the 2004 growing season to run a classification using TM data alone. For this reason, ETM+ SLC-off gap filled data were utilized in the classification. The segment based approach used to fill missing image pixels in the SLC-off imagery with coincident spectral data was relatively successful for the identification of crop fields. Errors likely occurred where field boundaries changed since 2000 (the date of SLC-on imagery) and where clouds existed in the SLC-on imagery. The technique was most successful for crops like citrus or sugarcane which did not change appreciably since 2000 (Maxwell *et al*. 2007). It should be mentioned, that the Florida 20 04 CDL would not have been able to be produced without the SLC-off gap filled imagery. Although there were errors incurred by the SLC-off gap filled imagery, it was considered the only option to produce the classification.

The masks that were created for this CDL will not be used in the future. It has been determined that adding the NLCD 2001 (non agriculture information) post classification is more appropriate than using the agriculture mask during the classification process. Date masks will no longer be utilized because single date images or single date mosaics are preferable to using multi-date mosaics. Cloud masks will only be created when absolutely necessary. After significant testing with the See5 software, it has been determined that if more dates of cloud fee data exist for an area than clouded data then that the software will ignore the cloud (no data) information.

In the future, the FSA CLU data will be an excellent form of ground truth for acreage estimation in major agricultural producing states in the Midwest and the Mississippi Delta region. Additionally, even if JAS data were used as ground truth, the FSA CLU data can serve as an independent validation data set. However, the FSA CLU data were not optimal to derive ground truth for the specialty crops in Florida. The FSA CLU data were dominated by information on row crops rather than specialty crops that are common in Florida. Fortunately, NASS had access to the FASS GIS Grove Layer which served as an excellent training set for the 700,000 acres of citrus in the state.

The See5 Decision Tree software holds significant promise for acreage estimation of large area crops in major producing states and production of the CDL. The decision tree software is non parametric and makes no assumption regarding the distribution of the input data. It can accommodate large data volumes, be easily repeated, and is efficient. Additionally, the See5 software provides the option for "boosting" and has the ability to ingest a wide assortment of GIS data layers. The testing of See5 software and the use of the FSA CLU data were the most valuable lessons learned in the production of the Florida 2004 CDL and will be used in future state wide CDLs produced by NASS. The Florida 2004 CDL is available for sale on DVD on the NASS website at http://www.nass.usda.gov/research/Cropland/SARS1a.htm.

ACKNOWLEDGEMENTS

Special thanks to Susan Maxwell of the USGS EROS Data Center for supplying SLC-off gap filled imagery and for information regarding the methodology used to produce it. Thanks also to Rick Mueller for his valuable comments on the manuscript. And further thanks to Dave Johnson for sharing his expertise in the utilization of See5 Decision Tree software.

REFERENCES

Allen R., Hanuschak, G., Craig. (2002), History of Remote Sensing for Crop Acreage in USDA's National Agricultural Statistics Service. < http://www.usda.gov/nass/nassinfo/remotehistory.htm>. Accessed 2007 13 Feb.

Anderson T, Payne T, Clarke K, and C. Ernst (2005). USDA Service Center Agencies Geospatial Data Management Team Data Management Plan for Common Land Unit Data February 2005. < http://www.itc.nrcs.usda.gov/scdm/docs/DMP-CLU-DataManagementPlan.pdf >. Accessed 2007 13 Feb.

Congalton, R.G. and K. Green (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Boca Raton, FL: CRS Press, Inc.

Craig, M.E. (2001). The NASS Cropland Data Layer Program. Proceedings of the Third International Conference on Geospatial Information in Agriculture and Forestry, Denver Colorado, November 5-7, 2001, CD-ROM.

Day, C. (2002), A Compilation of PEDITOR Estimation formulas, RDD Research Paper RDD-02-03, USDA/NASS, National Agricultural Statistics Service, Washington, D.C. Jan. 2002.

Friedl, M., and C. Brodley (1997). Decision Tree of Land Cover from Remotely Sensed Data. *Remote Sensing of the Environment*, 61: 339-409.

Hansen, M., Dubayah, R., & R. Defries (1996). Classification Trees: An Alternative to Traditional Landcover Classifiers. *International Journal of Remote Sensing*, 17, 1075-1081.

Hanuschak, G., Allen, R., and W. Wigton (1982). Integration of Landsat Data into the Crop Estimation Program of USDA's Statistical Reporting Service. In: *Proceedings of the 1982 Machine Processing of Remotely Sensed Data Symposium*, Eighth International Symposium, LARS, Purdue University, West Lafayette, IN-USA, pp. 45-56.

Hearld, J. (2002), ESRI ArcUser Online. April-June 2002. USDA Establishes a Common Land Unit, <http://www.esri.com/news/arcuser/0402/usda.html> Accessed 2007 13 Feb.

Homer C, Huang C, Yang L, Wylie, B. and M. Coan (2004) Development of a 2001 National Land-cover Database for the United States. *Photogrammetric Engineering and Remote Sensing* 70:829-840

Johnson, D. (2004). Modeling the Florida commercial Citrus Grove Infrastructure within a GIS. <http://gis.esri.com/library/userconf/proc06/papers/papers/pap_1891.pdf>. Accessed 2007 16 Feb.

Lawrence, R., Bunn, A., Powell, S and M. Zambon. (2004) Classification of Remotely Sensed Imagery using Stochastic Gradient Boosting as a Refinement of Classification Tree Analysis, *Remote Sensing of the Environment*, 90, 331-336.

Maxwell, S.K., Schmidt, G. L., and J. C. Storey (2007) A Multi-scale Segmentation Approach to Filling Gaps in Landsat ETM+ SLC-off images, *International Journal of Remote Sensing*, In Press.

Mueller, R. (2000). Categorized Mosaicked Imagery from the National Agricultural Statistics Service Crop Acreage Estimation Program. Available on the CD: *Proceedings of the ASPRS 2000 Conference,* ASPRS, Bethesda, MD. May 2000.

Ozga, M., and Craig, M (1995). PEDITOR-Statistical Image Analysis for Agriculture. Presentation at the Washington Statistical Society (WSS) Seminar, April, 1995, USDA, National Agricultural Statistics Service, Washington, DC-USA.

Quinlan, J.R. (1996). Bagging, Boosting, and C4.5. *In Proceedings AAAI-96 Fourteenth National Conference on Artificial Intelligence*, Portland OR, 1996.