CATEGORIZED MOSAICKED IMAGERY FROM THE NATIONAL AGRICULTURAL STATISTICS SERVICE CROP ACREAGE ESTIMATION PROGRAM

Rick Mueller
Remote Sensing Analyst
USDA National Agricultural Statistics Service
Spatial Analysis Research Section
3251 Old Lee Hwy, Suite 305
Fairfax VA 22030
rmueller@nass.usda.gov

ABSTRACT
The Acreage Estimation Program (AEP) at the National Agricultural Statistics Service (NASS) has researched acreage estimation using remote sensing since the mid 1970's. The AEP traditionally produces acreage statistics on a county and state level for major crops in several highly intensive agricultural states. The AEP has gone through a variety of phase changes since inception, ranging from research to production mode based on Agency funding and sensor related issues. Currently, the AEP is in production mode encompassing the States of Arkansas, Illinois, Mississippi, New Mexico, and North Dakota, with the focus this year of producing digital categorized geo-referenced output products. The Agency developed image processing package PEDITOR, was utilized to perform the categorization, and the images were exported to commercial GIS output formats. Experimental digital geo-referenced output products have been developed in Arkansas, and North and South Dakota for 1997 - 1998 that contain near statewide coverage of categorized Landsat 5 TM imagery. The individual scenes were mosaicked and distributed via CD-ROM, along with ESRI’s ArcExplorer GIS browser. Currently, adjacent scene and year to year registration methods are under investigation to accurately co-register scenes for the 1999 crop year.

INTRODUCTION
NASS has done extensive research integrating ground survey data with satellite data into the crop AEP (Allen and Hanuschak 1988, and Ozga and Craig 1995). The methodology of making accurate acreage estimates from the Landsat sensor used a sampled regression approach to correct for the inaccuracies of the categorized images. By combining both satellite imagery and ground survey data (i.e., June Agricultural Survey), NASS was able to produce better crop acreage estimates, rather than relying solely on using the area expansion derived from the ground data (Allen 1990). This methodology has not changed since the program’s inception. The real gains in the program were the increased efficiencies associated with faster and cheaper hardware, and continuous enhancements to the programs written for PEDITOR (Ozga 2000).

Since 1997, the AEP encompassed Arkansas, and North and South Dakota. It became necessary to obtain external partnerships to support a program of this magnitude. Creating a data sharing partnership in 1997 with the Foreign Agricultural Service and the Farm Service Agency (FAS/FSA) with Landsat 5 TM imagery was the first step for program expansion. The FAS/FSA possess a large archive of satellite imagery covering major crop producing areas around the world, and the Production Estimates and Crop Assessment Division of the FAS was one of the longest running operational users of commercial satellite imagery (Bethel 1998). The FAS/FSA/NASS partnership was a natural fit for expanding the AEP. For the 1999 crop year, three new states were added, Illinois, Mississippi, and New Mexico. Staffing partnerships were added with the State Governments of Illinois and Mississippi, and the State Conservationist of the Natural Resources Conservation Service in New Mexico. Traditionally, the ground truth processing was performed in NASS’s State Statistical Offices, and transferred to the Spatial Analysis Research Section (SARS) in Fairfax, VA for the analysis phase of the project. This was the first time dedicated personnel working in the State Statistical Office’s would handle the entire project. The goal of these partnerships was to produce geo-referenced categorized images of their state for crop year 1999, while producing a geo-referenced categorized mosaic and adding
an acreage estimate for crop year 2000. The SARS will produce acreage estimates and the categorized mosaic for crop year 1999, however, these processes will be moved to the State Statistical Offices for crop year 2000.

Since the inception of the AEP, nearly all of the processing and management functions were performed on site in the SARS. However, the ground truth data collection and field digitizing were processed in the Agency’s State Statistical Offices since the early 1990’s. These staffing partnership programs were the first steps in moving the entire AEP to NASS’s State Statistical Offices. These partnerships will allow the SARS to take an advisory/supportive role for project management, while investigating alternative remote sensing technologies. During the fall of 1999, the partners were trained for two weeks on using a Microsoft Visual FoxPro application called the Remote Sensing Project or RSP, and PEDITOR software. The RSP application manages the ground data collection efforts, and makes system calls to PEDITOR for handling any image processing function, while PEDITOR performs the clustering and classification, and produces final state and county acreage estimates. Next fall the partners are scheduled to be trained in acreage estimation.

The focus of this study is the mapping aspect of the classification rather than acreage estimation. The demand for supplying large area categorized imagery has greatly risen the past few years as more interests compete for limited land resources, and these images provide a historical snapshot of land cover for cropland areas. In 1997, the SARS created an experimental categorized digital geo-referenced mosaic of Arkansas, and North and South Dakota. One of NASS’s primary goals as an Agency, is producing county estimates, and the regression estimator produces a supplemental indication of county estimates. However, when the categorized images are mosaicked into one geo-referenced image, it’s possible to determine spatio-geographical relationships that were not previously identifiable, such as, in what areas certain concentrations of crops are grown, using temporal analysis to determine the standard crop rotation patterns, identify crop production/migration trends, and acreage changes from farm loss to urban sprawl. The Acreage Estimation’s Program mapping methodology will be discussed as well as future directions of the program.

METHODS

There are six general steps used to create acreage estimates and a final mosaicked image. First, the NASS collected ground truth samples or “training” pixels are clustered by land use/cover type. Second, the training pixels are assigned to a land use class and classified by using maximum likelihood. Third, the entire scene is classified using maximum likelihood. Fourth, the training pixels are regressed with reported crop acreages. Fifth, acreage is estimated with a regression estimator using classified pixel counts as the independent variable and the training data as the dependent variable. Sixth, the categorized images are mosaicked together. The first five steps were previously discussed in (Graham 1993, and Ozga and Craig 1995), the remainder of this study will discuss the process leading to and including the sixth step.

Scene Selection

Scenes were selected for the project based on their inherent quality and phenological timing for maximum land cover separation associated with the growing season. Once a scene was selected for usage in the project, it must be reformatted before it can be used by any of the PEDITOR programs. PEDITOR’s reformat program was used to import the scenes in band interleaved by pixel format, and extract the registration supplied with the scene. The reformattting process does not alter the pixel values, and much effort was extended throughout the entire acreage estimation process to ensure that image integrity was maintained. The PEDITOR system allows for using a variety of sensors for analysis, including; Landsat 5 TM, Landsat 7 ETM+, Indian IRS LISS, and French SPOT MSS. The data sharing partnership agreement established with the Foreign Agricultural Service and the Farm Service Agency allowed access to nearly every usable scene observed by the Landsat 5 TM during the growing season. Scenes were selected based on three main criteria; optimum period for crop separation, data quality, and atmospheric effects (i.e., haze, cloud coverage). Data quality was an important issue when the goal was to classify an entire state. It was necessary to minimize the effects of image defects such as line drops, and the influence of haze and clouds on each scene’s training sites.

It was preferable to have multitemporal coverage over an agriculturally intensive area, as a spring and summer observation can provide adequate spectral separation from row crops versus small grains, but also crops from all other vegetation types. PEDITOR’s overlay program performed multitemporal overlays of two-date scenes (Ozga 2000). The
spring scene usually gets designated the primary scene, while the second scene, normally the summer scene was designated as the secondary scene. The secondary scene was then registered to the primary scene. If only one scene is available during the growing season, then the summer scene is preferred.

To control for phenological and atmospheric effects, each state was divided into analysis districts based on scene overpass dates. The analysis districts comprised a group of counties or parts of counties that were entirely contained within a single Landsat TM pass. Historically, some states had as many as 14 districts (i.e., North Dakota 1998), while some had as few as one (i.e., Arkansas 1997). If a scene in a district was unitemporal, then each scene in that district would have the same observation date. If the scene in a district was multitemporal, then both dates of images would have the same observation date in that district. Each district was processed separately with respect to clustering, classification, estimation, and mosaicking.

Cluster & Classify

Each analysis district was clustered with a modified ISODATA algorithm (Bellow and Ozga 1991) and then classified using a maximum likelihood classifier. It became apparent after the 1997 analysis was completed, that there was a need to collect extra signatures for additional features that were not included in the training data, such as clouds, water, and urban or built up areas. Additional functionality was built into PEDITOR that allowed for feature extraction. Extra signatures were collected on each scene in 1998 to assist with the separation of water, clouds, and urban or built up areas. Each feature type was clustered individually and then processed with the full training set of that analysis district. This provided an efficient method to account for features that were dynamic in nature, such as; clouds and sometimes water, and urban or built up areas that were mostly static. The SARS did not normally account for these features in the acreage estimation process, however, it was necessary to account for these additional features when planning to distribute a visual product.

Image Export

Once all images were categorized, it was necessary to create a method that would convert the PEDITOR image format into a commercial format. Erdas Imagine was the chosen commercial vendor format. An export conversion program was written in PEDITOR that took the original registration coefficients and output an ascii text file that contained the row and column coordinates and their corresponding UTM coordinates. The corresponding reference points were used to compute the transformations, in order to re-project the image. Each categorized image within each state was then rectified to the UTM coordinate system.

The next step involved a reassignment of the categorized image class values. A master set of image categories was established for each state, and all categorized scenes were reassigned to match the master category set. Many categories of the same class (e.g., 10 categories of spring wheat) were combined to one class. Other classes with similar land use characteristics were collapsed together (i.e., idle crop land and summer fallow) into one class. Once the image was properly reassigned, the categories were assigned to a predefined color for each class. When all scenes were mosaicked together for a state, the final output image will have no more than ten to twenty categories.

All scenes within the same analysis district, had the same number of categories. Each analysis district had the same date of observation(s), whether unitemporal or multitemporal. In 1998, North Dakota had 14 analysis districts, whereas in 1997 North Dakota had only eight districts (See Figure 1). The overall classification for 1998 in North Dakota was better than the one performed in 1997. This was simply a function of better observation dates. The 1998 imagery for North Dakota contained nearly all clear multitemporal scenes, with optimum observation dates, but there were no contiguous scenes observed on the same date. Therefore, every scene was a separate analysis district. While the 1997 imagery provided multi row analysis districts, not all dates were optimum for spectrally separating crops. For crop acreage estimation, however, large contiguous analysis districts are preferred.

Once category reassignment was completed, it became necessary to re-project the states that contain more than one UTM zone such as North and South Dakota to one UTM zone. The western edge scenes in North and South Dakota were re-projected from UTM zone 13 to zone 14. A master set of colors was preselected for all of the categories, and applied to each image. A uniform set of colors was now present, and the images were ready for display and mosaicking.
Mosaicking

The image analyst determined the final image priority, based on their knowledge of the classification, the observation dates, cloud/haze coverage, and how the original analysis districts were laid out. The analysis districts were set before the acreage estimation process began, however, the analyst had the ability to re-prioritize image placement based on the visual appearance of the categorized scene, and their knowledge of the area and other available scenes. If the analyst determined that a categorized scene possessed a quality accuracy assessment and was visually appealing, (i.e., crop and non crop fields possess an overall contiguous look rather than a mottled look) then the analyst will bump up the priority of that image over adjacent scenes. Images with cloud problems were assigned a low priority and any overlapping image of better quality were assigned a higher priority, and therefore overlaid onto the lower priority image. Unitemporal images that had poor observation dates for crop separability, such as planting or post harvest, were considered low priority, whereas unitemporal images that cover the crop phenological stages of emergence through senescence were considered a higher priority. Multitemporal scenes have the highest priority as they provide the best means to separate crops and non agricultural features accurately.

The images were stitched together using the previously discussed method for image overlays. As the images were added to the Erdas Imagine Mosaic Tool, the images were not altered or smoothed to account for seam edges. The final product had a somewhat jagged edge like appearance when crossing from an analysis district with a high quality classification to that of a lower quality classification. However, when there were two adjacent analysis districts that had outstanding accuracies and contain quality classifications, then it was very hard to determine seam edges on the mosaic. The SARS continually puts forth great effort to avoid altering the original raw imagery, starting from image import through the final classification. So the intention was to carry that methodology through to the end, regardless of the final appearance. Once the mosaic was completed, the scenes were clipped to fit the state boundary, and exported to Erdas Imagine’s GIS image format.

Distribution

Since the Agency had never produced a digital categorized geo-referenced mosaicked image of an entire state before, the next step was to distribute the images to the cooperating State Statistical Offices and to a limited number of researchers. In 1997, all of the scenes for Arkansas, North and South Dakota were mosaicked. Before distributing any mosaicked images, it was a requirement to bundle the imagery with a self-contained application that could display and browse Erdas type images. This requirement was necessary as the State Statistical Offices lacked the capability to read any type of Erdas format. ArcExplorer from ESRI Inc., was investigated in 1997, and determined to be the easiest and most cost effective method for publishing an “experimental” limited distribution digital mosaic of these states. ArcExplorer comes with a “freeware” type licensing agreement, and allows for the display and browsing of a variety of raster and vector formats. The distribution was small enough to fit each state onto a separate CD-ROM. Included in the distribution were ArcExplorer, the categorized mosaiced image, the land use stratification produced by the Agency’s Area Frame Section, statistics files from the classification that were created in HTML format and written onto the CD-ROM. Statistical information such as the original cover type signatures, the regression analysis by crop, the percent correct and kappa by analysis district, the sampling and area frame information and the Landsat TM coverage/analysis districts were published. The Acreage Estimation Project methodology was prepared in HTML format and published on the CD-ROM.

There was a great deal of interest in the research community for this product in 1997, and that became the driving force to again produce this product in 1998. The SARS released an “experimental” product to the entire Agency, soliciting each State Statistical Office to obtain staffing partnerships with another Federal/State entity that could perform the Acreage Estimation Project in their own state. The same states were again mosaicked for the 1998 crop year. ArcExplorer was again used to publish the data. This time two years of imagery and statistical information were published together for each state on CD-ROM. The States of Illinois, Mississippi, and New Mexico responded back by obtaining funding to support a remote sensing analyst position. Currently, these states are all nearing completion of the classification phase of the project for crop year 1999, and are going to send the SARS the completed categorized imagery. Plans are to mosaic the imagery this year in the SARS, and perform the mosaic and estimation process in the partner’s office for crop year 2000.
DISCUSSION

North Dakota

The North Dakota Land Use Categorization depicted in Figure 2 shows the location of the fourteen unique scenes from this classification, twelve scenes were Landsat 5 TM, and two were from the Indian IRS III 1C. The scene observation dates chosen for this classification were the best available and optimized for crop separation across the state. The image legend in Figure 2 exhibits North Dakota’s crop variety, with some of the lesser crops such as; canola, flaxseed, safflower, and mustard seed combined into another crop category. In the eastern edge of the state, the fertile soils allow for an abundance of different crop varieties to grow, whereas in the western part of the state, small grain production dominates. For display purposes, all other small grains were collapsed into one category, because of difficulty in accurately separating the crop based on their spectral characteristics and growing properties, with the exception of durum wheat. Durum remains a separate category, because it tends to be grown in regions of North Dakota that were spatially separate from any competing small grain.

Changes in production patterns will be evident, as more crop seasons are mapped. Currently, there is an emerging trend for growers producing durum wheat across the state. Durum production has been migrating westward over time. The Agency has noticed this trend statistically the past ten years, but it was never mapped until 1997. The migration can be attributed to the effect of new strains of blight and disease that have forced the growers to change crop varieties in certain areas. Another emerging trend among spring wheat growers, is the total acreage has been down the past few years, the result of an economic downturn in farm prices, and growers moving toward more cash crops. Trends like this can be documented with crop mapping, and modeling where the next big trend may possibly occur.

South Dakota

The South Dakota Land Use Categorization shown in Figure 3 exhibits the concentrations of corn, soybeans, and sunflower on the eastern side of the state. Clouds proved to be a major problem in 1998, as three of the six analysis districts were unimodal, and clouds were impossible to eliminate. The available observation dates in 1998, was not optimal for small grain separation. Therefore, all small grains were collapsed for this state. Other crops such as sorghum, potatoes, and flaxseed, were collapsed into an “other crop” category because of sparse field level training data and for visual mapping purposes. There appears to be an abrupt drop-off in agricultural production once you cross the Missouri River heading west, where rangeland and non agriculturally intensive land were the predominant land use types, and therefore not classified. Also, the NASS sampling scheme west of the Missouri River was too sparse to perform a crop type classification, as indicated by the land use stratification map.

Arkansas

The Arkansas Land Use Categorization depicted in Figure 4 shows the agricultural production in the Mississippi Delta region. The Landsat sensor performs very well in this region, by consistently and accurately separating the predominant crops of rice, cotton and soybeans, given favorable spring and summer observation dates. The classification of unimodal district 03 was over classified with soybeans, but this was due to the late observation date on 9/2/98. The regression estimate was not negatively affected by this over classification, as there was sparse agriculture once you move away from the Delta region in Arkansas. The regression estimate used Agency collected ground truth data in areas where the land use stratum was less than 25 percent cultivated rather than use the categorized images, therefore the ground collected data was used in this area to fill in the estimate.
ISSUES

Registration

When a new product such as this was released, some glitches were expected. One of the SARS’s chief customers, Ducks Unlimited (DU) in Bismarck, North Dakota noticed that the mosaicked images were not lining up uniformly across the entire state. DU discovered that there were inconsistencies in the way the 1997/1998 mosaicked images were geo-referenced. Change detection analysis determined that there was a significant shadowing effect occurring between know features from one year to the next. Also, the shift was not uniform across the state. Some areas of the mosaic were displaced, a few pixels in one direction, while in other areas, the displacement was in a different direction. Determining the cause, and developing a fix was of paramount importance.

DU volunteered to research improved image registration methods and forward the results back to the SARS. DU began searching the North Dakota Department of Transportation’s road GIS data layer 1:126,720 for candidate intersections within a Landsat TM scene. The goal was to obtain approximately 80 usable intersections per scene out of nearly 10,000 total intersections. The first step was to eliminate the minor intersections by choosing only roads that met the class types of; paved, drained, or graded and drained. This left several thousand candidate intersections per scene. The next step was to obtain regularly spaced points throughout a scene, and one way to accomplish this was by sampling the Public Land Survey System’s township, range, and section GIS data layer 1:24,000, also provided by the North Dakota Department of Transportation. The corner of every township was selected and any of the previously selected intersections that fell within 200 meters of a township corner were selected for future consideration. Approximately 100 to 150 intersections were now selected per scene. One last iteration of point selection was necessary to produce a final set of ground control points. The USGS’s 7.5 minute quadrangle Digital Raster Graphics (DRG) was selected to perform the final ground control point reduction. A script was written in ESRI’s ArcView to visually process the remaining road intersections within a Landsat TM scene with DRG (Blunck 2000). The script evaluated each remaining intersection based on how well the road intersection corresponded with the DRG intersection. If the corresponding points were within 15 meters of each other, the intersection was retained as a final ground control point. If the road intersection was just outside of acceptable tolerance, it was shifted to match the DRG. If the intersection exceeded acceptable tolerance, it was dropped from consideration. This left between 90 and 120 ground control points. The output of this process was a database consisting of 90 - 120 quality ground control points, corrected to the 1:24,000 DRG. These points can now be applied to registering a raw base image. The categorized image was then registered to the base image. Any future categorized images will be registered to the raw “base” image.

Categories

The Agency defines non agriculturally intensive land use categories very broadly. The June Agricultural Survey (JAS) sets the acreage numbers for all major commodities, and the AEP training data is based on this survey. The AEP performs very well where high intensity agriculture is present, and accurately separates crop land vs. non crop land. However, the Agency has a few broadly defined land use classes that contain non agricultural, range, or pasture type fields. The JAS non agricultural field classes (e.g., farmstead, waste, woods, non agricultural, pasture, range land, and CRP) are very intermixed within each other. Any one of the previously listed fields may look like each other; pasture land looks like range land, or CRP, or non agricultural, and non agricultural fields can look like just about every other category. These non intensive agricultural areas provide classification problems where the training data is not specific enough to separate these areas. A variety of different vegetation covers can be found growing in these fields (e.g., hay, alfalfa, grass, oats, or scrub vegetation). It is very difficult to separate these non agricultural classes visually on the ground, and just as hard to separate with clustering and classification. Each non agriculturally intensive class was used for training to ensure as much separability as possible from crop fields. All non agricultural categories are aggregated into a composite non agricultural class for visual display purposes. This composite non agricultural class produces a more contiguous look for these fields, rather than a mottled look if the entire non agricultural classes are kept separate.

In the next few years, with the help of the state partnership programs, the SARS hopes to expand the scope of the land use categories by having the partners coordinate efforts on collecting ancillary data at specific locations throughout their own state. Specifically, we need to address the non agricultural areas where more specific land use classes need to be broken out. One possibility is to perform field survey observations to collect perennial and other vegetation types
in areas that continually classify poorly, while using GPS receivers to specifically identify these additional land use classes. Accessing the GIS resources of state and federal governments or private entities may provide another way of obtaining additional ground truth. One method could be to “cookie cut” various GIS data layers from different entities, such as; one designated as cropland by NASS, another designated as forest by the USDA Forest Service, and other land covers provided by the USGS and state governments. The SARS hopes that one or a combination of one of the previously identified ideas can contribute to creating a better overall classification.

CONCLUSION

The extension of the Acreage Estimation Program’s mapping methodology was described through the exporting, mosaicking, and distribution of the images. The results were favorable for large area crop mapping on a seasonal basis. The SARS plans for the 1999 crop year are to register all scenes in the five states to their respective raw base image, and then register the categorized image to that base image, rather than use just the precision corrected coordinates that are distributed and extracted from the raw image. The mosaicked imagery will again be distributed via CD-ROM format, accompanied by accuracy assessment statistics, and ArcExplorer. The full State of Illinois, the eastern edges of New Mexico, and the western edge of Mississippi will be the newest mosaicked products. Further research will be done to accurately control the registration of the categorized scenes to a known base, and to investigate the collection of ancillary data to improve the overall classification.

ACKNOWLEDGMENTS

The SARS is very thankful to Ducks Unlimited Inc., specifically Duane Pool, and Darin Blunck for their tireless efforts in helping to produce an image registration methodology, and for assisting in the registration of the raw and categorized scenes that cover North Dakota.

The SARS is also very thankful to our data sharing partners at the Foreign Agricultural Service/Farm Services Agency for their continued support, specifically Allen Vandergriff, Glenn Bethel, and Brad Doorn, and to our staffing partnership programs with the State Governments of Illinois and Mississippi, and the State Conservationist of the Natural Resources Conservation Service in New Mexico, not only for their support, but for their willingness to try alternative technologies.
REFERENCES


Figure 1. North Dakota Analysis Districts. Note that districts 03 and 06 in 1997, and district 15 in 1998 are either post harvest or in the early emergence stage for crops. These classifications had the lowest accuracies of all districts.
Figure 2. North Dakota Land Use Categorization. Analysis districts 01 and 11 used Indian IRS III 1C imagery, providing some of the best accuracies.
Figure 3. South Dakota Land Use Categorization. Analysis performed using Landsat TM. Land usage east of the Missouri river is dominated by row crops with some small grain production.
Figure 4. Arkansas Land Use Categorization. Analysis performed using Landsat TM. Note that despite Analysis District 01 being unitemporal, an accurate categorization was possible because the observation date was optimal for crop separation.