A History of the Cropland Data Layer at NASS

By Michael Craig

5/10/2010

The first vision of the Cropland Data Layer began in 1971 when a National Agricultural Statistics Service (NASS) researcher was chosen as a Principal Investigator of the LANDSAT 1 satellite program by the National Aeronautics and Space Administration (NASA). This was in response to a 1970 outbreak of southern corn leaf blight which affected the US corn crop. The disease problem presented an ideal scientific research opportunity, and NASS was selected to study the potential of the new satellite to provide information pertinent to the blight research and analysis. Specifically, NASS was to focus on developing methods for identifying crop species and estimating crop areas.

Analysis on LANDSAT 1 began at NASS shortly after its launch in 1972. Early LANDSAT data consisted of 80x60 meter pixels from the Multi-Spectral Scanner (MSS), each having energy readings from four spectral bands, including near infrared. The first project selected study sites in Idaho, Kansas, Missouri, and South Dakota. NASS took the lead for USDA and worked closely with NASA and the Laboratory for Applications of Remote Sensing at Purdue University in program planning and execution. In total, some 17 Federal and State agencies and more than 1,000 people participated in the effort. High-altitude photography was collected in addition to the satellite data. NASS June Area Survey segments of approximately one square mile were used as the data collection units within each study site. Overall, satellite data crop-classification results in Kansas and Missouri were nearly as good as photo interpretations based on aircraft data.

Some basic conclusions had become obvious. Nearly cloud-free satellite data would be needed for acceptable crop classifications. Atmospheric conditions varied enough from day-to-day and from one satellite overpass to the next; specific training was needed for each satellite-data acquisition, rather than applying classification parameters from one date or one satellite scene to other dates and scenes. The most encouraging conclusions were that June Area Survey segments were ideal for developing training samples for analyzing satellite data, and that most satellite scenes in the North Central States had large numbers (20 or more) of segments per scene.

In the mid-1970's NASS was asked to participate in and help evaluate a project called the Large Area Crop Inventory Experiment (LACIE). This effort primarily focused on wheat, and it selected large segments (5 miles by 6 miles) that would be studied each year. At the same time, improvements were quickly made in the procedures for digitizing field boundaries and handling such large data files. A joint agreement was written with the Illinois University Center for Advanced Computation (CAC) to design and enhance a software system named EDITOR to process such large files. Large computers at the Bolt, Beranek, and Newman Data Processing Center (BBN) in Cambridge, MA, were employed; in the late 70's even larger supercomputers at the NASA-Ames facility in California were used. Both the BBN and NASA-Ames computers were accessed through the ARPANET, which was the original version of the present Internet.

Classification of satellite imagery into types by a computer consists may be accomplished in several ways. The NASS supervised approach is to use a sample of known labeled pixels to train

the computer in recognizing and distinguishing the types; and then using the trained maximum likelihood algorithm to categorize all pixels in a satellite scene. A regression estimation approach was used to form acreage estimates for each crop type, which were based on the relative percentages of pixels of each type in the whole population relative to the training data. The regression estimator was one feature which set the NASS software system apart from other remote sensing efforts.

Scenes covering different areas can be merged to give larger area, such as state, coverage. A major 1975 research project was the collection and classification of satellite data for the entire State of Illinois. The Illinois pilot research project took nearly two years to complete. To determine if satellite data could be obtained and processed in time for final end-of-season estimates to be published in early January, a full State study was conducted in Iowa in 1978. Estimates of major crop acreages were created and available to the Iowa State statistical office and the Agricultural Statistics Board (ASB) in time for the "Annual Crop Production Report." Improvements in data handling and other procedures between 1975 and 1978 meant that the Iowa pilot study cost only 40 percent of the Illinois research effort conducted three years earlier.

At this time in the CDL process, desktop computer graphics and Geographic Information Systems did not exist. In 1977-78 the only graphics product was something called a "Dicomed", created by the NASA Ames Cray Supercomputer. A dicomed was a scene classification printed in color on photographic paper; it could also be transferred to a transparency for slide show presentations. These were quite expensive to obtain, both in computer time and materials. However, they gave a glimpse of what was to come.

During the late 1970s and the 1980s, NASS continued to develop and enhance its ability to utilize full LANDSAT scenes to improve estimates of acreages of major crops in important producing States. During this period, the NASS remote sensing efforts were coordinated as part of the Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing Program (AgRISTARS). In the early 1980s, a number of technical improvements were made. Starting in 1981, a video camera and image processing system were used to capture segment and field boundaries instead of having to manually digitize the boundaries. By the late 1980s, technology had evolved to use a super-microcomputer workstation to overlay crop field boundaries onto a satellite-data graphic representation.

In 1983, a multi-temporal approach (which combined a spring satellite scene with a summer scene in order to control for trees and other permanent vegetation) was used for the first time in order to improve late-season crop estimates. Many preprocessing steps could now be done in NASS offices using a combination of mini- and microcomputer components that saved considerably on mainframe computer costs. Another key advancement in 1983 was the start of the conversion of the EDITOR software to a more portable program language. The new version would allow the software to be used on many computer platforms. The new system was referred to as P-EDITOR (for Portable EDITOR). When it was finished by 1986, many routines could be run on personal computers of the era.

AgRISTARS was originally planned as a six-year effort—it would run October 1, 1979 to September 30, 1985, but was later extended an extra year to September 30, 1986. Eight separate

AgRISTARS projects were created; one NASS-led topic was the Domestic Crops and Land Cover (DCLC) Project. In 1980, the DCLC project included two states, and from 1985 to 1987, eight States were included. Early work was targeted to wheat, corn, and soybeans; but other crops such as rice and cotton were the subject of research studies. State level crop acreage estimates were completed by mid-December or so each year. From 1985 on, county acreage calculations for major crops were available by February for use in setting official county estimates.

As the AgRISTARS program came to a close, less research funding was available, and there were serious concerns about future availability of LANDSAT satellite multi-spectral scanner (MSS) data. In addition, a number of foreign satellite systems were being developed that would provide higher resolution data—but at higher costs. A decision was made to discontinue (at least temporarily) the full State crop-estimate projects, and to concentrate on evaluating other satellites and sensors in some smaller test areas.

Three new sensors were available to test in the late 1980's as replacements for the MSS. The latest LANDSAT satellites carried a Thematic Mapper (TM) sensor which acquired data in a 30 square meter pixel with seven spectral bands. The French had launched the SPOT series of satellites with a MSS type sensor (i.e. four bands) but with 20 meter resolution. Finally, the Indian Resource Satellites system had a LISS sensor with 25 meter resolution and four bands. The LANDSAT TM sensor won out in all head to head matches, the increased spectral resolution (3 extra bands, two of which were mid-infrared) made up for the slightly worse spatial resolution (of 30 meters versus 20 or 25).

Full state production began again in 1991, focused on Arkansas and Mississippi; Louisiana was added in 1992. However, when the NASS budget became even tighter and LANDSAT data costs skyrocketed (due to the Government's decision to privatize sales of LANDSAT data), by 1995 NASS had reduced the research efforts to concentrate mainly on 'Delta' counties in Arkansas. These counties have extensive acreages of cotton, soybeans, and rice. Another emphasis in the early 1990s was to explore uses of remote sensing classifications in conjunction with other Geographic Information System (GIS) data layers. When LANDSAT data costs declined and it became possible to return to analysis of data for entire States, NASS researchers started creating a Cropland Data Layer (CDL) product by merging the various classifications in a State.

The CDL was a complete, geographically referenced classification of all satellite data pixels within a State by crop or land use. By using Landsat scenes from multiple times of the year, the CDL did an excellent job of classifying pastures, trees, and other permanent vegetation separately from annual crops. The CDL product would be valuable to a wide variety of researchers and policy officials looking at land-use planning, water quality, environmental issues, and other conditions. Before NASS released any CDL products, it carefully considered whether any confidentiality or proprietary data relationships were being compromised and decided they were not. Another key policy decision was implemented in regards to the CDL. Since a person buying a CDL product for a State could summarize all pixels and essentially create crop county estimates, NASS decided to not release any CDL products until official crop

county estimates were released for that crop season. (Those estimates are normally issued in February following the end of the crop season).

The CDL program proper began in 1997 with a two state startup (the entire state of North Dakota and a twenty-two county 'Delta' region in Arkansas). In late summer 1999, the Enhanced Thematic Mapper on LANDSAT 7 began acquiring data operationally, giving eight day intervals for Landsat type coverage. LANDSAT 5 TM and & ETM imagery were available to NASS via an agreement with the USDA's Foreign Agricultural Service (FAS), which had established an imagery archive and was brokering imagery purchases for the entire Department. NASS had created an extremely powerful remote sensing land-use classification system, but it hadn't the staffing, budget, or mandate to produce many State-level GIS data products.

A 2001 remote sensing paper suggested that the Cropland Data Layer (CDL) Partnership Program might expand to cover the top 15 to 20 cropland States. Indeed, this idea did seem to have strong interest from many sources, particularly State government agencies looking at watershed issues. By 2001, data analysts were located in Illinois, North Dakota, Mississippi, and New Mexico; Section staff completed Arkansas, Indiana, Iowa, and Missouri that year. There also were ongoing negotiations for a partnership with Florida A&M University and one to create a CDL product for the Middle Atlantic States. However, the partnership program did not expand as anticipated. If one person was designated and trained for the CDL project, there was usually no technology transfer when the analyst moved on to other assignments—and the entire effort would close.

Through 2006, LANDSAT TM based CDL products had been created for Arkansas, Florida, Idaho, Illinois, Indiana, Iowa, Louisiana, Mississippi, Missouri, Nebraska, North Dakota, and Wisconsin (with multiple dates for all but Idaho, Louisiana, and Wisconsin). In addition, combined States products are available for the Midwestern States, which cover crop years 2005 and 2006 (the State of Washington is included in the 2006 product), and the Mid-Atlantic States for the 2002 crop year. Products were available over the Internet through the USDA Natural Resources Conservation Service (NRCS) Geospatial Gateway in 2007.

A new problem had surfaced in May of 2003. The LANDSAT 7 ETM sensor developed a Scan Line problem and became unusable under the then current software system. This problem caused missing data strips throughout the image. Using LANDSAT 5 alone would cause data gaps and poor classifications. NASS had to find a new source of imagery and/or software. The Resourcesat-1 satellite launched by India in fall 2003 offered an AWIFS sensor which had 56 meter resolution, possible five-day turnaround and four bands (one of which was mid-infrared). Research studies in 2005 and 2006 reviewed this imagery for application to NASS systems. By 2007 it was considered ready for operational use.

Two other factors affected the CDL program around this same time. Research into uses of GIS-registered data files, such as USDA Farm Service Agency's (FSA) digitized field boundaries for all farms signed up for Federal farm programs was encouraging for CDL work. This would provide a training set of pixels independent of NASS Area Survey Segments. Also, a new artificial intelligence system built on the classification and regression tree (CART) analysis approach and using commercial software was tested and implemented to replace the workhorse

PEDITOR system. The CART approach can handle many satellite imagery dates at once (PEDITOR was limited to two) and can combine other data sources, such as slope and elevation data, at the same time. The CART approach will also be able to accept cloudy satellite scenes by ignoring cloud problems, assuming there are other non-cloudy scenes in the same analysis. It handles the missing LANDSAT 7 ETM data the same way.

Remote sensing interpretations of current-season crop acreages for major States were available by October 1 for the first time in 2007. Those interpretations were based on samples selected from FSA's digitized field boundaries and certified crop uses for 2007 season. AWIFS imagery was used operationally. Increased efficiencies from the software and imagery process led to extension of the CDL in 2008 to all states except Hawaii, Alaska, New Mexico, California, Oregon, Washington, Montana, Idaho, Florida, and the six New England states.

In 2009, a partnership with the Environmental Protection Agency has allowed NASS to extend the CDL program to all the lower forty-eight conterminous states for the first time. An on-board energy problem limited the amount of July 2009 imagery collected by AWIFS, so it was supplemented with LANDSAT 5 and LANDSAT 7 data, where available.