Operational Remote Sensing Program Requirements

Jeffrey T. Bailey*
United States Department of Agriculture, National Agricultural Statistics Service, Washington DC, United States – jeff_bailey@nass.usda.gov

Rick Mueller
United States Department of Agriculture, National Agricultural Statistics Service, Washington DC, United States – rick_mueller@nass.usda.gov

Remote sensing methods and technologies have advanced significantly in recent years and are becoming embedded and operational in some Agricultural Statistical Agencies. The United States Department of Agriculture's National Agricultural Statistics Service (NASS) has used remote sensing to internally produce crop acreage estimates in an operational mode since 2007. Today in-season acreage estimates are generated for 20 different crops in over 27 of the major producing states. This paper will outline the basic requirements for an operational acreage monitoring remote sensing program; satellite imagery, ground reference data, image processing and geographic information systems (GIS) software applications, highly trained staff, robust geospatial and statistical methodologies, and deadline adherence. First and foremost, a reliable, frequent, cost effective and timely source of medium resolution satellite imagery must be available for the area of interest. Second, a source of robust annual inventories of ground reference data is essential for the current year crops as well as at least quinquennial updates for the non-agricultural land cover area. Third, adequate computer and software resources are required with readily available Commercial off–the-shelf (COTS) applications to process and derive the remote sensing land cover classifications and calculate area estimates. A trained workforce with software knowledge and fluencies in the methodologies are prerequisites. Accurate estimates must be produced in a timely manner to be useful for agricultural stakeholders, with final public dissemination of the geospatial land cover data completing the process.

Keywords: satellite imagery; remote sensing; land cover; ground reference data.

1. Introduction

The mission of the National Agricultural Statistics Service (NASS), an agency of the United States Department of Agriculture (USDA), is "*to provide timely, accurate and useful statistics in service to US agriculture*". Towards this goal, NASS conducts hundreds of surveys every year collecting information on virtually every aspect of agricultural activity. In addition to traditional surveys, NASS has a remote sensing program to supplement the survey program.

NASS initiated its remote sensing acreage estimation program in the 1970s and early 1980s with the Large Area Crop Inventory Experiment (LACIE) and Agriculture and Resources Inventory Surveys through Aerospace Remote Sensing (AgRISTARS) (Craig, 2009). These programs assessed if crop acreage estimates could be derived using imagery and ground reference data. Overall they were successful at generating unbiased statistical estimates of crop area and more importantly reducing the statistical variance of acreage indications from farmer reported surveys (Craig, 2009).

These early research programs were successful in producing crop area estimates, while their early efforts were quite expensive, labor intensive, and covered only small regional areas. After LACIE and AgRISTARS completed, the NASS remote sensing acreage monitoring program went through varying phases of research and development as well as operations based on funding support or satellites were no longer available or deemed too expensive. The late 1990's and early 2000's were the dawn of the modern Cropland Data Layer (CDL) program. That research and development phase investigated improved software, enhanced imagery, desktop computing, and began disseminating geospatial CDL data products for the first time. NASS began remote sensing operational monitoring starting in 2007. Since then the operational program began slowly expanding. Today, in-season acreage estimates are generated for 20 different crops in over 27 of the major producing states. Starting in 2011, NASS began serving CDL products to the public via an online dissemination portal called CropScape.

2. Operational Requirements for Acreage Estimation

This paper will outline the key components of an operational remote sensing monitoring program. These components are 1) sustainable satellite image collections, 2) ground reference data, 3) image processing and geographic information systems (GIS) software applications, 4) robust geospatial and statistical methodologies, and 5) highly trained, multi-disciplined, and motivated staff with the capacity to meet strict deadlines and produce accurate defensible estimates. First and foremost, a reliable, frequent, cost effective, and timely source of medium resolution satellite imagery must be available for the area of interest. Second, a source of robust annual inventories of ground reference data is essential for the current year crops as well as at least quinquennial updates for the non-agricultural land cover area. Third, adequate computer and software resources are required with readily available Commercial off-the-shelf (COTS) applications to process and derive the remote sensing land cover classifications and calculate area estimates (Boryan 2011). Accurate estimates must be produced with statistically defensive methodology in a timely manner to be useful for agricultural stakeholders. A highly trained workforce with both geospatial and statistical software competencies, scripting knowledge and fluencies in the methodologies are prerequisites and essential in the production of agriculture estimates. Investigating partnership opportunities with collaborators who have competencies in the development of public dissemination portals of the geospatial land cover data products, promotes open government and serves the agricultural community.

2.1. Satellite Imagery

Satellite image collections are obviously the backbone of any remote sensing monitoring program. Any operational program needs a reliable continuous source of imagery collections throughout the growing season. If an agency's published data is derived partly from remote sensing estimates, it is critical to meet the public’s expectations of timeliness and accuracy for the data. It is also important to develop a defensible, transparent, and repeatable process that can supplement the production of survey based statistical acreage estimates. Interruptions in the delivery of published statistical data can be problematic and even be disruptive to the commodity markets. Having multiple imagery providers constitutes one of the best ways to enhance reliability, program continuity, and enhance the probability of obtaining cloud free collections. When a satellite fails, service is disrupted, or the life expectancy expires, it is always advantageous to have multiple reliable redundant satellite data streams. Additionally, multiple sources will enable comparisons between multiple satellite imagery providers, perform satellite inter comparisons, optimize spectral band selections, and evaluate varying view angle/geometry opportunities, and provide opportunity to analyze and improve the crop estimates. Some image providers now are operating satellite constellations, reducing the risk of relying on a single system without redundancy and one point of failure.

The satellite providers need to provide frequent or high temporal monitoring of the agricultural area of interest. Ideally capturing the entire season of a crop with enough images provides the fullest phenological profile of the crop; thus offering the best opportunity for crop identification and separability from other crops as well as the non-crop areas. Twice a month collections or better are preferred to overcome cloudy conditions and variable climatic regions with seasonal cloud cover issues to enhance the classification accuracies. Landsat’s 16 day repeat cycle can be too infrequent to capture peak crop phenological conditions when clouds are abundant obstruct clear observations when crop separability is optimal. Depending on the climate in your region the imagery provider may need to make multiple satellite passes over certain regions to obtain cloud free imagery throughout the growing season. Imagery providers with a constellation will have greater capability to offer repeat coverage. Some systems are able to point their sensors which gives them extra flexibility, however the off-nadir angles may be large, can add uncertainty, and reduce classification accuracies.

Costs for satellite imagery monitoring was a major limitation in remote sensing up until a few short years ago. Historically, a single scene could cost thousands of dollars. Thankfully some governments are now providing or planning to provide imagery for free. In the United States, Landsat became freely available at the end of 2008 (USGS 2015). Since that time, as of January 2015, over 22 million scenes have been downloaded (USGS 2015). A report of the Landsat Science Team stated “The economic value of just one year of Landsat data far exceeds the multi-year total cost of building, launching, and managing Landsat satellites and sensors.“ In November 2013, the European Space Agency announced that the Sentinel satellite data will also be freely available (ESA 2015). The agency plans to launch the Sentinel-2 satellite in June of 2015.

Some countries have set up their own ground receiving stations. This option necessitates a large upfront investment and it can reduce the overall cost of the images. The Mexican federal government’s Ministry of Agriculture, Livestock Farming, Rural Development, Food and Fisheries (SAGARPA) has built a receiving station for SPOT satellite data. The China Brazil Earth Resources Satellite (CBERS) program which is a joint Chinese/Brazilian partnership delivering free coverage to their countries for purposes of agricultural monitoring. Another way to reduce costs is to have several organizations work together for purposes of purchasing imagery, thus lowering each units cost. The USDA has established a Satellite Image Archive managed by the Foreign Agricultural Service where several agencies collaborate to share costs for satellite imagery purchases.

Statistical agencies typically have publication calendars to inform the public the day and time that statistics will be made available and published. In order to meet the published deadlines, remote sensing data must be made available in a timely manner. The data must be available as scheduled and the time from the image collection to delivery must be rapid; within 24 hours of collection. The rapid delivery time is required to provide the remote sensing analyst adequate time to view the latest collection and determine whether they should use the latest image in their analysis, as it can take two to three business days to complete analysis of one state in the United States.

Determining the optimum spatial resolution as well as optimum scene size are important considerations when investigating potential satellites to use for agricultural monitoring programs. Spatial resolution is the size of the surface area being measured on the ground by each pixel. The larger scene size correlates inversely with spatial resolution. Today improved sensors allow higher spatial resolution with larger scene size. Typically a medium resolution satellite that covers between 10-30 square meters is optimal, see Table 1. Balancing spatial resolution and scene size will depend upon the size and type of the area to be covered and the processing capability of agency. Smaller sized fields may require higher spatial resolutions. There are tradeoffs when considering scene size and analysis area. NASS prefers large scene swaths for two reasons; the large swaths cover more unit area per scene, while too many input scene/image layers can impede the image processing routines, and larger scenes tend to frequently overlap, providing for rapid repeat cycles.

Consideration needs to be given to the number of bands or spectral resolution that a satellite collects. Satellites have sensors for different bands of the visible light frequencies and higher wavelength infrared bands. The primary bands NASS utilizes are the red, blue, green and near infrared bands. The optical based satellites referenced in Table 1 are calibrated and optimized to discriminate different vegetative plant cover types based on how much light is reflected or absorbed, thus providing a unique spectral signature for a given crop during the growing season.

Table 1: Sample of Available Satellite Imagery Options

Satellite	Scene Size (KM)	Temporal Resolution (days)	Spatial Resolution (Sq M)	Bands	Costs
LandSat8(US)	185 KM	16 day	30	7	Free
DMCii(Europe)	600 km	4	22	4	$$
Sentinal 2(Europe)	290KM	10	10-30	13	free
Resourcesat-2(Indian)	740	5	56	4	$$$

2.2 Ground Reference Data

To produce accurate crop area estimates a source of robust annual inventories of ground reference data is essential. It is theoretically possible to perform crop classifications based on historic or a library of spectral signatures, but there are no two growing seasons alike with each growing season having differing planting/emerging times, stress conditions (i.e., drought, flood, disease) varying atmospheric characteristics (i.e., haze, humidity, water vapor), and cloud cover. Many environmental factors in the current year will impact the accuracy of the classifications. The timing of the crop plantings may vary by several days to weeks depending on the precipitation, temperature, and other weather conditions. Continuous cloud cover and the atmospheric conditions will impact sensor collections. Collection of ground reference data is expensive and statistical agencies should look for a partner or source of administrative data to avoid the cost of collecting ground reference data. Many agriculture departments have farm programs to assist farmers and participating farmers may have to report crop data. If partnering opportunities present themselves, it is quite advantageous to collaborate with a partner who captures grower information in a useable geospatial framework that is easily ingestible into your data stream. If there are no sources of ground data then it may be necessary to fund costly ground data collections.

At USDA, the Farm Service Agency administers farm support programs and participating farmers report crop data to the Farm Service Agency in an attributed geospatial framework. NASS has been systematically obtaining Farm Service Agency data since 2007, one of the initial steps enabling an operational program. It's imperative to get updated ground reference data throughout the growing season, capturing the latest plantings and capturing double cropping actions.

For optimal crop classifications all types of land cover should be classified requiring training data for non-agricultural land cover. This is to improve land cover class separability. The most accurate classifications normally have the right proportion of each land cover. When a certain land cover is not well represented, it will likely be underrepresented in the final classification; and can bias classification results. Non-agricultural land cover doesn't change much from year to year, so quinquennial updates would be sufficient.

In the United States, the United States Geological Society (USGS) creates the National Land Cover Database (NLCD) (MRLC 2015). This data set is available every 5 years, but is quite sufficient since the non-agricultural land change is minimal. NASS began using the 2011 NLCD with 2014 CDL production. Other ancillary data layers, such as imperviousness layer, forest canopy, and elevation are also useful, and are also produced by USGS.

2.3 Computer and Software Resources

Historically, processing of remote sensing data required the use of super computers and custom built software (Ozga 1984). Today personal computers are much faster and are also capable of processing large volumes of data. Agencies may want to obtain high end workstation class computers which will give them a boost in performance to handle the data volume and meet operational deadlines. With the wide adoption of Geographic Information Systems (GIS) there are many COTS options for processing of remote sensing data. NASS uses a suite of software to perform CDL processing tasks: For ground reference data preparation, we use the commercial ESRI product ArcGIS; for image preparation/processing - ERDAS Imagine; for ground reference sampling and classification - See5, which is integrated with Imagine via the NLCD Toolkit and uses regression trees for classification; and for regression estimation - SAS, a commercial statistical software application.

2.4 Methodology

A sound geospatial and statistical methodology is critical for a consistent and accurate estimation of crop acreage research. All efforts of a statistical organization require well developed sound methodologies that are well researched, accepted, published, and available for review. After crop classification is complete, a statistical methodology is needed to produce accurate estimates which have measures of uncertainty or variances. For estimation, NASS uses simple linear regression using the classified pixels and data from the NASS June Area Survey. These are overlaid and together these give the best estimate of acreages. When only pixel counting is used, there may be an undercounting bias that is not measured and you will not have statistical measures of uncertainty. The regression approach corrects for classification bias, and reduces the statistical variance of the acreage estimate,

2.5 Trained workforce

A multi-disciplined, motivated and highly skilled workforce with software knowledge, scripting ability, and fluencies in diverse competencies are prerequisites to create accurate statistically defensible estimates and meet strict deadlines. Remote sensing staff need to have skills in geography, remote sensing, information technology, and mathematical statistics. For many statistical agencies, these may be new job series and require special recruiting and additional staff training.

3. Geospatial Dissemination

In addition to official estimates generated with remote sensing a geospatial data layer is a valuable output by-product. The Cropland Data Layer (CDL) has been used for many different uses ranging from research on agricultural sustainability studies, environmental issues, land conversion assessments, crop rotations, decision support, disasters, farmer surveys, carbon, bioenergy, ecology, and biodiversity (Mueller 2013). NASS delivers the CDL via an interactive web portal known as CropScape at http://nassgeodata.gmu.edu/CropScape which provides open access, visualization, and geospatial analytics to the user community (Han 2014). CropScape also supports the ethos of data democracy, providing free and open access to digital geospatial data layers. CropScape utilizes open web standards, thereby supporting transparent and collaborative government initiatives. CDL is available from 1997 to 2014 with full 48 state coverage available from 2008 to 2014.

4. Conclusions

Today, remote sensing offers a viable component of a statistical organization's methods to produce publically consumed statistics. With the key components of: 1) satellite imagery collections , 2) ground reference data, 3) image processing, geographic information systems (GIS) and statistical analysis software applications, 4) robust geospatial and statistical methodologies, 5) highly skilled and trained staff remote sensing can be used operationally in a statistical organization. Remote sensing offers several advantages. First, it can provide a quick turnaround time in output of statistics since image collects are now available the next day. Second, with remote sensing there is a reduced need to talk to the farmers and we can reduce the respondent burden placed upon the agriculture producers. This is a critical concern in the United States with a smaller population of producers and falling response rates.

References

Boryan, C., Yang, Z., Mueller, R., and Craig, M. 2011, Monitoring US Agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. Geocarto Intl. 26 (5), pp. 341-358.

Craig, M. (2009). A Brief History of the Cropland Data Layer at NASS http://www.nass.usda.gov/research/Cropland/CDL_History_MEC.pdf, Accessed April 17, 2015.

ESA (2105). http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Free_access_to_Copernicus_Sentinel_satellite_data, European Space Agency webpage, Accessed on April 17, 2015

Han, W., Yang, Z., Di, L., Yagci, A., and Han, S., 2014. Making Cropland Data Layer data accessible and actionable in GIS education. Journal of Geography, 113(3), 129-138.

MRLC (2015), http://www.mrlc.gov/, Multi-Resolution Land Characteristics Consortium webpage, Accessed on April 17, 2015.

Mueller, R., and Harris, M. (2013), Reported Uses of CropScape and the National Cropland Data Layer Program, Presented at the Sixth International Conference on Agricultural Statistics, June 2013

Ozga, M. (1984). Experience with the use of supercomputers to process Landsat data. http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/GIS_Reports/Experience with the Use of Supercomputers to Process Landsat Data.pdf. Accessed on April 17, 2015.

USGS (2015), http://landsat.gsfc.nasa.gov/?p=9654 Landsat Science team webpage. Accessed on March 6, 2015.