



منظمة الأعدية
والزراعة
للأمم المتحدة

联合国
粮食及
农业组织

Food
and
Agriculture
Organization
of
the
United
Nations

Organisation
des
Nations
Unies
pour
l'alimentation
et
l'agriculture

Organización
de las
Naciones
Unidas
para la
Agricultura
y la
Alimentación

APCAS/10/21
April 2010

Agenda Item 8

ASIA AND PACIFIC COMMISSION ON AGRICULTURAL STATISTICS

TWENTY-THIRD SESSION

Siem Reap, Cambodia, 26-30 April 2010

The Use of Remote Sensing for Area Estimation

by

Robert Hale, Mathematical Statistician
International Programs Office
National Agricultural Statistics Service
United States Department of Agriculture

The National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) is responsible for producing all of the official statistics on all aspects of U.S. agriculture. Production and supply of food and fiber, prices paid and received by farmers, farm labor, and chemical use are some examples. These statistics are produced by gathering information from farmers and agribusinesses dealing directly with the commodity or item of interest. Representative farm operators or agribusinesses are surveyed and the results are used to set estimates of the various agricultural items of interest. NASS also uses remote sensing techniques to supply additional information when estimating area devoted to crop production.

Following is an overview of how to use remote sensing to estimate crop area with the specific example of the methods used in NASS.

Some background information is useful for those who may consider using remote sensing techniques. Some terms will be used describing the attributes of various satellites sensors. First is spatial resolution. Each satellite senses reflected light from a square of the Earth's surface. The digital representation is called a pixel. The size of the pixel varies depending on the purpose of the satellite and can range from more than 1000 meters square to less than 1 meter square. Typically, satellites with a spatial resolution between 30 and 60 meters are used for agricultural applications. This gives a reasonable balance between sensing small fields and covering the large areas associated with agricultural activities.

Second is temporal coverage. This refers to how often the satellite passes over the same spot on the Earth. Temporal coverage can vary from once every day to more than 200 days. The temporal coverage is closely related to the spatial resolution. The smaller the spatial resolution, the less frequent the temporal coverage. Generally, a more frequent temporal coverage is useful for agricultural applications. There is a need to have imagery while the crops are actively growing to be able to distinguish one crop from another. Every time the satellite passes over the area of interest during the peak growing season the greater chance of obtaining useful information. Also highly related to the temporal coverage is the swath width or how wide is the path the satellite senses as it passes over the Earth. The swath width can vary from as little as 16 km to more than 2400 km.

Third is spectral range. This refers to the wavelengths of the electromagnetic spectrum sensed by the satellite. Your eyes are only capable of seeing a very small part of the spectrum. Satellite sensors can be designed to 'see' many wavelengths from very long radio waves to very short gamma rays. Each satellite will be able to sense energy in several wavelengths, each called a band. The number of bands and which wavelengths each cover will vary on each satellite. For agricultural applications, two or more bands in the visible range are useful and at least one band in the near infrared band is needed. The data provided by the infrared bands will help distinguish crop types.

The various sensors are located on different satellites. The satellites are placed into classes based on the spatial resolution of the primary sensor. Some low resolution satellites are used to monitor crop condition or plant vigor. Some of these are; the National Oceanic and Atmospheric Administration (NOAA) series of satellites with the Advanced Very high Resolution Radiometer (AVHRR) sensor, the French Spot Image Corporation's Spot series satellites with the Vegetation sensor and the Terra/Aqua satellites with the Moderate Resolution Imaging Spectrometer (MODIS) sensor. Each have a spatial resolution of more than 250 meters, several spectral bands, a very wide swath width and most importantly, temporal coverage of one day. These attributes give a frequent wide look of the plant conditions, but very little detail.

The second category of sensors is called medium resolution. These can be the most valuable for agricultural applications because the spatial resolution allows enough detail to separate fields and also cover enough land to be efficient. The primary satellites and sensors are the Thematic Mapper (TM) sensor on the Landsat series of satellites, the Multispectral Sensor (MSS) on the

Spot series of satellites and the Advanced Wide Field Spectrometer (AWiFS) on the ResourceSat 1 satellite. Each has spectral bands in the range to distinguish crop types. The Landsat series was the most used for many years, however, the two remaining satellites have technical problems that limit their use. The Spot series still functional, but the narrow swath width of 60 km makes it difficult to join together enough images to cover larger areas. The primary imagery currently used in the U.S. comes from the ResourceSat 1 satellite. The wide swath width and five day temporal coverage make it an excellent choice to be able to collect imagery while the crops are actively growing.

All of these sensors are called passive sensors. That means they sense reflected sunlight from the Earth. A problem occurs when there are clouds. The reflected sunlight comes from the top of the clouds and not from the surface of the Earth. Clouds can cause major challenges when using remote sensing. There are only a limited number of chances to have imagery of the crops during the peak growing season. Therefore, the satellite with the most temporal coverage has the advantage.

There are several steps that must be completed to produce area estimates. The first is to choose the primary satellite imagery. This requires deciding which satellite/sensor best fits the needs of the project. One of the three mentioned early would be likely candidates. Next is to find imagery that has very little or no clouds over the area of interest. This has to be balanced with the time of the year of the imagery. At least one cloud free image is needed during the time the crops are actively growing. This is necessary to distinguish between the various crops. A second image in the spring when the trees and grasses are turning green but before the crops have emerged can help differentiate between crop and non-crop areas. Additional imagery can help, but the marginal benefits decrease with each additional image. In some cases finding cloud free images during the growing season can be a challenge. You would expect areas with enough moisture to produce crops would have cloudy, rainy days during the growing season.

The next step is to gather ground data information. The ground data is necessary to develop a method to determine the crops within the satellite imagery. One of several methods can be used to mathematically relate the ground information to the digital information in the satellite imagery. It is important to have ground information on every crop to be estimated as well as non-crop land covers in the area of interest. The methods to determine crops in the satellite imagery can only work for crops included in the ground information. The ground data can come from test sites, ongoing surveys, or other sources. The information has to be geo-located so it can be aligned with the satellite imagery or other input information.

The primary satellite imagery and ground data are required to use remote sensing techniques for area estimation. There can be other data included that will add to the process to improve the results. These are called ancillary data. This information can be anything that will help distinguish between crops or between crop and non-crop lands. The ancillary data could be satellite imagery that can provide extra information, but would not be sufficient to produce the desired estimates directly. It could be information about non-agricultural areas to eliminate these areas from consideration.

Each of these sources of information must be prepared and aligned to a common base. That base is usually the primary satellite imagery. This means all of the data must be recalculated to the same spatial resolution and put into the same cartographic projection. This will guarantee all the data for any particular location within the area of interest will align. This is not a simple task, but several commercial software packages are available capable of completing the task.

Once all of the data sources are aligned, each pixel in the primary satellite imagery must be classified to one and only one crop or land cover. There are several methods to accomplish this task. Two of the more popular are maximum likelihood classification and decision rule classification. In maximum likelihood the pixels from the satellite imagery aligned with the ground data are used to develop statistical categories representing each crop or land cover. Then each pixel in the primary satellite imagery is compared to every category. The pixel is assigned to the category that is most similar to the pixel based on statistical probabilities. The process continues until all of the pixels in the primary satellite imagery have been classified. A second method is based on decision rules. In this method decision rules are developed based on the ground data and the associated pixels from all of the input data. These rules are a series of “if” statements such as if condition 1 and if condition 2 and if condition 3 then classify the pixel to this crop. Pixels are compared to the if statements until all are classified to one and only one crop or land cover.

When the classification is complete, the pixels are tabulated by crop to prepare for estimation. Some users will simply multiply the area of each pixel by its size to estimate area. This is not the best method. A better method is to build a regression relationship between the area reported to each crop in the ground data information and the area of the pixels classified to each crop in the ground data locations. This relationship can then be applied to all pixels in the area of interest to estimate the crop or land cover area.

NASS has used remote sensing as an additional way to estimate crop area since the early 1980s. The size of the program has varied over the years. The program currently covers the major crops in the lower 48 conterminous states. Estimates are prepared for different crops at different times of the year, depending on the need.

The primary satellite imagery used is from the AWiFS sensor on the ResourceSat 1 satellite. The wide swath width of 739 km is useful to cover the widely scattered agricultural areas of the U.S. The five day temporal coverage increases the probability of having cloud free imagery within the prime growing season even in areas with a high likelihood of cloudy days. There are only four spectral bands, but they are in the ranges well suited for agricultural use. Finally, the 56 meter pixel size is reasonable for the typical field size in the U.S. NASS will still use Landsat imagery in cases where the AWiFS data is not available.

NASS has two sources of the ground data. First, NASS maintains an area sampling frame in the lower 48 conterminous states. Each year an area based sample is selected from the frame and a survey is conducted to produce crop planted area estimates. The selected pieces of land are called segments. These segments are geo-located on maps with the boundaries stored in a digital format to provide the exact location of the land. Field enumerators are given aerial photography and questionnaires and are required to determine the size and crop or land cover for all the land

in the segment. This provides a crop to match on the satellite imagery and an accurate measurement of the area. The area information is used in the estimation process.

The second source of ground information come from another USDA agency called the Farm Service Agency (FSA). The FSA controls the farmer support programs of the USDA. If a farmer wants to apply for government support, he must register his land with FSA and report what crops he is growing. Several years ago the FAS undertook a project to digitally represent every farm field in the U.S. within a geographic information system (GIS) database. The fields are referred to as common land units (CLU). FSA updates the current crop in the field when the farmer registers. NASS has access to the database and uses it as a source of ground data. This source allows for representative ground data for less frequently planted crops that may be rarely found in the area survey.

NASS uses several forms of ancillary data to improve the resulting estimates. One of the most important is the National Land Cover Dataset (NLCD). The NASS and FSA ground data are an excellent source for agricultural information, but not well suited for non-crop land areas. The United States Geological Survey (USGS) along with several other federal agencies maintains the NLCD. The NCLD covers the entire U.S. and has all land classified into 21 different categories. This dataset specifies where to find other types of land covers within the satellite imagery such as grassland, forests, or urban areas. These are needed for the classification procedures to work properly.

Imagery from the MODIS sensor on the Terra/Aqua satellites are also used. The spatial resolution of 250 meters is too coarse for effective area estimation, but the frequent temporal coverage allows for use of multiple dates for a view of the Earth's surface between the primary imagery dates.

NASS typically uses three other forms of ancillary data that can help with the classification process. First is elevation data. Certain land covers are more common at lower elevations and others at higher elevations. Knowing the elevation of a particular pixel in the primary image can help distinguish the cover. Second is slope. Some crops are more likely planted in relatively flat areas and others in hilly areas. No crops will grow where the slope is too steep. Finally, NASS uses a dataset that shows the location of impervious features. Impervious features are things like paved roads, buildings or urban areas. When this dataset is aligned with the primary imagery, any pixels covered with an impervious feature can be excluded from consideration as an agricultural pixel.

All of these datasets are processed and aligned to the first AWiFS image. The AWiFS images have a 56 meter pixel size. The NLCD, Elevation, Slope and Impervious Features dataset have a 30 meter pixel size. The MODIS imagery has a 250 meter pixel size. These datasets must be resampled to match the 56 meter pixel size of the AWiFS imagery. This is a computer intensive process, but it is easily accomplished on a well equipped desktop computer.

Once this is accomplished the areas corresponding to the ground data are separated out of the aligned data. NASS uses a decision tree classifier to classify the imagery. The data associated with the ground data from FSA is used by the decision tree software to build the rules for

classification. When the rules are set, they are applied to the entire image. Every pixel in the primary imagery will be classified to one of the crops or land covers included in the ground data. A classified dataset is saved with the location of every pixel and the crop or land cover assigned to each pixel.

Tabulation is run across the imagery to count the number of pixels classified into each crop or land cover. NASS takes advantage of its area frame in each state in the estimation process. All of the land in a state is stratified into land use strata based on the percent of the land cultivated and the density of buildings. The tabulation is a cross tabulation of land covers within land use strata. Each of the land use strata had a sample of segments for the area survey. During the data collection process the crop and crop areas were recorded. The segment boundaries are aligned to the classified dataset and the pixels within the boundaries are tabulated by crop and land cover. A regression relationship is constructed between the individual crop areas and number of pixels classified to that crop within each segment. This relationship takes the form

$$y = a + bx$$

where a is a constant, b is the regression parameter, x is the number of pixels and y is the resulting area estimate. This form of estimation helps account for any misclassifications and has a smaller variance than the area frame only estimator.

The estimates produced are used by the NASS Agricultural Statistics Board (ASB) as additional information to assist in setting the official crop estimates for the U.S. The remote sensing based estimates use satellite imagery later in the year than the main survey to estimate crop area. Therefore, the remote sensing based indications make use of more current information. In addition, the remote sensing indications have a smaller variance.

NASS maintained its own in-house software to manage and process the data to produce the estimates for many years. As the commercial remote sensing and GIS software improved, NASS changed completely to a suite of commercial off the shelf software to produce the remote sensing based estimates. NASS uses Earth Resource Data Analysis System (ERDAS) Imagine to process all of the imagery, get all of the datasets in the same base and cartographic projection and align these together with the ground information. Environmental System Research, Inc. (ESRI) ArcGIS is used to process the boundary information from the NASS segments and FSA CLU data. Rulequest See 5.0 is the decision tree software is used to process the input data and classify the pixels from the primary imagery to the various crops and land covers. ERDAS Imagine is also used to produce a GIS ready data layer and accuracy diagnostics. Finally, Statistical Analysis System (SAS) is used to generate the regression equations and produce the final estimates. NASS has made considerable effort to facilitate the movement of data files between the software packages and to automate the flow where ever possible.

NASS has produced estimates using remote sensing techniques for many years. There were satellite images classified to crops and land covers produced every time. However, they were not used of anything but producing the estimates. As the GIS software became more powerful and widely used we realized the byproduct of the process is quit valuable to other people. Any study that needs the location of particular crops can benefit. For example, a chemical use study

within a particular watershed can benefit from know which crops are being grown in the watershed. Also agribusinesses can use the location information to determine the best place for a processing facility. The number of possibilities is limited only by the needs of the potential users.