Using Remote Sensing Cropland Classification Data to Update Area Sampling Frame

Xinhua Yu

Rural Survey Department, National Bureau of Statistics, Beijing, China

E-mail: yuxh@stats.gov.cn

ABSTRACT

The National Bureau of Statistics (NBS) of China has begun to use remote sensing technology to estimate crop planting acreage. With the limitation on data availability of current satellite imagery, however, it is more proper to use semi-current or near-historical imagery classification results to update an area sampling frame which is constructed by combining latest land-use census data and the latest agricultural census data. The new methodology has been used to produce cultivated land data (CLD) based land-use data and remote sensing classification data. Based on the CLD, a new area sampling method based on multivariate stratification using classification data was tried and two different methods to estimate using design-based inference and model-based inference was compared for illustrate the advantage of the new design. The further improvement of the new design was also discussed.

Keywords: cultivated land, land-use, sampling frames, crop area

1. Introduction

Area frame sampling is one of the major sampling methods currently used in a large range of countries for agricultural statistics and other statistical survey programmes. The National Bureau of Statistics of China (NBS) began to bring in the area frame sampling method in its crops survey programme recently. There has been a series of pilot surveys implemented since 2010. In 2010, the pilot was conducted in six counties. In2011, province level pilot surveys were conducted in three provinces. The province level pilot survey expanded to five provinces.

Following the rapid development of China economy and society, the rural economic and agricultural structure also changes dramatically. These changes bring in large changes of rural statistical works in aspects of work environment, work conditions, and survey objectives. In this new situation, the agricultural statistical work faces with more difficulties and challenges. The statistical data quality is also affected by these rapid changes.

The conventional crop acreage statistics, which was mainly based on completely reporting system, are facing the difficulties of weakening statistical reporting infrastructure and data susceptible to interferes. After the reform of township administrative institution, the statistical workforce at base level weakened. There is no more specialized statistical staff or institution at base level in many places. The administrative records are not with good integrity and coverage. Some complete statistical reporting from administrative village does not have sound quality. Besides, since the data are reported level by level, it is susceptible to interferes. For example, farmer based survey is susceptible to bias because of agricultural taxation and subsidy policy. When there is taxation, crop area and production are reported less; when there is subsidy, crop area and production are reported more.

Sampling survey are facing the difficulties the sample attrition and representation level. In the recent years, the agricultural structures are changing rapidly. The sample counties of crop sample survey, which were selected in 1983, becomes less representative for some crop kind, such as cotton and oilseed. The biases of some estimates are originated from this cause.

The crop sampling survey using mainly farmers as respondents becomes more and more susceptible to bias because of fast mobilization of rural workforce recent years. Many rural families immigrate outside. Their cultivated land is not easy to trace if survey is based operators' interview. The survey base data were highly affected by these changes. Moreover, the population of frame becomes uncertain, which entailing bias to estimates too.

Farmer based sample survey is difficult to designed to meet sample size constraints and multi-purpose survey objective because of the small parcel operation and decision behaviour diversification of farmers. The outdated survey vehicles and tools, such as tape rulers and measuring dividers, also hampered and affected the survey conducts and results greatly. It really needs big improvements in this aspect.

To cope with all these difficulties, though no all of them can be resolved immediately, some reform can be taken with proper direction. The application of area frame sampling and remote sensing methods should be right direction.

With the assistant of remote sensing satellites, based on the 2nd National Land-use Survey and 2nd National Agricultural Census results, NBS has began to develop a system of Crop Area and Production Estimation using Remote Sensing and Area Frame Sampling Survey (CAPERAS) to produce the planting area and production of crops at provincial level. The CAPERAS has been taken the considerations of the characteristics and difficulties encountered to Chinese crop production such as plant scattered, complex planting structure, division of the south and the north, multiply planted crops in a year, frequently clouding and raining. Taking the pilot survey design in 5 provinces as an example, this paper illustrates the frame updating procedures of the CAPERAS. The result from the output of sampling design has demonstrated that it could meet the requirements of sampling design for CAPERAS by the best use of remote sensing technology. The design results from pilot CAPERAS could completely satisfy the precision requirement of national agricultural statistics in China.

In 2012, the National Bureau of Statistics indicated the need of obtaining estimates at county level for large counties of food production. To meet this request, the decision was made to start with new materials and reconstruct the entire area frame. This will allow the correction of problems that had been identified when using the original area sampling frame. In 2011 and 2012, a complete new frame needs to be constructed using all new materials and the sample designed and selected to provide estimates for counties.

2. Methods and Results

Although the methodological research of remote sensing monitoring on crops was implemented, its operational application at large scale for monitoring crop planting area and estimation directly based on remote sensing data source still faces some difficult problems:

Remote sensing data cannot cover area of interest completely during crop growing season. The cost to acquire suitable data is still high and unaffordable for agricultural agency. Even if data are available, it also is not easy to process in the sure time. The method through complete coverage with current RS data is not applicable for remote sensing crop area and production estimation.

Because the weather factor's influence, it is difficult to acquire high resolution remote sensing data in proper time period.

In the current land tenure system, the right of land use is decentralized. The land for agricultural production is fragmentized. The planting is scattered. Inter cropping and reply cropping is prevalent.

Therefore, if we want to conduct crop area monitoring operationally at national wide, the sampling methods combined with remote sensing must be adopted. It will not only resolve the difficulty of whole coverage remote sensing data, and saving money, but also can carry on quality controls and error analysis. Sampling techniques have been used in large agricultural survey projects worldwide. The MARS project of the European Union adopted a stratified area sampling method. The AGRISTARS of the United States also used area sampling frames.

2.1 Objective

With the support of moderate or high resolution remote sensing data, the 2nd national land use survey data and the 2nd national Agricultural Census data, targeting on major crops (wheat, corn, rice) and other related land cover categories, the methods on area sampling frame construction, the methods of area sampling combined with remote sensing classification and the methods for estimation are developed. After assurance of statistical analytical accuracy, the field sample are arranged, which includes the sample size, the shape of sampling unit, survey content, spatial dispersion and the formula for universe and sampling error estimation. The spatial sampling methods are also developed to support crop planting area measurement and cross-validated with current sampling survey technology.

The area frame sampling survey method for crop planting area will mainly utilize the feature of moderate spatial remote sensing data which can be acquired for large area, periodically and precisely, to resolve the specific problems in conventional sampling surveys. The acquiring capability of moderate spatial resolution at large area and in almost real time can be used to resolve the low data currency problems with base data for crop planting area sampling survey. The advantage of high spatial resolution remote sensing data can be used to relieve the workload of fieldwork in conventional survey.

2.2 Methods

The area frame sampling survey method for crop planting area based on high resolution remote sensing data includes the following steps:

Step1. The cultivated land in land use datasets is in vector polygons format and geo-referenced. The village level administrative boundaries are also included in the same datasets. But with the low data currency of land use data, remote sensing classification was used to update cultivated land in land use datasets. The updated cultivated land data are used to construct the area sampling frame.

Step2. Using the current, multi-phase moderate spatial resolution data, combined with field survey data, the planting area for major crops are extracted with automated classifications.

Step3. Stratified PPS sampling methods was adopted to draw enough administrative villages as sample. Because of the clear boundaries and convenience of field survey, the administrative village is chose as sampling units.

Step4. ZY1-02C high spatial resolution data (5 meter) were used for classification, based on interpreting signatures from the ground truth. An object-based high spatial resolution imagery extraction method was used for crop planting area identification within sampling villages.

Step5. The estimates of county level crop planting area are computed with PPS direct expansion estimator and regression estimator separately.

Step6. The sampling variance and relative error (CV) were computed and analyzed at same time. The effectiveness of the two estimators was compared. The estimates from area frame sampling were also compared with the conventional survey estimates.

The schematic flow chart of the pilot area frame sampling method is as following:

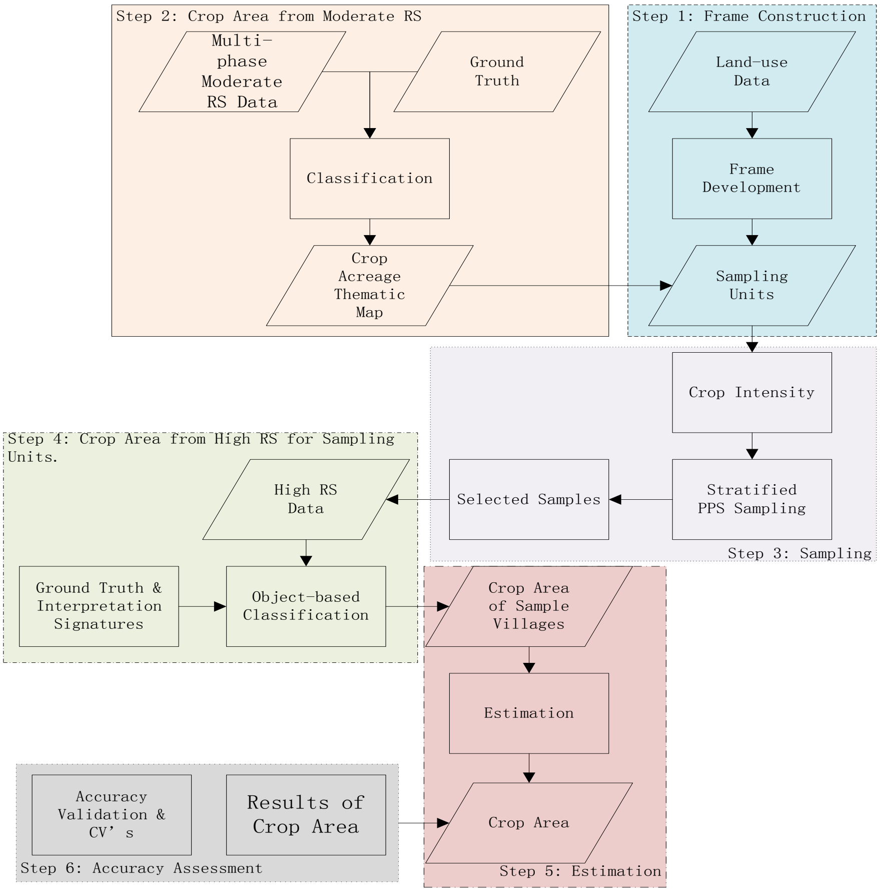

Figure 1 Technical Flow Chart

2.3 Pilot Application.

Taking the year 2012 crops planting area estimation of Beizheng in Liaoning Province as an example, three major crops which are wheat, corn and rice are measured by using area frame sampling survey based on remote sensing data and land use data. The major aspects of the pilot application were described as following:

2.3.1 Crop planting area identification using moderate spatial resolution HJ1A/1B data.

Since the variations in colours and tones of different crops in the moderate resolution images from the same period are not obvious. The major task of crops area extraction from moderate spatial resolution remote sensing is identification of crop type. Major steps include: the choice of remote sensing data time, the processing of data, ground truth interpretation criteria, signature extraction and classification.

The choice of remote sensing data time phase: Because of the differences of crop phenologies, the extraction of crop planting area at county level is mainly depended on the crop phenologies. By selecting the moderate spatial resolution data at different key time period for crop identification during crop growing season and using the characteristic differences of crop types in phenologies, the classification was carried on. According the crop phenologies of Beizheng in Liaoning province, the differences of wheat, corn and rice are evident in April, June and August. In April, rice and wheat were sown, but corn not yet. In June, rice is in irrigation. In August, wheat is harvested, but rice and corn are in maturation. Therefore, those three months are chose for crop planting area extraction.

The processing of remote sensing data: The pre-processing of remote sensing data includes: geometric correction, radiometric correction, and the choice of key vegetation index. Geometric correcting HJ data with SPOT5 (2.5 meters) data as base map, the correcting error will be less than one HJ pixel. Envi 5.0 flaash will be used radiometric correction. Because the time serial NDVIs can reflect the growing progress of crops. So three NDVI were computed in three periods and crop classifications are based on time series NDVI.

The setup of ground interpretation criteria: The setup of ground interpretation criteria is through field survey. Different signatures at remote sensing images for different crop types were created. The cultivated land polygons were overlaid on NVDI data, obvious paddy and field can be identified and obvious differences when field crops turn green.

Signature extraction and classification: The cultivated land polygons were used to create AOI for crop identification, Spectral Angle classification method is used for extraction of crop planting area of Beizheng.

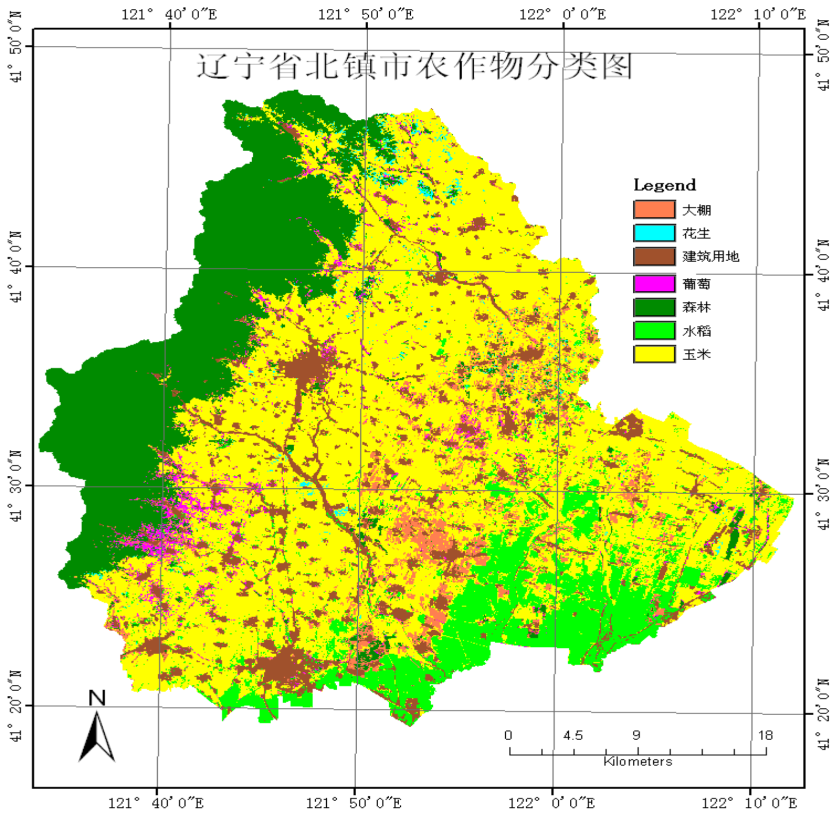

Figure 2 The Crop Classification Map of Beizheng.

2.3.2 Multivariate stratified PPS Sampling, using cultivated land as measure of size.

Multivariate stratified PPS Sampling method was used when sampling villages. 24 sampling villages were selected from Beizheng pilot survey. There are 226 villages in the population.

2.3.3 The extraction of crop planting area within sampling villages.

Object-based extraction method was adopted in this step.

2.3.4 The population estimation and accuracy assessment.

Using direct expansion estimators, the CV of rice is 6.9%， the CV of Corn 5.4%. The wheat is a minor crop in Beizheng, its CV was reasonably large and not reported here.

Using regression estimators, the CV of rice is 5.3%， the CV of Corn 3.7%.

Comparing with the conventional statistics, the ratio of the pilot estimates are 92.48% and 92.86%.

3. OUTLOOK

With the development of remote sensing methodology in China agricultural agencies and the availability of usable multi-source remote sensing data, especially high resolution satellite imagery, it is becoming more liable to the operational application of remote sensing. It is a good manner to improve the area frame sampling with remote sensing, whatever on frame construction, sampling and estimation. But many aspects in area frame sampling should be studied more deeply in China, such as small administrative unit as sampling unit, clustering and ordering of sampling units, small area estimation using remote sensing data as auxiliary variables and model-based estimation.

References

Benedetti, R., Bee, M., Espa, G., Piersimoni, F. Cotter, J. Davies, C., Nealon, J. Roberts, R.,(2010) “Chapter 11.Area Frame Design for Agricultural Surveys,” *Agricultural Survey Methods*, John Wiley & Sons, Ltd.Published Online: 25 March 2010.

Carfagna, E., and Gallego, F. J.(2005) “Using Remote Sensing for Agricultural Statistics,” *International Statistical Review* ,73, 389-404.

European Commission, Joint Research Centre, MARS.() *Best practices for crop area estimation with Remote Sensing.*

Food and Agriculture Organization (1996) *Multiple Frame Agricultural Surveys*, FAO statistical development series 7, Vol. 1, Food and Agriculture Organization of the United Nations, Rome.