

A Sampling Design for Monitoring of the Cultivated Areas of Main Crops at National Scale Based 3S Technologies in China

Quan Wu^{1,1}, Zhiyan Pei¹, Fei Wang¹, Hu Zhao¹,
Lin Guo¹, Juanying Sun¹, Lijuan Jia¹

¹ CAAE, Chinese Academy of Agricultural Engineering, Remote Sensing Application Centre. 41, Maizidian Street Chaoyang District, Beijing, China
wuquan95@tom.com ; peizhiyuan@agri.gov.cn ; wangfei@agri.gov.cn

Abstract. Remote Sensing Application Centre (RSAC) has been working on monitoring the main crop cultivated areas as operational task and research project for several years. The main problems is that the time and features of background data used in sampling maybe produce system errors. The sampling method called stratified sampling adopted by RSAC is also used in this study. But there is an important difference that the background data used to stratify is the same year's cultivated land data derived from an analysis result to MODIS data, containing crop cultivated area proportion information, instead of cultivated land data. The Area Proportion Enlargement Estimation (APEE) model designed in this study is different with Area Enlargement Estimation (AEE) adopted by RSAC. The new APEE model was set up through calculating with twice weighted averaging. A case study limitedly presents the process of the sampling design and testifies feasibility of the method.

Keywords: Sampling, Stratified sampling, Sampling error, Cultivated land, Crop cultivated area proportion, RS, GIS.

1 Introduction

Sampling methods are widely used to estimate main crop cultivated areas by many countries, such as US, China, EU, etc., in the world. But there is much difference among these countries about technologic content introduced below[1][2][3][4][5][6].

1.1 In European Union (EU)

Monitoring Agriculture with Remote Sensing (MARS) is a project facing Europe in order to obtain crop yield information constituted by European Union Committee[7].

¹ Corresponding author: Quan WU, senior engineer, CAAE, mobile phone:+8613621043045

It is a kind of three-stage sampling based on unsupervised classification using multitemporal RS data[8]. The first-stage sampling unit consisted of 60 sites is designed to square with the side length of 40 kilometer. There are 16 component parts in each site and each part has 40 sampling points. Because unsupervised classification method is used with RS data to cover the third-stage sampling points 5-6 times every year, all crops in sampling units are recognized and then the acreage and yield will be worked out after statistical calculation from the third-stage sampling points to the first-stage sampling units.

1.2 In America

In America, the prediction of total crop yield is acquired from crop acreage and crop yield per unit. The crop acreage data is gotten by June Agricultural Survey (JAS) [3][7]. Two different sampling units used by JAS are area frame covering America and name list frame. The name list consists of registered farmers. Every year about 2400 investigators contact more than 120 thousand farmers in the first two weeks of June in order to get crop acreage data [3].

In NASS, these parcels are usually about 6 to 8 square miles in size and are called PSUs, or Primary Sampling Units. These are further divided into 1 square mile parcels, called segments. During data collection, enumerators determine who operates the land within the randomly selected segment and what kinds of agriculture occur on that land. Notice that the subject PSU is yellow and the selected segment is red.

1.3 In China

The operating prediction of crop cultivated area is mainly provided by RSAC. RSAC adopts two methods to obtain the acreage of main crops such as wheat, corn, cotton, soybean, rice, etc.. One method is stratified sampling using RS and the other is ground random sampling using GPS [2][3][4]. The Stratified Sampling Method (SSM) using RS is the major one.

To use SSM, the first step is to select a sampling unit according to the surveyed crop and to order appropriate RS images covering all sampling units in the region. The second step is to discriminate the crop on RS images and get the area of the crop distributed in the sampling units. The last step is calculating the result with AEE (Area Enlargement Estimation) model [3][4].

2 Problems

It is evident that the US' sampling method does not adapt to China for lack of base data while the one of EU also does not meet China for lack of money [4]. However, one problem coming from the monitoring system is that the background data used to stratify is too old to reflect the survey year's ground truth [9]. The time intervals vary from three years to ten years based surveyed crops. So, as a sampling error the systematic error is theoretically produced. Another problem is that the background

data is cultivated land data while the monitoring object is one kind of crop among many crops, such as cotton, corn, soybean, etc., distributed on the cultivated land. So, in the abstract, another systematic error also probably exists [1][4][10][11].

3 Objectives

It is the objective of the research to theoretically set up a new sampling system, including the APEE model, based cultivated land data with crop cultivated area proportion information, for RSAC.

4 Methods

In order to achieve the goal RSAC made up a workgroup to research the above problems. At the same time, designing a new sampling method is also a research project authorized by CAAE (Chinese Academy of Agricultural Engineering), launched by RSAC in 2012.

4.1 General and basic points

Because RSAC's operational system of crop cultivated area monitoring has been working well for many years while the technologic and working systematic structure has come into being [7][12], the research key point should be firstly to theoretically modify the technologic system at some technologic points according to the actual demand produced by change of the background data from cultivated land data layer to crop data layer, not throw it away. It is feasible and reasonable to do so based RSAC's work condition and experience. So, a new sampling method may be set up through rebuilding the old one.

The main task is theoretically to set up a new sampling systematic structure in former half year of 2012 and experiment using the new sampling method in next half year. So, this paper has only given detailed account of the theoretical research for the new sampling method and a case based simulate data. The technologic flow chart is below show in figure 1.

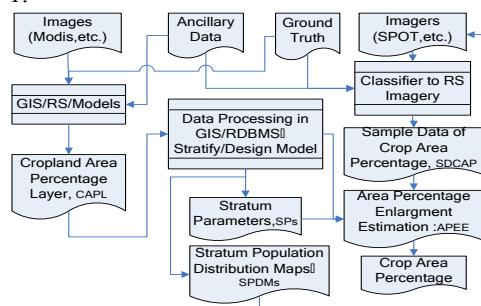


Fig. 1. The flow chart of the designing method

4.2 Sampling units

Sampling units are basic to a sampling system. It is directly related to sampling accuracy and sampling efficiency. At the same time, sampling methods are also involved with sampling units. County administrative boundaries, as a sampling unit, was adopted by RSAC before 2004. In order to decrease the sampling error derived from the different area and shape among sampling units and increase sample size, it was changed to quadrangle frames of relief map at national scale from that year. So, the sampling unit has been used for many years although it also holds systematic errors[4]. In the research, the quadrangle frame of relief map was acted as a sampling unit.

4.3 The Stratified Sampling

Adapting to spatial distribution of Chinese main crops and surveying condition, RSAC has selected a sampling method named SSM acted as main technical method to monitor crop cultivated area for many years. The SSM was also adopted in the research, but the stratified index was changed.

Stratified indexes. A stratified index should describe main characters of surveyed objects. Optional stratified indexes in the research are mainly crop cultivated area, crop type amount and objective crop cultivated area percentage, etc., which are derived from the analysis result of low spatial resolution RS images, such as MODIS, spatially distributed on the surveyed geographic regions[14][15]. The research selects objective crop cultivated area percentage to act as a stratified index for an experiment which will be performed in back half year of 2012.

Confirming an objective crop. Because different crops distribute on different geographic regions with different cultivated habits and area proportion, it is necessary to stratify crop cultivated area proportion only aiming at one kind of crop. So, the stratum parameters vary with crop types. The investigated regions and imaging time of RS images can be confirmed when a kind of investigated crops is acted as an objected crop.

Resampling. As basic data used to stratify is produced of low spatial resolution RS images, crop cultivated area proportion is derived from estimation to pixels and that pixel area is generally not equal to sampling unit area. So, it is necessary to resample crop cultivated area from pixel level to sampling unit level.

Based one important hypothesis that is the objective crop evenly distributed in pixel space, the resampling can be performed. The resampling includes two area transferred steps which orientation is opposite. The first step is to transfer objective crop area proportion to objective crop area according to polygons derived from pixel geographic boundaries. The second step is transfer objective crop area to objective crop area proportion according to polygons derived from sampling unit geographic boundaries.

Stratification. It is simple to stratify directly based crop cultivated area proportion within sampling units. The stratum threshold values can be designed to 30%, 50% and 100%. In this article Frequency Accumulation Stratification (FAS) was used, with which RSAC mainly stratify sampling units in crop cultivated area survey. The essence of the method includes dividing data into groups according to a certain step length such as 0.1%, between two groups in ascending or descending data array, calculating the amount of sampling units of each group, which is called frequency, accumulating frequency and square root of frequency of each group, getting the total accumulating value of square root, obtaining another step length between two layers through dividing the total accumulating value of square root by layer amount such as six, which is an equal step length method. The stratification process is shown figure 2.

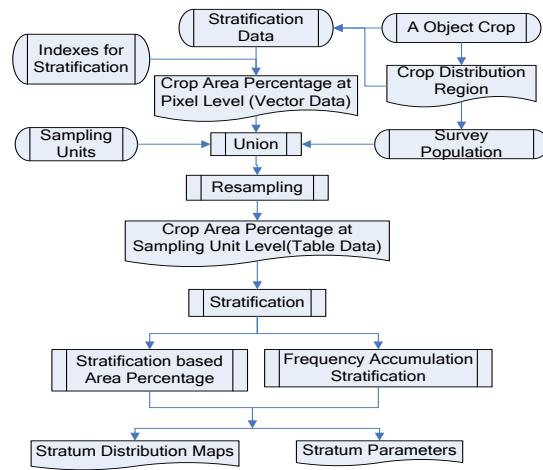


Fig. 2. The flow chart of the stratification process

4.4 Parameters and equations

When stratification is finished, the following task is to calculate stratum sample average and variance based stratum population average and variance. Sample size is a very important parameter. The parameter should be calculated before performing a crop cultivated area sampling investigating.

Stratum sample average and variance. Stratum sample average and variance can be calculated by following equations[3].

$$\bar{y}_h = \frac{1}{n} \sum_{i=1}^{n_h} y_{hi} \quad (1)$$

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 \quad (2)$$

Where \bar{y}_h = the sample average of the h layer
 n = the sample size
 n_h = the sample size of the h layer;
 y_{hi} = the SOV of the i unit of the h layer
 s_h^2 = the sample variance the h layer

Sample size. So, the least sample size is determined by the follow equations[3].

$$n = \frac{\sum_{h=1}^L N_h s_h^2}{W_h} \quad (3)$$

$$V + \sum_{h=1}^L N_h s_h^2$$

Where n = the sample size
 $W_h = \frac{N_h}{N}$
 L = the amount of layers
 V = the variance of estimated value

The others are as same as the above.

Distribution according to proportion:

$$n = \frac{n_0}{1 + n_0 / N} \quad (4)$$

$$\text{Where } n_0 = \frac{N \sum_{h=1}^L N_h s_h^2}{V}$$

The others are as same as the above.

The surveying accuracy signed δ is set to a certain value such as 0.95, and the reliability signed $(1-\alpha)$ is also set to 0.95 or other value, the value of V should be calculated according to the follow equation:

$$V = (d/t)^2 \quad (5)$$

Where $d = (1-\delta)Y$, Y = the total value of the population

$t = \mu_{\alpha/2}$ When the sample size is bigger than 45

4.5 Area Proportion Enlargement Estimation (APEE)

The stratification outcome is used to design a model for estimating crop cultivated area proportion based SOV derived from surveying year. The model is below shown with equations (12).

$$\hat{Y} = \sum_{h=1}^L (W_h \times \bar{y}_h) \quad (6-1)$$

$$\hat{Y}_h = \sum_{i=1}^{n_h} (p_{hi} \times y_{hi}) \quad (6-2)$$

$$p_{hi} = a_{hi} / \sum_{i=1}^{n_h} a_{hi} \quad (6-3)$$

Where \hat{Y}_h = the population average of crop cultivated area proportion

y_{hi} = the sample average of crop cultivated area proportion of the h layer

$$W_h = N_h / N$$

P_{hi} = the cultivated area weight of the i unit of the h layer

y_{hi} = the crop cultivated area proportion of the i unit of the h layer

a_{hi} = the cultivated area of the i unit of the h layer

The others are as same as the above.

The process of designing the model APEE is presented by figure 3.

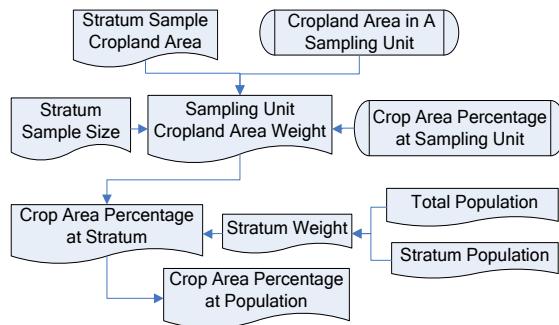


Fig. 3. The model of APEE

5 A case presentation

The case presents the working process including SSM and APEE model using simulative data. According to the process of the sampling designed above, the FAS of SSM was performed on the assumption that the used background data was derived from resampling to low spatial resolution RS images. Figure 4 below shows the result of stratified sampling, in which sampling units are the quadrangle frames of relief map with the scale 1: 50000 and the stratum number is 6. There are totally 3083 sampling units mostly covering 9 provinces' early rice planting region of China in this map.

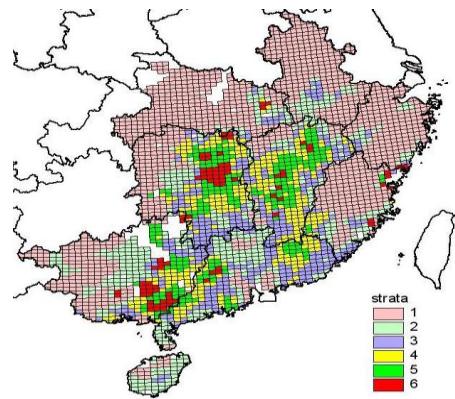


Fig. 4. Stratum distribution of early rice area

The table 1 below shows main parameters produced by FAS of SSM using early rice cultivated area proportion , produced by the latest land-use data and 2009's early rice cultivated area data which is derived from MOA, acted as a stratified index when the population is 9 provinces.

It is evident that the early rice cultivated area in main paddy land is relatively small while early rice plots separately distribute on paddy fields with other crops such as peanut, potato and vegetable, etc..

Table 1. The parameters of sampling

Layers	Population	Minimum sample size	Area percentage means (%)	Sampling rate
1	1483	42	2.99	
2	532	15	15.83	
3	404	12	24.99	
4	339	10	33.29	0.0285
5	232	7	41.32	
6	93	3	57.61	

Using the APEE model with the above parameters and early rice cultivated area proportion data produced by RSAC' monitoring result to early rice cultivated area in 2011 and cropland area data derived from the latest land-use data, the 2011's early rice cultivated area percentage is below shown in table 2.

Table 2. The outcome derived from APEE

Region	Minimum sample size	Practical sample size	Area percentage means (%)	Sampling rate
9 Provinces	88	212	1.80	0.0285
Central China	69	160	1.63	0.0379
South China	84	52	1.82	0.0655
Main productive provinces	48	134	2.86	0.0289

6 Conclusion and problems

Through the case based theoretically calculation and simulate data, a primary conclusion is that the sampling design for monitoring a crop cultivated area, consisting of SSM and APEE, is feasible. But it is very necessary to experiment using practical data in operational productive work.

A problem in the case is that the simulate data used to stratify and estimate crop area is directly produced by calculating with other data. The resampling designed in the sampling for producing stratified parameters was omitted because it is too complicated to simulate the process. It is maybe needed further discussing based another topic.

Acknowledgement

This study is supported by the project named Preceding Study to Agricultural Remote Sensing Monitoring and Evaluating Subsystem which number is E0201/1112-1.

References

1. Wu, Q., Yang, B.J., Pei, Z.Y., Wang, F.: Influence of small features on crop area estimation at a nation scale using remote sensing and a double sampling method. Chinese Journal of Agricultural Resources and Regional Planning. 2004,20(3), pp. 130--133(in Chinese)
2. Jiao, X.F., Yang, B.J., Pei, Z.Y.: Design of sampling method for cotton field area estimation using remote sensing at a national level. Transactions of the Chinese Society of Agricultural Enginerring, 2002,18(4), pp. 159--162(in Chinese)
3. Wu, Q., Sun, L.: Sampling methods using RS and GPS in crops acreage monitoring at a national scale in China. In: Chen Jun, Jiang Jie, Alain Baudoin eds., the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Beijing, 2008, Vol. XXXVII, Part B7, [WG VII/7]* : 1337--1342
4. Wu, Q., Sun, L., Wang, F., Jia, S.R.: Theory of double sampling applied to main crops acreage monitoring at national scale based on 3S in China. EDITORS: D. Li, Y. Liu, and Y. Chen(Eds.), Computer and Computing Technologies in Agriculture IV, CCTA2010, Part III, IFIP AICT 346, PP. 198--211
5. Zhao, T., Pravesh, Alfred, S.: Geochemical sampling scheme optimization on mine wasters based on hyperspectral data. In: Chen Jun, Jiang Jie, Alain Baudoin eds., the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Beijing, 2008, Vol. XXXVII, Part B7, [WG VII/7]* : 1529--1532
6. Xie, H., Tong, X., Jiang, Z.: The quality assessment and sampling model for the geological spatial data in China. In: Chen Jun, Jiang Jie, Wolfgang KAINZ, the International Archives of the Photogrammetry,Remote Sensing and Spatial Information Sciences, Beijing, 2008, Vol. XXXVII, Part B2,[WG VII/7]* : 819--824
7. Yang, B.J.: Monitoring the agricultural condition using RS. China Agriculture Press, Beijing, 2010, pp. 19--29(in Chinese)
8. Duda T.: Unsupervised classification of satellite imagery: choosing a good algorithm. Remote Sensing, 2002, 23(11), pp. 2193--2212
9. Wu, Q., Sun, L., Wang, F.: The applications of 3S in operational monitoring system of main crops acreage in China. EDITORS: D. Li, S. X. Yang, Computer and Computing Technologies in Agriculture - II , TSI Press, USA, 2010, Volume 24, pp. 319--324
10. Sun, L., Wu, Q., Zhang, S.L., Wang, F.: Estimation of area ratios of small features in crop area obtained by remote sensing technique in Hebei Province. Transactions of the Chinese Society of Agricultural Enginerring, 2006, 22(2), pp. 79--82.(in Chinese)

11. Huang, H.P., Wu, B.F.: Analysis to the relationship of feature size, objects scales, image resolution. *Remote Sensing Technology and Application*, 2006, 21(3), pp. 243--248. (in Chinese)
12. Wu, Q., Pei, Z.Y., Zhang, S.L., Wang, F., Wang, Q.F.: The methods for monitoring land-use change with RS at non-large scale. China Agriculture Press, Beijing, 2010, pp. 33--40 (in Chinese)
13. Yang, X.: Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. *Remote Sensing*, 2002, 23(9), pp. 1175--1798
14. Zeng, Y., Wu, B., Li, G.: Using Data Grid technology to build MODIS data management system in agriculture application. EDITORS: D. Li, Y. Liu, and Y. Chen(Eds.), *Computer and Computing Technologies in Agriculture IV*, CCTA2010, Part III, IFIP AICT 346, PP. 241--248
15. Thomas M. Lillesand: *Remote sensing and imagery interpretation*. John Wiley & Sons, Inc, New York, 2002, pp. 576—586