Determining Crop Area Proportion Based on the Specific Sampling Method Applied

Shengjun Wu (Corresponding author)
Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China
Tel: 86-27-6888-1075 E-mail: wsj@whigg.ac.cn

Bingfang Wu
Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China
Tel: 86-10-6485-5689 E-mail: wubf@irsa.ac.cn

Qi Feng
Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China
Tel: 86-27-6888-1075 E-mail: fengqi@asch.whigg.ac.cn

Huaiping Xue
Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China
Tel: 86-27-6888-1075 E-mail: xuehp@whigg.ac.cn

Yun Du
Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China
Tel: 86-27-6888-1075 E-mail: duyun@asch.whigg.ac.cn

Yuanzheng Li
Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China
Tel: 86-27-6888-1075 E-mail: rushfuture@sina.com

Received: July 28, 2011 Accepted: August 8, 2011 Online Published: December 29, 2011
doi:10.5539/jas.v4n3p123 URL: http://dx.doi.org/10.5539/jas.v4n3p123

The research is financed by China Hubei Provicial Science & Technology Department, No2010DEA003 and No2008CDA093

Abstract

The aim of this study was to provide a method for acquiring Crop Area Proportion (CAP). Several processes were employed to achieve this objective. A two-level general regional structure was first applied to establish the crop planting structural zone in which to provide a scientific basis to obtain CAP. Sampling units were then ascertained based on the two-level structural zones. Random samples were taken according to the standard procedure of the two-stage stratified sampling method, and "sampling trees" were built from them. GIS was used to calculate proportion by way of landscape imagery of the various crops within the sampling database. Statistical methods were then applied to convert these proportions into "sampling tree" planting proportions. High resolution remote sensing images taken at three to four sample sites were selected to test the accuracy of CAP. Results show that this method produces a high-level of performance when applied to large agricultural areas.

Keywords: Crop area proportion, Two-level regional structure, Sampling tree

1. Introduction

Determining crop area is the basic step in remote sensing crop estimation. After nearly 30 years of development in remote sensing crop estimation, however, no suitable means has yet been devised to acquire crop area accurately and in a timely manner. Formulating crop area is a challenging problem in China owing to its complex natural topography, planting structure variety, small and dispersed block divisions of crops, and the different agricultural techniques applied (Wu, 2006). Nowadays, satellite remote sensing data are used to estimate crop area. Nevertheless, accurately estimating crops in this manner fluctuates in terms of cost due to the cost of spatial and temporal resolutions and the cost of acquiring satellite data (Zhao et al., 2005; Wood, 2003; Doraiswamy, 2004; Rasmussen, 1992; Hayes, 1996).

Crop area proportion (CAP) specifically relates to the proportion of land area that a particular crop occupies as a ratio of the entire agricultural area in which it is located (Wu, 2004). It is typically represented by a ten-point system. For example, one CAP equals 10%. Obtaining CAP is important in obtaining the crop area for crop estimation. If CAP and the total area of arable land are known in a region, for instance, the area of each crop can be easily ascertained. Sampling techniques used in this study were based on the principles of statistics. This makes the large-scale acquisition of CAP possible.

2. Theoretical basis and technical procedure

2.1 Theoretical Basis

2.1.1 The basis of statistics

Sampling is a part of statistical practice concerned with the selection of an unbiased or random subset composed from observations taken from a population of individuals with the intention to yield some knowledge in relation to the population under consideration, especially for the purpose of obtaining predictions based on statistical inference. Certain values within the subset can be used to determine CAP on a large- scale. Planting area proportion, for example, can be determined from the parent population by means of a precise examination of the unique representation of the values. This approach is in keeping with the procedure of statistical theory (Wu, 2004).

Stratified sampling is considered the best way to determine CAP owing to differences in crop planting structure and patterns in arable land practices as exhibited in different regions. This method also supports the inherent characteristics of CAP. Stratified sampling can be divided into different proportion allocation types according to arable land patterns and their unique planting structures. A simple random sampling method can be adapted for different allocation types with the goal of acquiring CAP. For this study, the level of stratified random sampling is the unit of the crop planting structure. Therefore, the regionalization framework of the crop planting structure is the key technique used in this research.

2.1.2 The basis of the zoning study

According to stratified sampling, samples are processed within different strata, and the division of stratification is based on the existing zoning system.

The regionalization framework is the division of a certain geographical space based on the law of zonalization. The geographical space, like the object, is divided into regional units that consist of multilevel structures, according to the basic zoning characteristics of spatial distribution. The regionalization framework carries out two tasks according to how it will be used. First, the geographical space is divided into different regions to retain unit features. Second, these divisions are subdivided into sub-regions based on the differences within each region. Thus, a regional rating system is formulated to reflect the distribution patterns of the elements of the regionalization framework. To ascertain CAP, which is the focus of this study, the regionalization framework is carried out on the crop planting structure in accordance to certain principles and practices. A sampling survey is then applied to each regionalization framework unit.

2.1.3 Technical foundations of “3S”

The appellative of geographic information system (GIS), remote sensing (RS), and global positioning system (GPS) is termed “3S.” Together, they make up the three supporting technologies used for the acquisition of spatial information, storage management, updating, analysis, and application in NASA’s Earth Observing System. 3S is an important technical practice used in the sustainable development of modern society as well as the rational planning and utilization of resources. It also supports urban and rural planning and managerial practices, dynamic monitoring, natural disaster control, etc., and is an important scientific method used in the quantification of earth science research (Li & Li, 1998).

GIS serves as the basic platform for this study. It not only carries out basic data integration operations but also performs functions such as data computing, input and output, and so forth. The positioning function of GPS provides spatial attributes for the sampling data, and RS supports a database for use in crop planting regionalization as well as allowing for data testing.

2.2 Technical procedure

The specific technical procedure to obtain CAP is provided in Fig. 1. The schematic is comprised of five parts in total and incorporates the crop zoning procedure, the sampling line transect layout procedure, the sampling procedure, the data processing procedure, and the testing procedure. The five sequences are relatively independent of each other. When integrated into the GIS platform, together they constitute a technology schematic to obtain CAP.

<Figure 1>

2.2.1 Crop structural zoning procedure

The purpose of zoning crop structure in this study was to provide a scientific basis for the method to obtain CAP. A general agricultural zoning system could not be used. For this study, the crop planting structure was the principle factor, and the supporting factors include topography, soil, climate, product, land use, etc. Furthermore, a combination of qualitative analysis and quantitative calculation is applied in this method with the support of GIS to establish crop planting zones. The initial step of the method is spatial analysis. Many additional factors were also considered. Therefore, not only is the continuity of crop blocks considered but also characteristics regarding crop yield estimation. Note that the method was not designed for a specific crop; rather, it is a comprehensive reflection of crop planting structure within a particular region. This crop planting structural zoning framework is necessary to follow the basic principles regarding integrated natural zoning. Special needs of crop planting structure must be taken into account at the same time.

According to the procedure of stratified sampling and the specific status of crop planting in China, a general two-level regional structure was used to enable the zoning of the crop planting structure. This will ensure that regional characteristics meet the requirements of sampling technology procedures while also avoiding the drawbacks caused by employing excessive zoning details to samples.

To begin with, the dynamic processes of various elements and the relationships between them were considered as were the rules guiding zonal differentiation. The primary benefit of this was in the selection of sample sites. Based on this principle, the first level was implemented according to the basic order of the factors that follow: regional location, geomorphology, climate, and arable land. After examining the first level indicators it was determined that crop planting effects and procedures in relation to orientation, topography, climate, and land type, etc., of the zoning units were the most important factors.

Azonal differentiation was regarded for the second level zoning, which primarily reflected on the statistical characteristics of crop products. The principal indicators were the percentage of agricultural land area to the total land area (the agricultural land area ratio) and the percentage of estimated sown area crop yield to total grain sown area (the sown area ratio for crop yield estimation). This approach allows for relative consistency between crop types and crop structures. It also provides reliable results in terms of crop interpretation and the treatment of information.

2.2.2 Sampling line design procedure

Sampling units were based on the second level units after they were determined and used to carry out random sampling according to the standard procedures of two-stage stratified sampling. Factors that were considered in the actual sampling process were the sample region, the sampling operation, sample size, and other related factors. The quadrat layout was put in place after the sample regions were selected.

Quadrat layout did not follow the traditional point mode or block mode nor did it follow the simple line mode. Instead, a dendritic sampling line was designed as a result of field work experience gained within the study area. A central line transect was first put in place, and then it was overlaid with supplementary line transects. Owing to its tree pattern, this line transect is referred to as a “sampling tree” sample method. Therefore, the central line transect was laid out like a tree trunk after which secondary line transects were overlaid like branching of a tree. The design of the line transects were based on rural roads. This means that the representative roads located in various sampling areas were in themselves chosen as line transects according to sampling techniques and procedures. If a state road that portrayed relevant characteristics was chosen as a central line transect, the interconnected county roads or township roads could be used as accessorial line transects. Both the central and accessorial line transects were sampled in the same manner and aggregated into the same database.

2.2.3 Sampling implementation procedure

The sampling method utilized for this study was attained through video streams of the agricultural landscape within the line transects sampling area. The video stream format was MPEG-4. Mainstream digital camera devices currently on the market all support this format. The line transects were transmitted to a computer sampling system by way of the video streams. Static landscape images were produced in the sampling system for which certain time intervals were applied, and then they were saved to the GIS platform database in the sampling system by means of the GPS instant access feature. This database was designated the sampling data.

2.2.4 Data processing implementation procedure

The core means used to obtain crop planting proportion in the sampling area was via data processing of the sample data. A data processing program in which to accomplish this was developed using the SuperMap® platform. The program was designed to calculate the proportion of landscape images of various crops in the sampling database, and turn this data into the planting proportion of the "sampling trees" by statistical methods. The proportion of the "sampling trees" then becomes representative of the planting proportion of its sampling area. Fig. 2 provides some sampling images.

<Figure 2>

2.2.5 Proportion accuracy test research procedure

Three to four sample areas were selected to test the accuracy of CAP. The test program incorporated both the high resolution remote sensing images and the actual survey data. High resolution remote sensing images (SPOT5 or QUICKBIRD) taken from the sampling area were selected. They were then interpreted, and cropland block maps were made. The area of each cropland block was determined with the support of GIS, and a field survey related to the crop planting status within each cropland block was generated. The different widths for each sample line, the crop area proportion for the various types of crops within the buffer zone, and the crop area proportion for the various types of crops in the whole sampling area were then calculated. At this point a test was carried out in combination with the proportion obtained by way of sampling to determine accuracy.

3. A case study

Hubei Province, which is located in the middle reaches of the Yangtze River, is situated within a subtropical monsoon region. It is one of the primary agricultural provinces in China. Accurate crop estimation in Hubei Province is important for agricultural activity throughout China since it reflects the country as a whole.

3.1 Research on regional crop planting structure in Hubei Province

In accordance with the actual requirements of the project, the principle and basis for crop planting structural development and, particularly, the basis for the regionalization of crop planting structural development in Hubei Province were established. Two levels of regionalization were set up in relation to crop planting structural development in Hubei Province in compliance with the technical approaches discussed earlier. The first level dealt with the dynamic processes of various elements and the relationships between them in combination with the rules guiding zonal differentiation. This was primarily instituted for the selection of sample sites. Based on this principle, the first level division was implemented according to the basic order of the following factors: regional location, geomorphology, climate, and arable land. The indicators were primarily the orientation of the zoning units, topography, climate, land type, and the like.

The second level dealt with non-zonal differentiation, which reflected on the statistical characteristics of crop production. The principle indicators were the percentage of agricultural land area to total land area (the agricultural land area ratio) and the percentage of estimated sown area crop yield to total grain sown area (the sown area ratio for crop yield estimation). Not only were the estimation of agricultural products in crop producing areas as well as products for the estimation of crop yield reflected, feature types and feature structures remained relatively consistent, providing evidence for the identification of crops and crop development. Landscape type, annual hours of sunshine, annual precipitation, multiple cropping indexes, crop yield per unit area, and the production value percentage of planting to the entire production of a farm were also selected.

Based on county-level administrative units, 82 basic units of division in Hubei Province were identified. After deliberating on several aspects such as terrain type, climate and weather, land use, etc., the crop structure zoning division was implemented based on the dynamic clustering approach. Five first level zoning units were established such as the humid hilly area used for rice production located in southeast Hubei Province, the sub-humid low hilly area used for both arid-land farming and humid-land farming production located in north Hubei Province, the semi-arid mountainous area used for arid-land farming production located in northwest Hubei Province, the sub-humid plains area used for both arid-land farming and humid-land farming production located in central Hubei Province, the humid mountainous area used for both arid-land farming and humid-land farming production located in northwest Hubei, etc., as well as 12 second level units (shown in Fig. 3).

<Figure 3>

3.2 Research on crop area proportion sampling line layout in Hubei Province

From a sampling perspective, the investigation of CAP falls within the sphere of spatial sampling due to its low proportion, large range, and high spatial heterogeneity. A line transect sampling framework is a typical two-stage stratified sampling method. In terms of layout, since line transect sampling frameworks set one highway-side buffer zone within the primary sampling units, the sampling level can be reflected by the length of the sample line after the width of the buffer and the sampling rate have been determined. Theoretically, the second phase sampling level should be based on the actual condition of each primary sampling unit. In actual sampling processes, however, since each primary sampling unit differs in terms of natural condition as well as in crop planting structure, there is no effective way to do so at present if calculating the sample size on the basis of the planting structure and distribution characteristics of the crops. Sampling accuracy can be ensured by means of the length of the specified minimum sampling line in the form of the lowest sampling rate. Throughout this article it has been demonstrated that the length and density of highways in Hubei Province can both meet the inherent requirements of line transect layout. Through the utilization of the random sampling method, the sampling line length that meets the requirements of crop samples in Hubei Province were calculated and then laid out.

3.3 CAP sampling methods applied in Hubei Province

The planting proportion of various crops on farmland was not what was seen when field samples were processed. What was seen was a variety of crops. Video capture devices were affixed in a moving vehicle that travelled at a speed less than 40 km/h in line with the sampling line. With the aid of a computer, the captured video images were processed into the database. At the same time, the corresponding geographic coordinates, sampling time, and zoning units of the imagery as well as traffic-related parameters, etc., were recorded.

3.4 Calculating CAP

Field photos captured by the sampling system were used to calculate CAP. The various planting proportions of the crops were determined in each photo by means of manual interpretation and then divided into different statistical units to establish the CAP of the designated statistical units. The formula used was:

$$P_j = \frac{\sum_{i=1}^{N} a_{ij}}{\sum_{i=1}^{N} A_i} \qquad j=1,2\ldots,M$$

where *Pj* is the planting proportion of J-crop; a_{ij} is the area proportion of J-crop in I-photo; *Ai* is the percentage of crops in I-photo; *N* is the valid photo record within the statistical units; and *M* denotes there are *M* kinds of crops within the zoning units. The planting proportion was calculated based on by the county-level administrative units and the planting structure units.

3.5 Sampling results

The summer crop harvest in Hubei Province was sampled in 2004 by means of the sampling system developed for this study. According to crop growth phenological characteristics in Hubei Province, mid-April was determined the optimal sampling time. The main crops monitored were rape, early rice, winter wheat, vacant land (fallow land), vegetables as well as other crops. Among these crops, early rice would be in the early stage of its development during this time where it would be just breaking ground or just being planted. Therefore, it would warrant preliminary monitoring.

Sampling was carried out in accordance with the established sampling lines, starting from Wuhan then passing through Xiantao, Honghu, Jianli, Qianjiang, and Jiangling as well as other counties and cities. Sampling time lasted for eight days. Total distance travelled was nearly 5000 km, and in that almost 2000 km was sampled. A total of 41,356 images, 15 database files, and 31,485 effective records were collected. The system performed well throughout the sampling process and, according to the scheme put in place, the sampling task was completed successfully.

Through the use of this method, data was collected and analyzed to establish CAP for each sampling administrative zoning unit, each planting zoning unit, and the province as a whole.

<Table 1>

Statistical data from 2004 was used to make a comparative study. An estimate was first established through the use of the statistical data according to the format provided for in Table 1. A comparative analysis was then carried out between the results and the proportions to establish the correlation coefficient. Due to the scope of this study, correlation analysis was carried out on only four main crops, that is, wheat, rape, early rice, and vegetables (cole, potato). Five correlation coefficients were obtained for the zoning units of level 1, 12 for level 2, and 28 for the counties sampled for a total of 45. Among these correlations, the largest was 0.9990 and the smallest was 0.1402. For all correlation coefficients, 31 number values were greater than 0.8, accounting for 69%, while 41 number values were greater than 0.5, accounting for 91.1%. The correlation coefficient was

0.9955 between the total sampling data and the average sampling proportion for the whole province. According to the correlation coefficient, a good correlation existed between them. Therefore, the regression equation could be built to carry out the inversion of the statistical data at least to a certain extent.

4. Outlook and conclusions

A 5 × 5 km^2 experimental area was selected to verify the precision in determining CAP with actual data from Gongan County, Hubei Province. The data was acquired by means of a detailed field investigation, and the method used to compare crop planting area to the field data was a comparative analysis using high resolution satellite imagery (integrated SPOT5 data of 2.5 M-resolution and 5M-resolution). Results show that the method used in this study offers a high level of success when used in conjunction with large crop areas. The three types of agricultural crops that occupied the largest area within the experimental zone were crops, vegetables, and cotton. The accuracy was 98.3%, 98.6%, and 93.9%, respectively, while the accuracy of satellite data was 88.6%, 84.7%, and 96.3%, respectively. Both the method used for this study and the use of high resolution satellite imagery showed greater accuracy in acquiring the proportion of large-scale crops. However, the accuracy in obtaining CAP by the method used for this study was greater compared to high resolution remote sensing data with regards to scattered crops located in small areas. The method is effective and works well in relation to the principal crops grown in a region, especially for the crops grown in large areas. Further study is required to monitor crops grown in smaller areas.

References

Doraiswamy, P.C. et al. (2004). Crop condition and yield simulations using Landsat and MODIS. *Remote Sensing of Environment,* 92, (4):548-559. http://dx.doi.org/10.1016/j.rse.2004.05.017

Hayes, M. J., & Decker, W. L. (1996). Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. *International Journal of Remote Sensing,* 17(16): 3189-3200. http://dx.doi.org/10.1080/01431169608949138

Li, D.R., & Li, Q.Q. (1998). The formation of Geo-spatial information science. *Advance in earth sciences,* 13(4):319-326. [Online] Available: http://d.wanfangdata.com.cn/Periodical_dqkxjz199804001.aspx

Rasmussen, M.S. (1992). Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from AVHRR. *International Journal of Remote Sensing*, 13, (18): 3431-3442. http://dx.doi.org/10.1080/01431169208904132

Wood, G.A., Taylor, J.C., & Godwin, R.J. (2003). Calibration Methodology for Mapping Within-field Crop Variability using Remote Sensing. *Biosystems Engineering,* 84(4), 409-423. http://dx.doi.org/10.1016/S1537-5110(02)00281-7

Wu, B.F., Li, Q.Z. (2004). Crop area estimation using remote sensing on two-stage stratified sampling, Proceedings of the XXth ISPRS(the International society for photogrammertry and remote sensing) Istanbul Congress, Istanbul. [Online] Available: http://www.isprs.org/Proceedings/XXXV/congress/comm7/papers/49.pdf

Wu, S.J. (2006). Acquiring Crop Area proportion Based on “3S” Technology--Taking Hubei as a Study Case. China University of Geosciences, Wuhan.

Zhao, C.J., Liu, L.Y., Wang, J., Huang, W.J, Song, X.Y., Li, C.J. (2005). Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress. *International Journal of Applied Earth Observation and Geoinformation*, 7(1), 1-9. http://dx.doi.org/10.1016/j.jag.2004.10.002

Table 1. Statistical units of CAP for summer crops harvested within Hubei Province in 2004

Zoning Unit	Proportion of Wheat	Proportion of Rape	Proportion of Early Rice	Proportion of Vegetables	Others
Xinzhou District	0.71	3.6	4.42	0.75	0.52
Jiangxia District	0.32	4.21	3.78	1.08	0.61
Ⅰ-1 Subtotal	0.51	3.87	4.08	1.06	0.48
Daye City	1.18	4.79	2.98	0.55	0.5
Chongyang County	0.25	3.6	4.95	0.34	0.86
Ⅰ-2 Subtotal	0.47	4.55	3.8	0.42	0.76
Xishui County	2.1	3.9	3.26	0	0.74
Qichun County	0.22	6.25	3.1	0.43	0
Ⅰ-3 Subtotal	0.43	4.78	3.67	0.35	0.77
Ⅰ Total	0.66	4.39	3.43	0.98	0.54
……	……	……	……	……	……
Province total	3.55	4.47	1.12	0.35	0.51

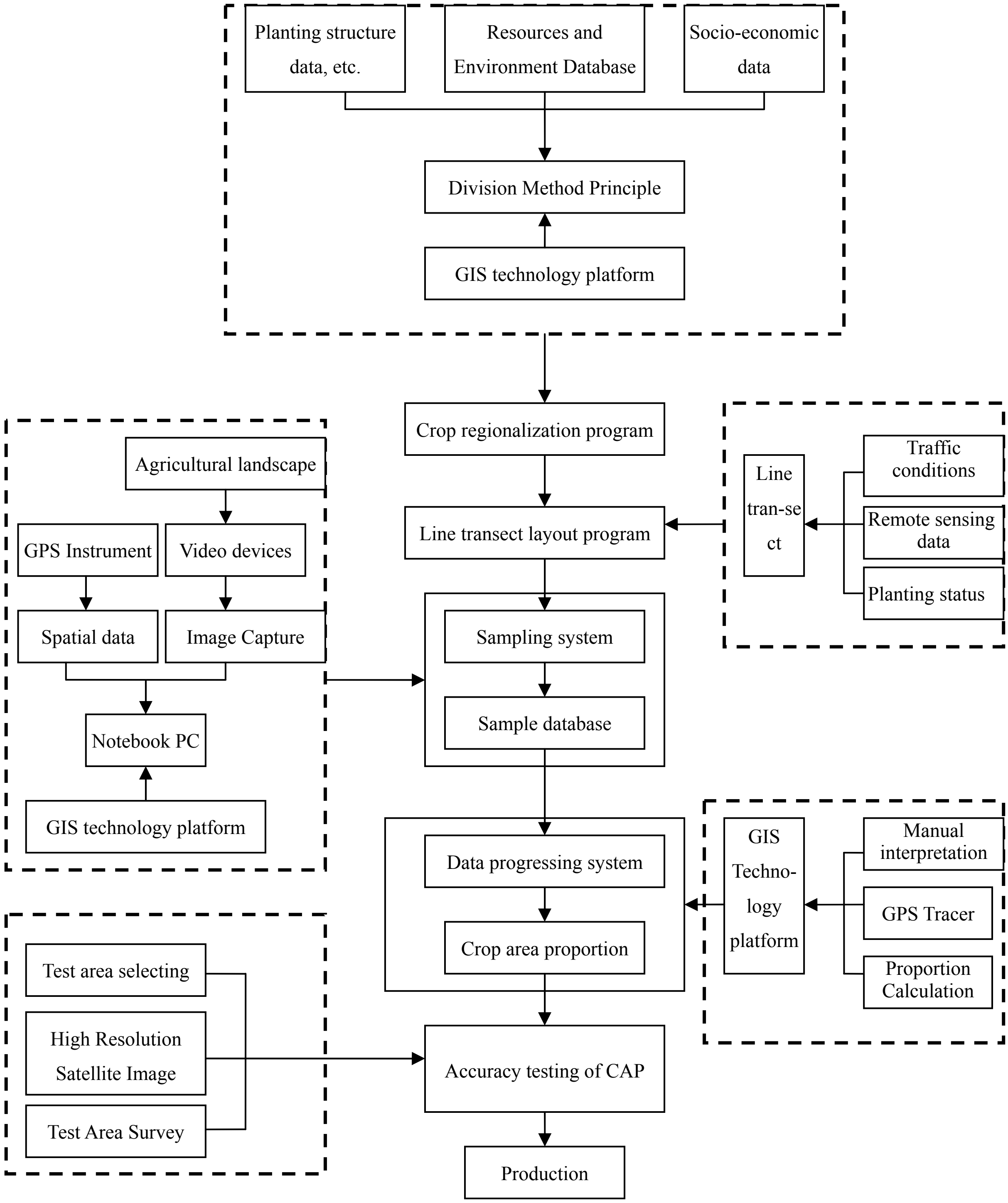

Figure 1. Technological schematic to obtain CAP based on the sampling technique

Figure 2. Sampling images

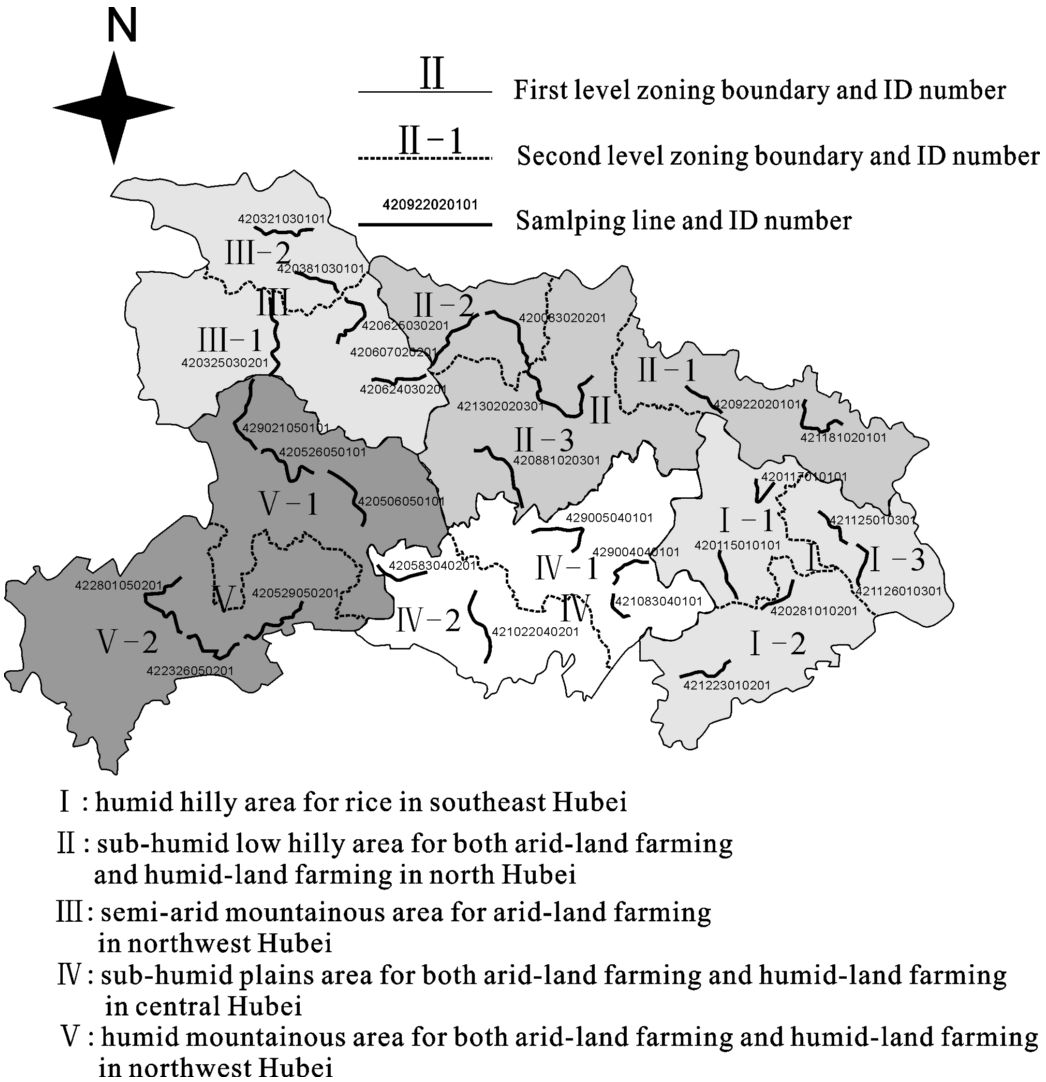

Figure 3. Crop planting structural zoning division in Hubei Provence and projections for sampling line layout