

Examination Method and Implementation for Field Survey Data of Crop Types Based on Multi-resolution Satellite Images

Yang Liu¹, Mingyi Du¹, Wenquan Zhu²,

¹ School of Geomatics and Urban Information, Beijing University of Civil Engineering and Architecture, Beijing, China

² College of Resources Science and Technology, Beijing Normal University, Beijing, China
liuyang@bucea.edu.cn

Abstract. In order to examine the accuracy of large amount of the field survey data with less accurate, an examination method based on multi-resolution satellite images was proposed in this paper. As there were so large amount of data, stratified random sampling was used to obtain effective samples. Firstly, vegetation index derived from low-resolution satellite images at different times has been adopted as analysis factor. And wave curve charts were drawn with the vegetation index. From those charts, the statistics law of wave curves for different crop types was recognized using for crop types' classification. Secondly, high-resolution satellite images were used to correct the area of crop types to get the final classification results. Finally, the accuracy of the field survey data can be calculated by comparing the original survey data with the final classification results. Moreover, for convenience using, a software has been developed according to the above examination method.

Keywords: Examination method and implementation, Stratified random sampling, Automatic processing, Crop types, Multi-resolution satellite images

1 Introduction

Planting area and yields of crops are the important basis for government's economic policy making. For a long time, the two kinds of traditional methods were used for the statistics of planting area of crops [1]. One way is the comprehensive statistical report coming from statistical and administrative units at various levels step by step [2]. Another way is to sample field survey of reported data. This method is suitable for a large degree on the discrete variables; frequency distribution was highly skewed socio-economic phenomena for investigation [3]. But, no matter whichever method was used, one step can not be omitted, that is examining the accuracy of large amount data of the field survey. Only in this way, the data can become the reliable argument for the decision analysis.

In recent years, with the extensive application of remote sensing technology, the study of estimation and examination of crop area using remote sensing techniques has made significant progress, and is steadily moving towards the direction of business

[4][5]. Compared with traditional ground survey and verification, the demand for manpower and costs are significantly reduced, using remote sensing image data [6]. However, from the level of remote sensing technology for carrying out, large-scale remote sensing examination of crop area is facing three problems, namely, precision, efficiency and cost problems [7] [8].

2 Research Process

This article aims to use multi-resolution remote sensing imaging technology (two kinds of resolution images: the high-resolution images, such as ASTER images, and the low resolution images, such as MODIS images). Utilizing their own advantage, combined with the traditional random, stratified sampling method [9], the examination scheme for field survey data of crop types was designed. In the scheme, stratified random sampling was used to ensure minimum sample size and stability, and the difference spectrum of low resolution remote sensing images was used to match the types of crops, and high-resolution remote sensing images were used to the visual identification of the types and area of crops. We try to solve the conflict of precision and cost, and try to provide methodological guidance for a large-scale examination for field survey data of the types and area [10].

The flow chart of technology route in this work, see Fig.1, is expressed as follows:

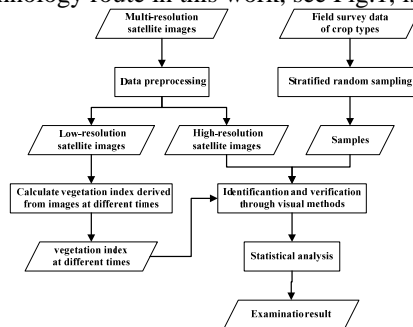


Fig. 1. This shows the flow chart of technology route.

In order to make the multi-resolution satellite images matched each other strictly and be in correspondence with the field survey data, the procedure of remote sensing data processing should be done. And stratified random sampling was used to obtain effective samples of field survey data of crop types. Vegetation index derived from low-resolution satellite images at different times has been adopted as analysis factor.

After all the data have processed, the identification and verification should be done through visual observation method. The statistics law of wave curves of vegetation index at different times was recognized using for crop types' classification of the samples. While, high-resolution satellite images were used to correct the classified crop area of samples to get the classification results. Finally, the accuracy of the survey data of crop types can be calculated by comparing the original survey data with the final classification results using statistical analysis methods.

In this paper, we focus on stratified random sampling in this work.

3 Stratified Random Sampling

The effectiveness of stratified random sampling can be evaluated along two dimensions. First, it might reduce the total quantity of measurements required. Second, it can do a reduction of measurements would imply cost savings [11].

3.1 Overview

In the sampling method, stratified sampling is one of the most effective methods. Compared with simple random sampling, stratified sampling has its several great advantages, such as the fewer number of samples, higher sampling precision and the lower cost. It is the effective way of large-scale statistical sampling surveys [12]. So stratified sampling was adopted in the paper.

Stratified sampling, also known as type sampling, is one of the most commonly used sampling techniques in practical work [13][14]. Firstly, overall sample is divided into several strata (groups) in accordance with certain rules in stratified sampling. Secondly, sampling was done within each stratum independently. The resulting sample is called stratified sample. Accordingly, the sample of each stratum is independent also [15]. Furthermore, if the sampling method of each stratum is simple random sample, the sample is called stratified random sampling. Thus the resulting samples are called stratified random samples.

3.2 Determination of Stratified Number

In practice, the number of strata can not exceed half of the sample size, because of ensuring that each stratum has at least two samples and having the need to calculate the standard deviation of each stratum in the method [16]. In this paper, we stratify the population into six strata.

3.3 Determination of Stratified Boundaries

It is usually to determinate the boundaries of strata in accordance with characteristics or features of population in the stratification [17]. However, when the obvious characteristics or features is not easily recognized, the general way to determinate the boundaries of strata is studying general distribution of some properties in population or studying a relative relationship of the variable X and the variable Y which are sub-features of population by means of some mathematical methods.

In this paper, the Optimal Stratification Method(also known as Accumulative Square Root Method) is used to determinate the boundaries of strata, which was proposed by Dalenius and Hodges [18], Firstly, a characteristic property of population should be determined, according to which population is stratified. Secondly,

population is sorted by the values of the property of population from small to large. Then, the sum of the values of the property of population is calculated, and the square root of the sum should be obtained [19]. Finally, population was divided into several strata by equal division method of the square root.

3.4 Calculation of Sample Size

There are two main steps in the calculation of sample size of each stratum as show in the below.

Computation of the total sample size. The total sample size calculations generally use the following:

$$n = \frac{\sum_{h=1}^L \frac{W_h^2 S_h^2}{w_h}}{\left(\frac{r\bar{Y}}{t}\right)^2 + \frac{1}{N} \sum_{h=1}^L W_h S_h^2} \quad (1)$$

In Equation (1), n is the total sample size, and h is the strata-specific variable which ranges from 1 to L . L is number of strata. W_h is the Weight of stratum h , which can be calculate by means of N_h (the population for stratum h) divide by the total stratified sample size N . S_h^2 is the population deviation of stratum h , which is calculated using Equation (2). And w_h is the sampling radio of stratum h , which is calculated using Equation (3). And r is the relative error limit (also as known the confidence interval) for (general set 95%). While, \bar{Y} is the population mean, and t is the percentile of the standard normal distribution ($z=0.05$ for 95% confidence).

S_h^2 is determined by:

$$S_h^2 = \frac{1}{N} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2 \quad (2)$$

Note that the value of sample i of stratum h is marked with a hat as y_{hi} , \bar{Y}_h is the population mean for stratum h . Other symbols are the same as Equation (1).

And w_h is determined by:

$$w_h = \frac{W_h S_h}{\sum_{h=1}^L W_h S_h} \quad (3)$$

Where, S_h is the population standard deviation of stratum h , namely the square root of S_h^2 . Also, other symbols are the same as Equation (1) or Equation (2).

Sample allocation in each stratum. For stratified sampling, the sample size of each stratum still need to determine after the total sample size is fixed. When doing the population estimation, the population variance which can be estimated is related not only to the variance of each stratum, but also to the sample size of each stratum. There are a lot of sample allocation methods in practice [20]. You can allocate the samples in accordance with the radio of between the sample size of each stratum and the sample size of population, or according to desired overall confidence interval with minimum total sample size.

The sample allocation of each stratum was done using the Optimal Sample Allocation, which was proposed by Neyman [21]. It is expressed mathematically as follows:

$$n_h = n \times \frac{N_h S_h}{\sum_{h=1}^L N_h S_h} \quad (n_h \text{ shall be as an integer}) \quad (4)$$

where, N_h is the sample size of stratum h . Also, other symbols are the same as Equation (1) or Equation (3).

However, the two problems should be noted in the Optimal Sample Allocation. There are:

- One is in the case of the sampling ratio $f = n / N$ is very large. This situation leads to S_h (the standard deviation of each stratum) is relatively large, and n_h (the sample size of each stratum) may be larger than N_h (the population for stratum h). At this situation, it required for 100% sampling to the population for stratum h .
- The other case is that n_h is less than 2 after the calculation using the Optimal Sample Allocation. The value of n_h needs to be set to 2 to reduce the impact of random error on the results and the stability when the sample size is 1.

4 Examination

The process of examination for field survey data of crop types can be mainly divided into the following four steps. Note that, the experiment of examination itself only use a small part of the test data.

4.1 Data

The entire test data used in this paper was shown in Fig. 2.



Fig. 2. This shows the experimental data in this work. From left to right is successively vegetation index at different times, high-resolution satellite images, field survey data and the overlay of the above three data which share the same region in the rectangular box.

We have collected the latest report information of field survey data of crop types of county-level, the multi-resolution satellite images which can cover the survey region in this step and prepare for the next step of Sampling.

The field survey data of crop types are Shape-file data, which should include the type and area property fields. High-resolution satellite data now is ASTER images. But if required, it can be higher resolution images. Vegetation index at different times were derived from MODIS time series data, which were formed as a multi-band image (namely one time corresponds to one band). And all test data was matched strictly

4.2 Sampling

Fig. 3 shows sampling result comparing with raw data, and the raw data was covered by the sampling data. So the visible parts of the raw data are not sampled.

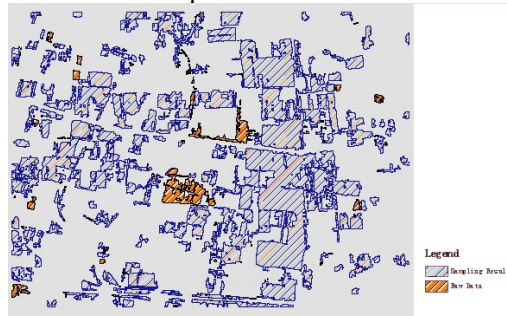


Fig. 3. This shows the sampling result comparing with raw data.

We also do the statistics to the sampling result which is convenient for customers to find out sampling information to decide whether to re-sampling, shown as the Table 1 below.

Table 1. Statistics of sampling data.

Index of Strata	Boundaries	Num of Population	Num of Samples	Standard Deviations	Means
1	100.00 - 890.13	109	2	119.42	388.07
2	890.13 - 1780.25	41	2	81.26	1212.19
3	1780.25 - 2670.38	22	2	61.56	2136.36
4	2670.38 - 3560.50	12	2	33.11	2974.99
5	3560.50 - 4450.63	2	2	14.76	4000.00
6	4450.63 - 3217100.00	181	166	273122.73	108427.62
Total is	100.00 - 3217100.00	367	176		

4.3 Crop Types' Identification and Areas' Correction

In this step, multi-resolution satellite images were used to determine crop types and area of the samples by visual observation method. So there are two factors as following:

Crop types' identification. Wave curve charts were drawn with the index of bands as X-axis and the value of vegetation index as Y axis, as shown in Fig. 4. There are 17 bands in this work.



Fig. 4. This shows the wave curve charts of vegetation index.

Fig. 5 show wave curve charts of four different positions. From those charts, the statistics law of wave curves for different crop types was identified using for crop types' classification.

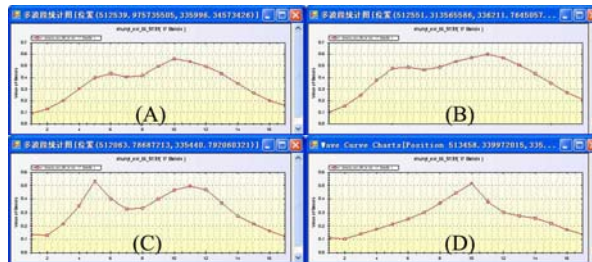


Fig. 5. This shows the schematic diagram of comparison with different wave curves.

Crop areas' verification. It provides a tool to draw polygons and calculate area of those. By intercomparison between high-resolution satellite images and samples of the same region, you can get the area of samples by measuring. The Area of all samples should be measured, which has been regarded as hypothetical real value.

4.4 Statistical Analysis

It needs to get error situation of the field survey data by doing statistics for result of type identification and area verification of samples.

As shown in Table 2, the number of qualified area, the number of qualified type and the number of qualified both of each stratum was obtained by statistics. Also, qualified rate of area, qualified rate of type, and overall qualified rate (area and type are both qualified) of samples were calculated by anti-deduction.

Table 2. Examination result.

Index Of Strata	Boundaries	Num of Population	Num of Samples	Num of qualified area	Num of qualified type	Num of qualified Both
1	100.00 - 890.13	109	2	2	1	1
2	890.13 - 1780.25	41	2	2	2	2
3	1780.25 - 2670.38	22	2	1	1	0
4	2670.38 - 3560.50	12	2	2	2	2
5	3560.50 - 4450.63	2	2	1	1	0
6	4450.63 - 3217100.00	181	166	83	81	23
Total is	100.00 - 3217100.00	367	176	91	88	28
Conclusion is	Qualified rate of area is 72.07%		Qualified rate of type is 56.63%		Overall qualified rate is 36.12%	

In statistics, we can find out the abnormal samples which are the types is right, but the rate of area change exceeding a certain threshold, as shown in Figure 6. In this works, the threshold was set to 10%. These abnormal samples should be report to the related departments of doing field survey to redo site verification.

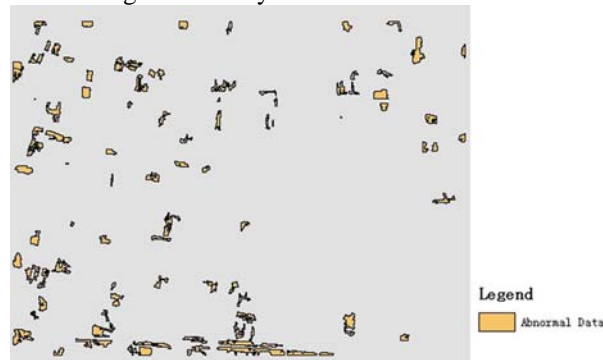


Fig. 6. This shows the abnormal samples.

5 System Implementation

An Examination System for Field Survey Data (ESFSD) has been developed based on the above examination method. The system (ESFSD) provides integrated and process-oriented functions, such as data processing, stratified random sampling, sample identification and verification, statistical analysis of samples, and so on. Using the system, you can easily complete the entire process of examination for field survey data of crop types.

The system (ESFSD) adopts C/S (client/server) software system structure, C# as the programming language and SQL Server 2000 as back database server. Using ArcEngine as development tools, the application program developed has been closely integrated with system by UG/Open MenuScript. Figure 7 presents the interface of ESFSD.

ESFSD also provides data management statistics information of examination result, and supply better decision supports for re-sampling of samples.

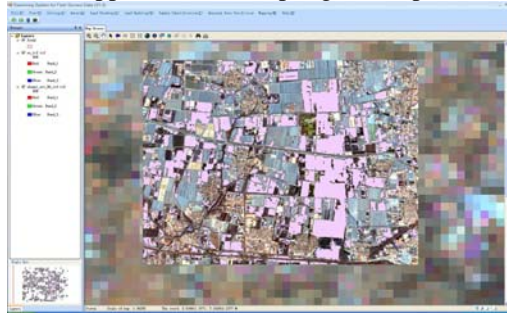


Fig. 7. This shows the system interface.

6 Conclusion

In order to examine the accuracy of large amount of the field survey data, an examination method based on multi-resolution satellite images was proposed in this paper. For reducing the number of data and cost in the examination, stratified random sampling was used to obtain effective samples. For determining crop types and area of the samples correctly and quickly, multi-resolution satellite images were used utilizing their own advantage. Low-resolution satellite images are used to identify the corps type of samples, and high-resolution satellite images were used to verify the area of samples. Finally, the accuracy of the field survey data can be calculated by comparing the original survey data with the final examination results by statistics. By the testing of actual data, the examination method is very effective to solve the confliction of precision and cost. And it also provides methodological guidance for a large-scale examination for field survey data of the types and area.

Moreover, for convenience using, the system (ESFSD) has been developed according to the above examination method. The results of system testing show ESFSD is comprehensive processing platform which has realized integrated and process-oriented functions for the examination. And it has strong robustness and stability.

Acknowledgments. The research performed is supported by Beijing Natural Science Foundation (No: 4102019), and is partially funded by Jurisdiction of Beijing Municipality under Grant No. PHR200906138, PHR200907127 and PHR 20070101. The work was performed under Beijing University of Civil Engineering and Architecture (No:100901405). We wish to thank Guoyin Cai for his careful proofreading of the manuscript.

References

1. B.F. Wu, Q.Z. Li: Crop Acreage Estimation Using Two Individual Sampling Frameworks with Stratification. *Journal of Remote Sensing*, 2004, Vol. 8, No. 6, pp. 551-569 (2004)
2. J.F. Wang, J.Y. Liu, and D.F. Zhuan, et al: Spatial Sampling Design for Monitoring the Area of Cultivated Land. *International Journal of Remote Sensing*, 2002, Vol. 23, No. 2, pp. 263- 284 (2002)
3. L.F. Li, J.F. Wang, and J.Y. Liu: Spatial Sampling Optimized Decision-making of National Land Remote Sensing Investigation. *Science in China Sec. D Earth Sciences*, 2004, Vol. 34, No. 10, pp. 975-982 (2004)
4. R.M. Narayanan, M.K. Desetty, and S.E. Reichenbach: Effect of Spatial Resolution on Information Content Characterization in Remote Sensing Imagery based on Classification Accuracy. *International Journal of Remote Sensing*, 2002, Vol. 23, No. 6, pp. 537-553 (2002)
5. S.B. Chen: *Bulletin of Wheat, Corn and Rice Yields Estimate Technique with Remote Sensing*. Chinese Science and Technology Press, Beijing (1993)
6. C.E. Woodcock, A.H. Strahler: The Factor of Scale in Remote Sensing. *Remote Sensing of Environment*, 1987, pp. 311-332 (1987)
7. H.Q. Liu: Sampling Method With Remote Sensing for Monitoring of Cultivated Land Changes on Large Scale. *Transactions of the Chinese Society of Agricultural Engineering*, 2001, Vol. 17, No. 2, pp. 168-171 (2001)
8. H.S. He, S.J. Ventura, and D.M. Mladenoff: Effects of Spatial Aggregation Approaches on Classified Satellite Imagery. *International Journal of Geographical Information Science*, 2002, Vol. 16, No. 1, pp. 93-109 (2002)
9. G.M. Foody: Status of Land Cover Classification Accuracy Assessment. *Remote Sensing of Environment*, 2002, pp. 185- 201 (2002)
10. X.F. Jiao, B.J. Yang, and Z.Y. Pei: Paddy Rice Area Estimation Using a Stratified Sampling Method with Remote Sensing in China. *Transactions of the Chinese Society of Agricultural Engineering*, 2006, Vol. 22, No. 5 pp. 105-110 (2006)
11. Y.K. Chen, R. Hu: The Theoretic Structure and Computerized Realization of Stratified Audit Sampling. *The Theory and Practice of Finance and Economics*, 2003, Vol. 24, pp. 78-80 (2003)
12. S.G. Wang, W.H. Chen, and L.D. Gao: *Probability Theory and Mathematical Statistic*, Science Press, Beijing (2000)
13. D.J. Brus, J.J. de Gruijter, J.W. van Groenigen: Designing Spatial Coverage Samples Using the K-means Clustering Algorithm. *Digital Soil Mapping: An Introductory Perspective*, Elsevier (2006)
14. F.J. Gallego, Stratified Sampling of Satellite Images with a Systematic Grid of Points, *ISPRS Journal of Photogrammetry and Remote Sensing*, 2005, Vol. 59, pp.369-376 (2005)
15. B. Minasny, A.B. McBratney, and D.J.J. Walvoort: The Variance Quadtree Algorithm: Use for Spatial Sampling Design. *Computers and Geosciences*, 2007, Vol. 33, pp. 383-392 (2007)
16. C. Kadilar, H. Cingi: A New Ratio Estimator in Stratified Random Sampling. *Communications in Statistics-Theory and Methods*, 2005, Vol. 34, No. 3, pp. 597-602. (2005)
17. Tetsuji Ohyam, J.A.Jimmy A. Doi, and Takashi Yanagawa: Estimating population Characteristics by Incorporating Prior Values in Stratified Random Sampling/Ranked Set Sampling. *Journal of Statistical Planning and Inference*, 2008, Vol. 138, No. 12, pp. 4021-4032 (2008)
18. C. Kadilar, H.Cingi: Ratio Estimators for the Population Variance in Simple and Stratified Random Sampling. *Applied Mathematics and Computation*, 2006, Vol. 173, No. 2, pp. 1047-1059 (2006)

19. C.C. Wu, R.C. Zhang: Empirical Likelihood Inference Under Stratified Random Sampling in the Presence of Measurement Error. *Acta Mathematicae Applicatae Sinica, English Series*, 2005, Vol. 21, No. 3, pp. 429-440 (2005)
20. S. E. Wright, R. B. Noble, and A. J. Bailer: Equal-Precision Allocations and Other Constraints in Stratified Random Sampling. *Journal of Statistical Computation and Simulation*, 2007, Vol. 77, No. 12, pp. 1081-1089 (2007)
21. V. Bosch, R. Wildner: Optimum Allocation of Stratified Random Samples Designed for Multiple Mean Estimates and Multiple Observed Variables. *Communication in Statistics-Theory and Methods*, 2003, Vol. 32 No. 10, pp. 1897-1909 (2003)