1. Publication Nº
INPE-2489-PRE/172

2. Version

3. Date
July, 1982

5. Distribution
☐ Internal ☒ External
☐ Restricted

4. Origin: *DSE/DES*
Program: *AREA*

6. Key words - selected by the author(s)
TWO PHASE SAMPLING
SEGMENT DIMENSIONS

7. U.D.C.: *528.711.7:633.61*

8. Title *INPE-2489-PRE/172*

ESTIMATION OF THE SUGAR CANE CULTIVATED AREA FROM LANDSAT IMAGES USING THE TWO PHASE SAMPLING METHOD

10. Nº of pages: *05*

11. Last page: *04*

12. Revised by
Nívea Teixeira Dias

9. Authorship *Carlos Alberto Cappelleti*
Francisco José Mendonça
David Chung Liang Lee
Yosio Edemir Shimabukuro

Responsible author

13. Authorized by
Nelson de Jesus Parada
Director

14. Abstract/Notes

A two phase sampling method and the optimal sampling segment dimensions are developed for the estimation of the sugar cane cultivated area. This technique employs visual interpretations of LANDSAT images and panchromatic aerial photographs considered as the ground truth. The estimates, as a mean value of 100 simulated samples, represent 99.3% of the true value with a CV of approximately 1%; the relative efficiency of the two phase design was 157% when compared with a one phase aerial photographs sample.

15. Remarks *This paper was accepted for presentation at the Sixteenth International Symposium on Remote Sensing of Environment, Buenos Aires, Argentina, June 2-9, 1982.*

ESTIMATION OF THE SUGAR CANE CULTIVATED AREA FROM LANDSAT IMAGES USING THE TWO PHASE SAMPLING METHOD*

Carlos Alberto Cappelletti
Francisco José Mendonça
David Chung Liang Lee
Yosio Edemir Shimabukuro

Instituto de Pesquisas Espaciais - INPE
Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq
Caixa Postal 515, 12200 - São José dos Campos, SP, Brasil

ABSTRACT

A two phase sampling method and the optimal sampling segment dimensions are developed for the estimation of the sugar cane cultivated area. This technique employs visual interpretations of LANDSAT images and panchromatic aerial photographs considered as the ground truth. The estimates, as a mean value of 100 simulated samples, represent 99.3% of the true value with a CV of approximately 1%; the relative efficiency of the two phase design was 157% when compared with a one phase aerial photographs sample.

1. INTRODUCTION

In this paper a statistical system to estimate the sugar cane (Saccharum spp.) cultivated area in a subregion of São Paulo State, Brazil, is presented. The region under study is known as the Great Region of Jaú, located in the central part of the state, within the parallels 22°00' and 23°00' South, and meridian 48°00' and 49°00' West, covering 5046 km².

For this region, the sugar cane acreage was determined by two different approaches. In the first approach, Koffler at al., (1980) made interpretations of panchromatic aerial photographs in the scale of 1:35000 and 1:45000, complemented with a rigorous field control. This information is considered as the ground truth.

In the second approach, Mendonça et al. (1981) used LANDSAT images in the scale of 1:250000 and visual interpretation, associating the spectral variations of the crop with temporal variations in different satellite pass. Channels 5 and 7 were used.

With this basic information, a sampling experiment was designed to estimate the sugar cane acreage and measure the error rate in relation to the ground truth. The applied sampling method was a two phase sampling with regression estimates.

* Presented at the Sixteenth International Symposium on Remote Sensing of Environment, Buenos Aires, Argentina, June 2-9, 1982.

2. STUDY OBJECTIVES

a) To determine segments size in order to have the maximum correlation between LANDSAT data and ground truth.

b) To estimate the sugar cane acreage with a sampling design using LANDSAT data and ground truth and considering the cost of each source of data.

3. SEGMENTS SIZE AND SAMPLING METHOD

a) Determination of segments size

With a grid of size 1 cm by 1 cm (corresponding to 2.5 km by 2.5 km) applied on a composition of LANDSAT images and on a mosaic of aerial photographs, the number of points with sugar cane in each unitary segment was determined.

From that information, a subarea of 60 km x 37.5 km was selected and a uniformity trial (Federer, 1967) was designed. The response variable was the number of hectares cultivated with sugar cane by segment of 6.25 km².

Ten different types of segments were considered:

Type	Size	No. of Segments
1 x 1	2.5 x 2.5 = 6.25 km²	360
1 x 2	2.5 x 5.0 = 12.50 km²	168
1 x 3	2.5 x 7.5 = 18.75 km²	120
1 x 4	2.5 x 10.0 = 25.00 km²	72
2 x 2	5.0 x 5.0 = 25.00 km²	84
2 x 3	5.0 x 7.5 = 37.50 km²	36
2 x 4	5.0 x 10.0 = 50.00 km²	36
3 x 3	7.5 x 7.5 = 56.25 km²	40
3 x 4	7.5 x 10.0 = 75.00 km²	24
4 x 4	10.0 x 10.0 = 100.00 km²	18

The criterion of selection was the maximum correlation between the variables "response to the LANDSAT images interpretation", X, and "ground truth", Y.

b) Sampling design

A two phase sampling design with regression estimates was applied combining the two sources of information.

This design takes advantage of the correlation between X and Y and the cost ratio for collecting the data in each variable. The sample selection was done in two steps. In the first step, a relatively large sample in the less expensive variable, X in this case, was selected, and in the second step a small sample in the other variable that is more expensive to be observed, Y, was selected. This bivariate information permits the calibration of the information from X.

When the correlation between X and Y is sufficiently large, a reduction in the estimator variance and in the sample size is significant in comparison with a single phase sampling design on Y alone.

Before applying the sampling design, the region of study was redefined in order to eliminate parts of it where sugar cane is not cultivated, and also to eliminate some geographical accident as well as some incomplete marginal segments.

The theoretical considerations of the sampling method applied in this study are developed in Cochran (1963), Loestsch and Haller (1973) and Jessen (1978), among other authors.

4. ANALYSIS AND ESTIMATION

The sample size in the two phases was set to achieve a sugar cane acreage estimate within 5% of the corresponding complete acreage evaluation with a 95% confidence level at a minimum cost.

To calculate k and n, the number of segments in the first and second phase, respectively, it was used σy and ρ calculated from the complete enumeration of available ground truth data. When this information is not available, it is necessary to select a pilot sample.

With the obtained values of k and n, it was simulated a sequence of one hundred samples in two phases by means of a simple random selection in each phase, being the second phase sample a simple random subsample from the first one.

The simulation produced a sequence of values of the random variables $\hat{Y}_R$, estimate of the sugar cane acreage; $\hat{V}$ ($\hat{Y}_R$), variance of the estimate; and of $D = \hat{Y}_R - \alpha$, where α = 133888 Ha is the total number of hectares cultivated with sugar cane from the ground truth complete enumeration. This value is assumed without error.

Finally, the simulated data were analysed statistically.

5. RESULTS

a) The selected segment size in the uniformity trial was the one with dimension 2 x 3, corresponding to 5.0 km x 7.5 km = 37.50 km^2 or 3750 ha. The maximum correlation coefficient between X and Y was r = .82 for this size, being r = .66 and .73 for segments of size 6.25 km^2 and 100.00 km^2 respectively.

b) The sampling frame, after being redefined (3 b), had 86 segments with an area of 3082 km^2, 16% of them incomplete, but with an area greater than one half of the complete one.

c) The standard deviation of Y was σy = 777.46 ha and the correlation coefficient ρxy = .82. The cost ratio considered $c_1:c_2$ was 1:13, where c_1 and c_2 are the unit cost of observing one Xi (in phase one) and one Yi (in phase two), respectively. The sample sizes were k = 58 and n = 11 for a relative variance of 5% with a confidence level of 95% and a minimum cost of C = 201 monetary units. The regression equation of ground truth and LANDSAT data was Y = 442 + .69X.

d) The results from the analysis of the simulated data were M ($\hat{Y}_R$) = = 132889 ha (99.3% of the ground truth value). SD ($\hat{Y}_R$) = 1013 ha (SD% $\simeq$ 1%). Range ($\hat{Y}_R$) = max $\hat{Y}$ - min $\hat{Y}$ = 29393 ha. $\bar{D}$ = M ($\hat{Y}_R$) - α = = -999 ha. The confidence interval for the sugar cane acreage at the 95% level was (130963; 134915) = (97.7%; 100.1%), and for the mean difference $\mu_{\bar{D}}$ at the 95% level was (-3025; 1027).

e) The relative efficiency of the two phase design with respect to a simple phase was 157%, what means a gain of 57% when LANDSAT data and ground truth data are jointly used in a regression estimator. Finally, the total acreage estimate from LANDSAT data only, with a sample of size k = 58, was $\hat{X}$ = 140266 ha (104.8% of the ground truth value). The 95% confidence interval was (138924; 141608) ≃ (103.8%; 105.8%). The regression estimate has a tendency to underestimate the true value, 99.3% of it, while using only LANDSAT data the estimate represents 104.8%. The estimation with aerial photograph data (ground truth data in this study) is not being taken into consideration because of high collecting cost. This is the main reason for implementing a sampling design that minimizes the use of that type of information.

6. CONCLUSIONS

The sampling design applied in this study shows the benefits, in cost and precision, relative to a complete inventory through a more expensive method of data collecting like aerial photography.

The authors understand that these results are limited to a small region like the one selected for this study. Besides that, much work has yet to be done in this sampling approach.

7. REFERENCES

Cochran, W.G. Sampling techniques. 2 ed. New York, Wiley, 1963. 413 p.

Federer, W.T. Experimental design, theory and application. New Delhi, Oxford, 1967. 544 p.

Jessen, R.J. Statistical survey techniques. New York, Wiley, 1978. 520 p.

Koffler, N.F.; Chiarini, J.V.; Gavalli, A.C.; Nogueira, F.P. Inventário canavieiro com auxílio de fotografias aéreas – a Grande Região de Jaú no ano safra 1979/80. 2(4): 1-22, ago. 1980.

Loestsch, F.; Haller, K.E. Forest inventory. 2 ed. BLV Verlags Gesellschaft Munchen Bern Wien, 1973. 436 p.

Mendonça, F.J.; Lee, D.C.L.; Tardin, A.T.; Shimabukuro, Y.E.; Chen. S.C.; Lucht, L.A.M.; Moreira, M.A.; Lima, A.M. de; Maia, F.C.S. Levantamento da área canavieira do Estado de São Paulo, utilizando dados do LANDSAT - ano safra 1979/80. São José dos Campos, INPE, mar. 1981. (INPE-2021-RPE/288)