FOREST INVENTORY USING MULTISTAGE SAMPLING WITH PROBABILITY PROPORTIONAL TO SIZE

David C. L. Lee; Pedro H. Filho, Yosio E. Shimabukuro
CNPq - Instituto de Pesquisas Espaciais - INPE

Oswaldo R. de Assis, José S. de Medeiros
Instituto Brasileiro de Desenvolvimento Florestal - IBDF

Comission VII

ABSTRACT

The objective of this study was to develop and evaluate a multistage sampling technique, with probability proportional to size, for forest volume inventory using remote sensing data. The study area (12 x 12 km^2) is located in the Southeastern Brazil. LANDSAT-4 digital data of September 11, 1982 of the study area were used in the first stage for automatic classification of reforested areas. Four classes of pine and eucalypt with different tree volumes were classified utilizing a maximum likelihood classification algorithm. Color infrared aerial photographs (1:10,000) obtained in July, 1982 were utilized in the second stage of sampling. In the third stage (ground level) the timber volume of each class was determined. The total timber volume of each class was expanded through a statistical procedure taking into account all the three stages of sampling. This procedure resulted in an accurate timber volume estimate with a smaller number of aerial photographs and reduced time in field work.

1. INTRODUCTION

In the last years, Brazilian government has been reducing oil importation with some programs for gradual substitution of this row material. Reforestation is one of the alternatives for traditional energy sources, contributing significantly to solve this problem.

Brazil has a reforested area of about 4 million hectares, which may supply 14 billion and 800 million liters of ethanol - the necessary percentage for gasoline mixtures per year. Therefore, it is necessary to know accuratly and quickly the timber volume in the artificial forest population.

Several conventional methods of forest inventory, using systematic and random sampling in the field, have been used frequently, but these methods are not quite precise because it is difficult to obtain the information about the planting homogeneity reforested land with a low cost and in a short period of time. Therefore aerial photographs and satellite images may be used to evaluate the forest quality.

The present work shows a methodology, using remote sensing data in multistage sampling, to inventory the planted forests; also it represents a step toward the investigation of new and modern technologies, which provide fast, accurate data for solving forest problems.

2. STUDY AREA AND DATA SOURCES

The study area, Mogi-Guaçu, is located in the humid subtropical climate zone of São Paulo State, Brazil with the coordinates at 22°15'S and 47°10'W

(Figure 1). This area, approximately 144 km², includes the Campininha pine experimental station of the Forestry Institute of São Paulo State (IFSP) and the Santa Terezinha eucalypt plantation of the Champion Celulose & Paper Company (CCP). The major *Pinus* species in Campininha are *P. elliottii*. Other species such as *P. taeda*, *P. caribaça*, *P. balamenois*, *P. oocarpa*, and *P. palustris* are also planted in small areas. The prominent *Eucalyptus* species in Santa Terezinha are *E. grandis* and *E. saligna*.

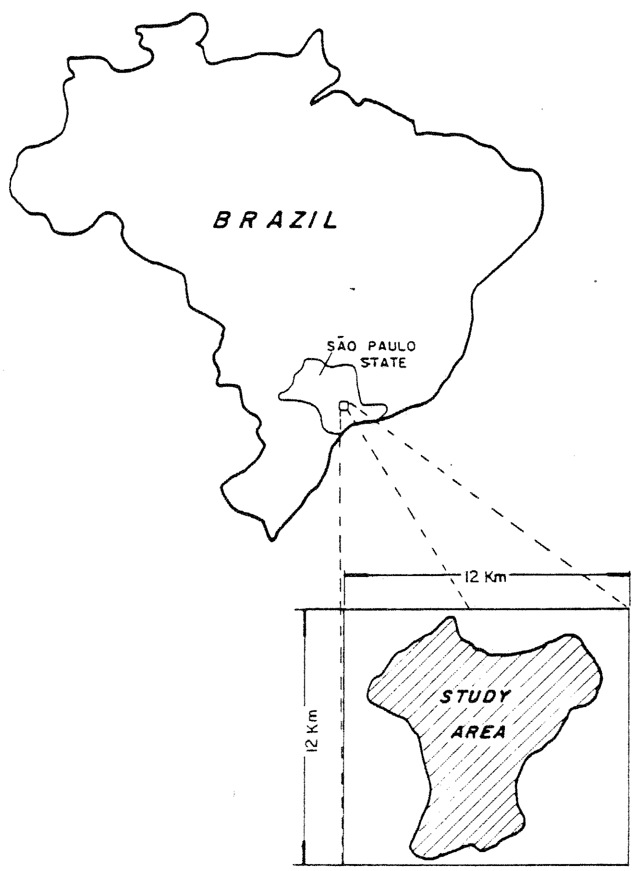

Figure 1 - Location of the study area.

LANDSAT multispectral scanner CCT's acquired on September 11, 1982 with bath/row annotation of 220/75 were processed with the G.E. IMAGE-100 system to classify *Pinus* and *Eucalyptus* species and their age groups.

Field data and forest cover maps were provided by IFSP and CCP and used for supervised classification. Spot field checks were also carried out.

2. METHOD AND MATERIALS

A multistage survey using LANDSAT MSS digital output, color infrared (CIR) aerial photograph and field measurement data was conducted in the artificially reforested area in the Southeastern Brazil. This technique, demonstrated by Langley (1975) as a sound method of obtaining estimates, is based on the integration of multistage inventory design and variable probability sampling with replacement.

2.1 - MULTISTAGE SAMPLING

This survey used a three-stage variable probability design where sample units are selected based on the probability proportional to size (PPS). In this study, the size was the percentage of reforested area estimated within the progressive subsampling units.

A generalized model for a three-stage estimator presented by Langley (1975) is:

$$V_t = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{P_i\ n_i} \sum_{j=1}^{n_i} \frac{1}{P_{ij}\ t_{ij}} \sum_{k=1}^{t_{ij}} \frac{V_{ijk}}{P_{ijk}}, \qquad (1)$$

in which

V_t is the estimated total volume of timber in the population;

V_{ijk} is the measured volume in the third stage unit;

P_i is the probability of drawing the i^{th} first stage unit;

P_{ij} is the conditional probability of drawing the j^{th} second stage unit given the i^{th} first stage unit;

P_{ijk} is the conditional probability of drawing the k^{th} third stage unit given in the first and second stage units which have been drawn;

m, n_j, t_{ij} are the sample sizes in the first, second, and third stages, respectively.

The unbiased estimate of sample variance can be obtained from the estimates at the first stage (Langley, 1975) using the following equation:

$$Var(V_t) = \frac{1}{m(m-1)} \left[\sum_{i=1}^{m} \left(\frac{V_i}{P_i}\right)^2 - m\ V_t^2 \right], \qquad (2)$$

in which V_i is the estimated timber volume in the i^{th} unit and the other terms are as defined earlier.

2.2 - LANDSAT IMAGE

Multispectral scanner digital data from the LANDSAT-4 image obtained on September 11, 1982 with path/row annotation of 220/75 were computer-aided analyzed using the G.E. IMAGE-100 system. Firstly, the study area (12 km x 12 km) was enlarged at a scale of 1:45,000 on the video monitor. Then a supervised classification scheme (maximum likelihood classification algorithms - MAXVER (Velasco et alii, 1978)) was used with the aid of aerial photographs and ground information. For this classifier a number of training sites which represented the various ground types on the LANDSAT image were selected. The objective of this classification was to generate classes that relate directly to stocked pine or eucalypt planted area. Finally four major forest classes were classified and used to reduce the variability of the population. These four forest classes were used to allocate samples for photointerpretation and ground data collection.

The computer-aided classification result was an alphanumeric thematic map at a scale of 1:10,000. The 144 km² study area was divided into primary sample units (PSU's) of about 493 x 519 m². Percentage of reforested area within each PSU was determined from the computer printout. The estimated percentages (x_i) and the cumulative totals (ΣX_i) of each reforested class were then listed. A sample of PSU's was randomly selected with probability proportional to the estimated percentage of the reforested area. For the total volume estimated expansion the probability of each selected PSU was $P_i = x_i/\Sigma X_i$. This procedure leads to a self-weighing estimator since each percentage in area has an equal chance of being selected in the sample.

Figure 2 shows how selected primary sampling units may be subsampled with progressively smaller sampling units in secondary and tertiary stages.

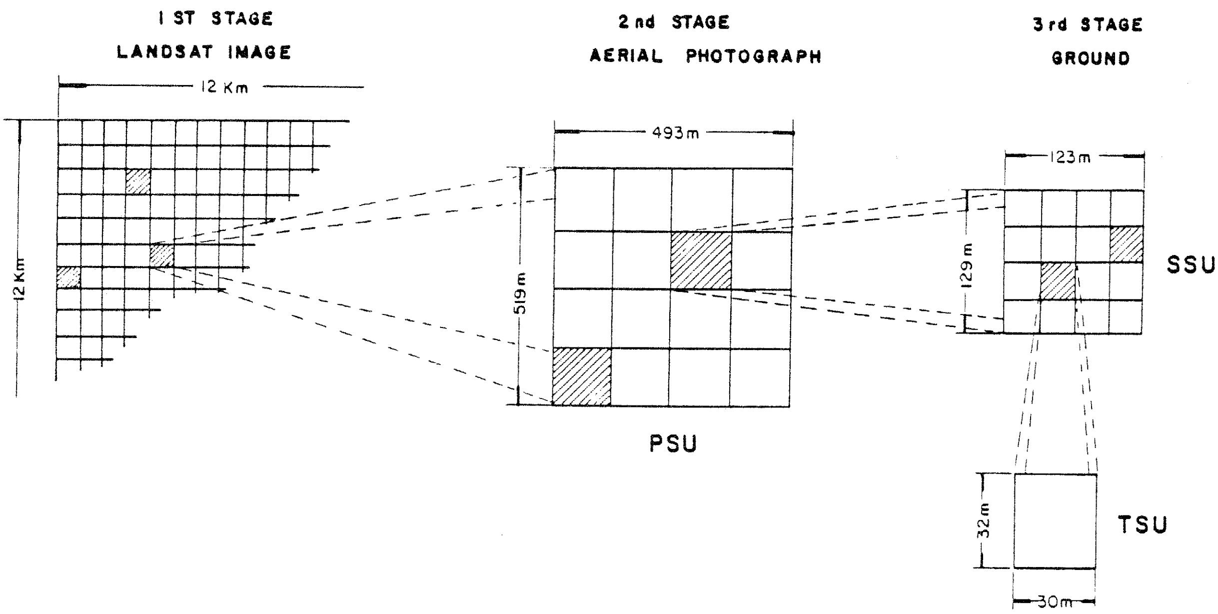

Figure 2 - Diagram of sampling design showing the selection of subsamples at three stages.

2.3 - AERIAL PHOTOGRAPH

Color infrared (CIR) photographs (Kodak Aerochrome Nº 2443) of the entire study area, obtained in July, 1982 by a RC-10 Camera at the scale of 1:10,000, were visually interpreted. The percentage of forest crown cover within each PSU was estimated and then the data were tabulated. These selected PSU's were precisely plotted on the appropriate CIR photo transparencies, and each PSU was then subdivided into 16 units (123 x 129 m² each) that defined the secondary sampling unit (SSU). These SSU's were interpreted for forest cover occupancy in percent and selected subsample with PPS from the second list.

The aerial photographs of the selected SSU's were enlarged 4 times and then each SSU was subdivided into 16 units (31 x 32 m² each) for tertiary sample selection (TSU). In that way, the crown density was evaluated and the tertiary subsample as before was selected with PPS.

2.4 - FIELD MEASUREMENTS

In the third stage, field work was carried out to determine the stand timber volume of the selected TSU. The mean volume per hectare from the sampled TSU was estimated using the Strand's Vertical Line Sampling method (Husch et al.,1971 and Loetsch et al., 1973). The formula is

$$V = \Sigma \frac{1}{10} d_j^2 \ . \ F,$$

in which

V is estimated stand mean volume in m³/ha;

d_j is diameter at breast height in cm of a tree enumerated in the vertical line sample;

F is form factor.

The TSU's volumes were then calculated and the total stand timber volume of each class was expanded back through the use of a statistically defined equation (Formula 1) taking into account all the three stage of sampling.

3. RESULTS AND DISCUSSION

In the first stage, the reforested areas were classified by computer-aided analysis in the following classes:

PINE-1: Pine species other than *P. elliottii*.
PINE-2: *Pinus elliottii*.
EUCA-1: *Eucalyptus spp.* from 8 months to 3 years.
EUCA-2: *Eucalyptus spp.* over 3 years.

Theoretically, sample size and allocation should take into consideration the effect of the sampling variance of the technique employed. The necessary number of sample size for the first stage was determined from the results of the computer-aided analysis presented on the alphanumeric thematic map. According to the coefficient of variation of each class and for an expected error of 20 percent, 6, 4, 10, and 7 units were required to the classes PINE-1, PINE-2, EUCA-1 and EUCA-2, respectively (Table 1).

TABLE 1

THE CALCULATED SAMPLE SIZE FOR THE FIRST STAGE

CLASS	C.V. %	EE %	Nº OF PSU's REQUIRED
PINE-1	50,7	20	6
PINE-2	42,3	20	4
EUCA-1	64,2	20	10
EUCA-2	54,7	20	7

For the classes EUCA-1 and EUCA-2 only 4 PSU's were used due to the monetary and time constraints for the field work. Two SSU's and two TSU's were selected for the second and third stage subsampling, respectively. For the pine classes the minimum sample size of 2 was used at each progressively subsampling stage. Therefore, a total of 48 TSU's were evaluated in the field. With the field data, it was possible to expand the total volume of each class.

For the entire study area, the total stand timber volume per class of pine and eucalypt was estimated and the data are summarized in Table 2. The results were obtained with the estimated sample errors of 11.27%, 4.78%, 42.87% and 13.58%, all with one standard derivation for Pinus-1, Pinus-2, EUCA-1, and EUCA-2, respectively. With the exception of EUCA-1, all fall well below the acceptatable sampling errors of 20 percent for the classes. These volume figures lead to the calculated stand timber volume per hectare, when the areas of classes were provided by forest cover map (Table 3).

TABLE 2

SAMPLING ERROR OF THE ESTIMATED TOTAL STAND TIMBER VOLUME

CLASS	TOTAL TIMBER VOLUME (m^3)	STANDARD ERROR (m^2)*	SAMPLING ERROR %
PINE-1	42,313.61	± 4,767.81	11.27
PINE-2	273,853.42	± 13,080.00	4.78
EUCA-1	12,972.56	± 5,561.41	42.87
EUCA-2	257,102.62	± 34,921.81	13.58

* one standard variation from the estimated volume.

TABLE 3

ESTIMATE OF AVERAGE STAND TIMBER VOLUME PER HECTARE

CLASS	TOTAL TIMBER VOLUME (m^3)	AREA * (ha)	AVERAGE TIMBER VOLUME (m^3/ha)
PINE-1	42,313.61	253.99	166.60
PINE-2	273,853.42	1,495.12	183.17
EUCA-1	12,972.56	323.42	40.11
EUCA-2	257,102.62	2,029.36	150.71

* data provided by forest cover map.

There was only about 0.1 percent forest area measured by an optical dendrometer on the ground at the 48 plots (TSU's). This sampling system used to estimate the total timber volume (per class) was very successful with only a minimum sample size. Therefore variable probability sampling based on a LANDSAT interpretation was demonstrated to be highly efficient for estimating reforested stand timber volume.

4. CONCLUSION AND RECOMMENDATIONS

It was possible to do a multistage sampling with the selection of sampling units based on the variable probability established from the interpretation of LANDSAT imagery, CIR aerial photographs and field measurements. LANDSAT MSS digital data can be effectively used to provide estimates of reforested land cover classes, that is:

PINE-1: Pine species other than *Pinus elliottii*.
PINE-2: *Pinus elliottii*.
EUCA-1: *Eucalyptus spp*. from eight months to three years.
EUCA-2: *Eucalyptus spp*. over three years.

For the third stage, an advantage by using the Strand's survey was the accurate and less time consuming procurement of ground data, i.e. to locate a plot (TSU) it is not necessary to delineate exactly its corners and boundaries, and within the plot four measurements were made to obtain an average volumetric result.

Although this study was not a perfect work with the optimal sample design, we demonstrated how to deal with the problems that might arise operationally as well as statistically. Techniques are developed for extracting better information from LANDSAT MSS data by computer-aided methods, and we will be able to inventory large reforested areas more rapidly and with a minimum field work. With this primary study done, we believe that a workable resources information system for forestry would be established. It is to be attempted that the data from this system would be provided to answer specific questions asked by management planners and policymakers. At that point, the system could be queried and answers returned almost instantaneously as to the quantity and distribution of forest resources in any part of the country.

Although the results of this study indicated that LANDSAT data and aerial photographs can be effectively used to inventory reforested timber volume, additional studies are necessary to determine: (1) the optimum size of sampling unit and the number of units required for each stage; (2) the stratification of heterogeneous study area to improve the LANDSAT classification results for a more efficient stratification of cover types; (3) large-scale air photos may be used to replace expensive ground sampling for timber volume estimates.

REFERENCES

Langley, P.G. Multistage variable probability sampling: theory and use in estimating timber resources from space and aircraft photography. Ph.D. Dissertation, University of California, Berkeley, 1975.

Velasco, F.R.D.; Prado, L.D.C.; Souza, R.G.M. Sistema MAXVER, Manual do Usuário. São José dos Campos, INPE, 1978 (INPE-1315-NTI/110).

Husch, B.; Miller, C.I.; Beers, T.W. Forest Mensuration. Second edition, John Wiley & Sons Inc., 1971.

Loetsch, F.; Zöhrer, F.; Haller, K.E. Forest inventory. BLV Verlagsgesellschaft mbH, München, 1973. Volume II.