AUTOMATIC CLASSIFICATION OF REFORESTED PINUS SPP. AND EUCALYPTUS SPP. IN MOGI-GUAÇU, S.P., BRAZIL USING LANDSAT DATA

Y. E. Shimabukuro, P. F. Hernandez, N. F. Koffler and S. C. Chen

Instituto de Pesquisas Espaciais (INPE)
Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)
12.200 - São José dos Campos, S.P. - Brazil

ABSTRACT

Single date Landsat CCTs were processed, by Image-100 to classify *Pinus* and *Eucalyptus* species and their age groups. The study area Mogi-Guaçu is located in the humid sub-tropical climate zone of São Paulo State. Prominent *Pinus* species are *P. elliottii* and *P. taeda*. *Eucalyptus* species are *E. alba* and *E. saligna*. The ages of these tree types range from 8 months to 20 years. The study area was divided into ten preliminary classes and feature selection algorithm was used to calculate Bhattacharyya distance (B-distance) between all possible pairs of these classes in the four available channels. The classes having B-distance values less than 1.30 were grouped together resulting in four classes: 1) class PE - *P. elliottii*, 2) class PO - *Pinus* species other than *P. elliottii*, 3) class EY - *Eucalyptus* spp. under two years, and 4) class EO - *Eucalyptus* spp. more than two years. For classification, single cell signature acquisition option was used. The low percentages of correct classification for class PO was due to its relatively small area (a total of 1.44 km^2) while for class EO error was caused by the similar signature responses of cerrado, gallery forest and heterogenous *P. elliottii* plots in the study area. The percentages of correct classification ranged from 70.9% to 94.12%. Comparisons of acreage estimated from the Image-100 with ground truth data showed agreement. The Image-100 percent recognition values for the above four classes were 91.62%, 87.80%, 89.89% and 103.30%. Greater precision of area estimation could have been obtained if the ground truth data based on the seedling area at planting time had been up-dated.

1. INTRODUCTION

Landsat is one of the potential tools which is meeting the increasing demands of government agencies and private industry for timely and precise inventories of forestry data. In the past few years various researchers have tested the possibility of identifying forest plant communities using Landsat data by conventional or automatic classification techniques. Sayn-Wittgenstein (1972) used a color composite of MSS band 4, 5 and 6 to separate coniferous forest (mostly spruce), hardwood (mostly poplars) and areas covered by willow and alder. López-Cuervo (1973) visually analyzed Landsat enlarged images and identified *Pinus* and *Eucalyptus* at the generic level according to their tonal differentiations. Gimbarzevsky (1974) also confirmed the separability of coniferous and deciduous forest using Landsat imagery. Lee et al. (1974) observed that

Landsat imagery could be used to distinguish forested and non-forested lands. Difficulties were encountered in separating coniferous species, even though age difference was recognized. Kalensky and Scherk (1975) reported that overall Image-100 classification accuracies for coniferous forest, deciduous forest and non-forested land ranged from 67% to 81% for single date imagery while for multidate imagery consistently above 80%.

The objective of this study was to classify Pinus and Eucalyptus species and their age groups, using the Image-100 automatic classification system.

2. STUDY AREA AND DATA SOURCES

São Paulo State contributes more than 50% of the paper and cellulose produced in Brazil. The state has an artificially reforested area of 6,144 km^2 predominantly in pine (22.15%) and eucalypt (76.48%). The study area Mogi-Guaçu is located at 22°15' S and 47°10' W (Fig. 1) and is a representative of pine and eucalypt plantations common to the region. This area includes the Campininha pine experimental station of the Forestry Institute of São Paulo State (IFSP) and the Santa Terezinha eucalypt plantation of the Champion Cellulose & Paper Company (CCP). The major Pinus species in Campininha are P. elliottii and P. taeda. Other species such as P. caribaea, P. bahamensis, P. oocarpa and P. palustris are also planted in small areas. The prominent Eucalyptus species in Santa Terezinha are E. alba and E. saligna. The tree age range from 8 months to 20 years.

Landsat multispectral scanner CCTs of September 13, 1975, orbital number 178, were processed for the classification of Pinus and Eucalyptus species and their age groups by the Image-100 system.

"Ground truth" information and forest cover maps (Fig. 2) were provided by IFSP and CCP and used for supervised classification and accuracy analysis. Spot field checks to verify the "ground truth" were also carried out.

3. IMAGE-100 ANALYSIS PROCEDURES

3.1 FEATURE SELECTION

Prior to the supervised signature extraction and classification, the CCTs were corrected radiometrically. A preliminary analysis of the study area was carried out through four B & W imagery and a false color composite displayed on the image monitor at an approximate scale of 1:82,000. The following ten preliminary classes were derived for the Image-100 study.

PE_a - P. elliottii over 13 years.

PE_b - P. elliottii under 13 years.

OP - Pinus species other than P. elliottii and P. taeda.

PT - P. taeda.

$E_{8mo. - 2yrs.}$ - Eucalyptus spp. from 8 months to 2 years.

$E_{2yrs. - 4yrs.}$ - Eucalyptus spp. from 2 to 4 years.

$E_{4yrs. - 7yrs.}$ - Eucalyptus spp. from 4 to 7 years.

$E_{7yrs. - 20yrs.}$ - Eucalyptus spp. from 7 to 20 years.

$RE_{< 2yrs.}$ - Regrowth Eucalyptus spp. under 2 years.

$RE_{2yrs. - 9yrs.}$ - Regrowth Eucalyptus spp. from 2 to 9 years.

Bhattacharyya distance (B-distance, Marill and Green, 1963) between these ten preliminary classes was calculated using INPE's feature selection algorithm. This algorithm assumes that each class has a multivariate Gaussian distribution. In this operation, 40 to 50 training samples of each class were delineated on the image monitor by the electronic cursor which was adjusted to the minimum size equivalent to a 4 pixels/sample. Then the B-distance value was calculated in the four available channels. The results of the B-distance values for the preliminary *Pinus* and *Eucalyptus* classes in four channels are presented in Tables I and II. The separability of the preliminary classes was defined according to the values of B-distance. Using the curve of Swain and King (1973) the estimated probability of correct classification is > 0.85 for the value of B-distance > 1.30. Thus those classes having B-distance values less than 1.30 were grouped together to form four classes as follows:

PE --- *P. elliottii*.

PO --- *Pinus* species other than *P. Elliottii*.

EY --- *Eucalyptus* spp. under two years.

EO --- *Eucalyptus* spp. over two years.

Table III lists the "B-distance" values of classes PE, PO, EY and EO in four spectral channels.

3.2 SIGNATURE EXTRACTION AND CLASSIFICATION

Single cell signature acquisition of the Image-100 was used for classification. This option creates a four-dimensional parallelepiped; each of the sides of which corresponds to the signature limits of the training areas in each channel. Training samples, independent from the test area, were selected from each of the four classes (PE, PO, EY and EO) using the electronic cursor. The limits of the histograms of the training areas were modified in the four spectral channels until a satisfactory classification was obtained. The spectral statistics of the histograms from the training samples of four classes are given in Tables IV to VII, respectively. A density slicing routine was also used to give pixel frequencies of the study area for all possible combinations of the four channels selected axes. An example of this result is shown in Fig. 3.

4. IMAGE-100 PERFORMANCE

.1 PIXEL CLASSIFICATION ACCURACY

After classification, the electronic cursor was again positioned on the image monitor. The "alphanumeric theme print" was executed for a cursored area of approximately 50 km^2. This printout was compared pixel by pixel with the IFSP and CCPs forest cover maps to produce a confusion matrix (Table VIII). This table shows the number of pixels classified according to its "true" class and Image-100's "chosen" class in the cursored area. The percentages given in parenthesis were calculated based on the total number of classified pixels in each class. Histogram modification thresholding in the four spectral channels tends to increase errors of omission and decrease errors of commission. Thus, percentages of correct classification which take into account errors of omission were not high. To give a more realistic view of the Image-100's pixel classification accuracy the following formulas were used to calculate the percentages of correct classification and errors of commission.

$$\text{Correct Classification (\%)} = \frac{\text{No. of correctly classified pixels in class A}}{\text{Total no. of pixels in the test area of class A}}$$

$$\text{Error of Commission (\%)} = \frac{\text{No. of pixels classified as class A but actually belonging to another class}}{\text{Total no. of pixels in the test area of non-A class}}$$

4.2 AREA ESTIMATION ACCURACY

The Image-100 estimated class areas were compared to their corresponding "ground truth" data provided by IFSP and CCP. The percentage of the estimated areas on ground truth data are presented in Table X.

5. DISCUSSION AND CONCLUSIONS

For the study, sufficiently large test areas of **pine trees less than** thirteen years old and trees over thirteen years old were further separated geographically into two classes, PE_a and PE_b, although a "B-distance" value of 0.22 indicated that no satisfactory separation was possible. P. taeda and the class OP were readily separable from P. elliottii but unseparable from each other (Table I). The morphological similarity among the Eucalyptus species made their differentiation difficult, although Image-100 successfully classified Eucalyptus under two years from the other preliminary age groups (Table II). This separability was due to the distinct spectral characteristics of the homogenous young Eucalyptus which aided the Image-100 identification.

For classification a hyperparallelepiped signature file was constructed for each of the classes PE, PO, EY and EO using a "single cell signature acquisition" program. The spectral statistics of the signature files are shown in Tables V, VI, VII and VIII, but separability of these four classes was most easily observed on "N-dimensional histogram two-dimensional print" (Fig. 3).The longer distance between genera Pinus and Eucalyptus confirmed their higher separability than species and age classes. The percentages of Image 100's correct classification for the classes PE, PO, EY and EO were found to be 89.14%, 79.26%, 94.12% and 70.99% respectively. These percentages were relatively lower compared to those calculated using the total number of pixels from each class classified as the denominator (Table IX). The lower percentage of correct classification was obtained from the class EO (70.99%). It was caused by cerrado, gallery forest and heterogeneous P. elliottii plots of the area which gave a similar spectral responses as class EO. The highest percentage of correct classification was found for class EY (94.12%). It may have been due to the homogeneity of the eucalypt under two years old.

In the area estimation study the misclassified pixels shown in Table VIII accounted for the low pixel classification accuracies of Table IX and gave area estimations of 17.28, 1.44, 4.00 and 22.51 km^2 for the classes PE, PO, EY and Eo respectively (Table X). The above areas estimated by the Image-100 when compared to the same ground truth areas gave recognition values (RVs) of 91.62%, 87.80%, 89.99% and 103.30%. The 3.30% overestimation of class EO was due to a high proportion of misclassified pixels (13.61%) which had produced spectral signature similar to class EO. The overall area-weight difference of 6.31 was achieved. The results of the Image-100 area estimations were encouraging, considering that the machine classification was carried out at the species level, for pine, and based on age groups, for eucalypt.

The ground truth provided by IFSP and CCP agencies were based on the seedling area at planting time. These data represented "theoretically reforested areas" under "optimum conditions" without damage and mortality. In the field check, various plant-failure areas were observed. The recognition value could have been greater if the ground truth data had been up-dated annually.

The following significant results were obtained from automatic classification of reforested areas of pine and eucalypt:

Based on "B-distance" values the following spectral separations were achieved:

1. P. elliottii was separated from the other pine species.
2. Eucalyptus spp. under two years old were separated from Eucalyptus spp. more than two years old.
3. Clear distinctions were made between pine and eucalypt.

Based on Image-100 area estimations, **RV ranged from 87.80% to** 103.30% for the four classes defined (PE, PO, EY and EO) when compared to the same information provided by forestry agencies.

These results point to the operational potential of computer automated processing of Landsat data to recognize tree species and calculate standing volume, especially in remote regions where ground and aerial surveys are often costly.

Before the methodology used in this study can be applied on an operational scale, more consideration should be given to signature extention to other geographical regions, and to spatial differences between areas of natural and artifical reforestations.

REFERENCES

Gimbarzevsky, P., 1974. ERTS-1 imagery in biophysical studies: Proceedings, Second Canadian Symposium on Remote Sensing, Ottawa, V. 2, P. 391.

Kalensky, Z. and L. R. Scherk, 1975. Accuracy of forest mapping from Landsat computer compatible tapes: Proceedings, Tenth International Symposium on Remote Sensing of Environment, Environment Research Institute of Michigan, Ann Arbor, Michigan, V. 2, P. 1159.

López-Cuervo, S., 1973. A study of agroforestry natural resources: Proceedings, First Panamerican Symposium on Remote Sensing, Panama City, EROS Data Center, Sioux Falls, P. 86.

Lee, Y. J., E. T. Oswald, and J. W. E. Harris, 1974. A preliminary evaluation of ERTS imagery for forest-land management in British Columbia: Proceedings, Second Canadian Symposium on Remote Sensing, Ottawa, V. 1, P. 87.

Marill, T. and D. M. Green, 1963. On the effectiveness of receptor in recognition systems: IEEE Trans. on Information Theory, V. IT-9, P. 11.

Sayn-Wittgenstein, L., 1972. A first look at Canadian ERTS experiments in forestry: Proceedings, Earth Resources Technology Satellite-1 Symposium, Goddard Space Flight Center, Beltsville, Maryland, P. 48.

Swain, P. H. and R. C. King, 1973. Two effective feature selection criteria for multispectral remote sensing: LARS Information Note 042673, Laboratory for Applications of Remote Sensing, Purdue Univ., W. Lafayette, IN.

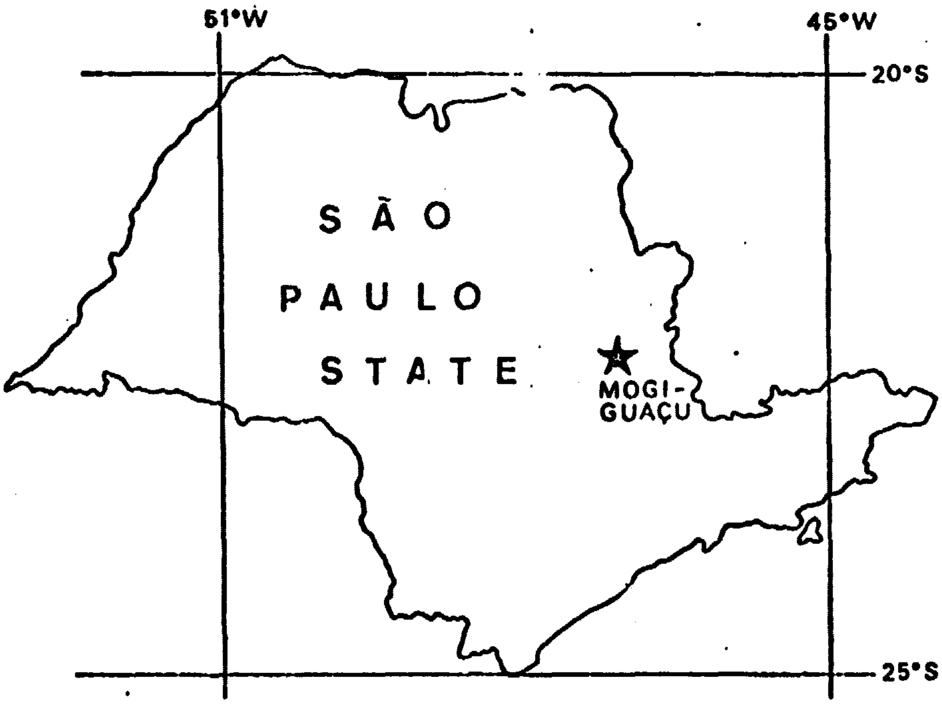

Figure 1. LOCATION OF STUDY AREA IN MOGI-GUAÇU, SÃO PAULO STATE

ORIGINAL PAGE IS
OF POOR QUALITY

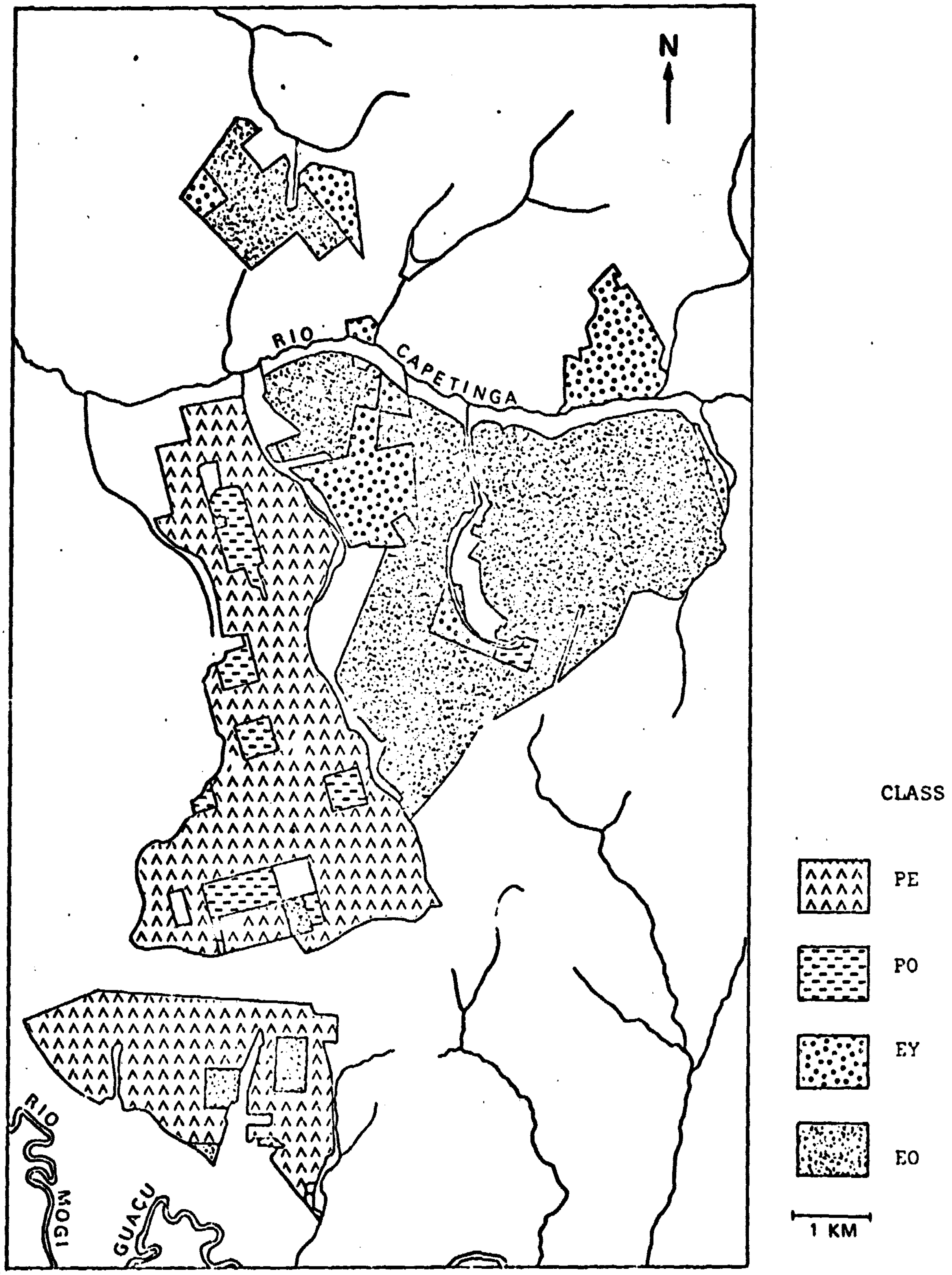

Figure 2. REGROUPED FOREST COVER TYPES OF CAMPININHA AND SANTA TEREZINHA

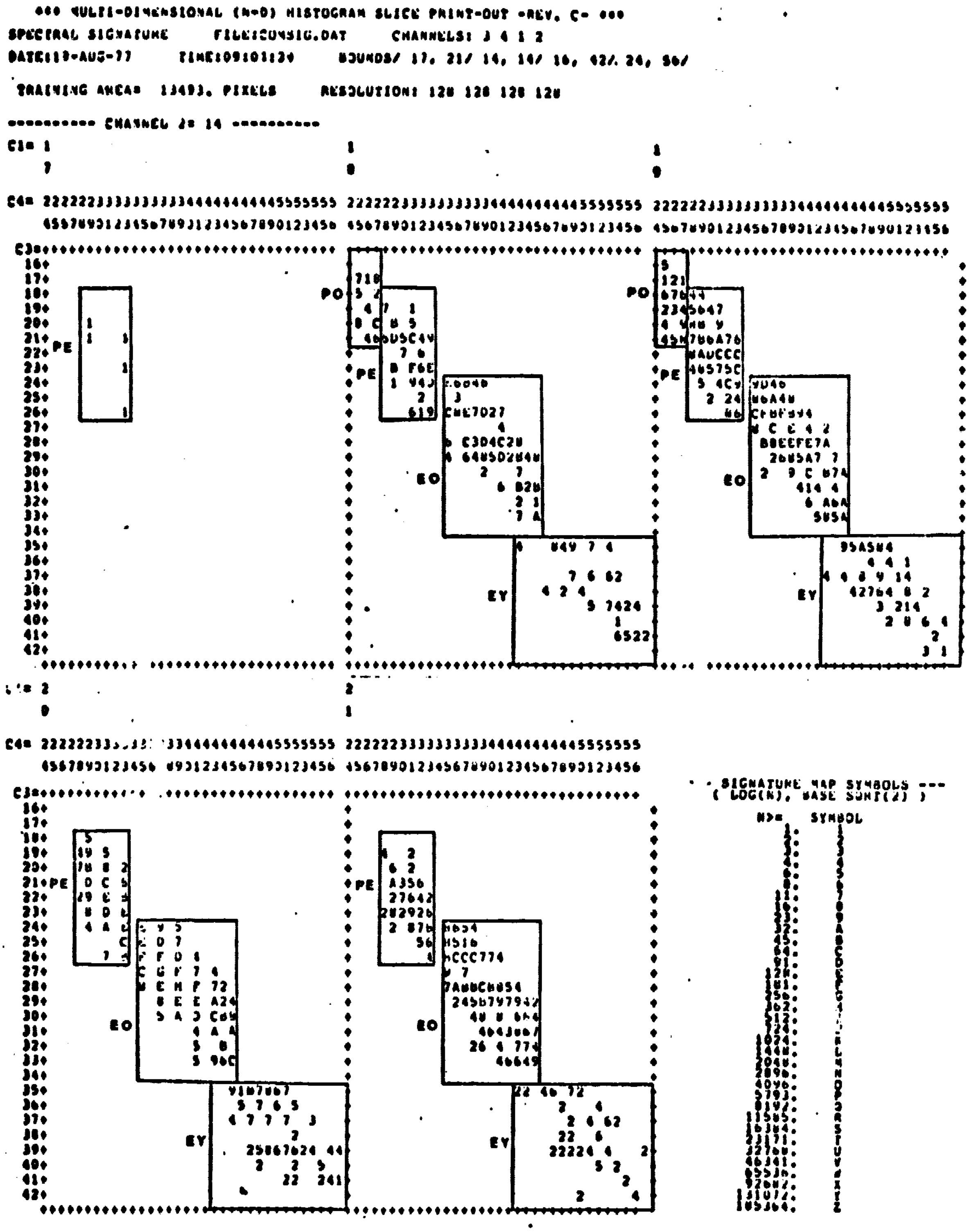

Figure 3. AN EXAMPLE OF "N-DIMENSIONAL HISTOGRAM TWO-DIMENSIONAL PRINT"

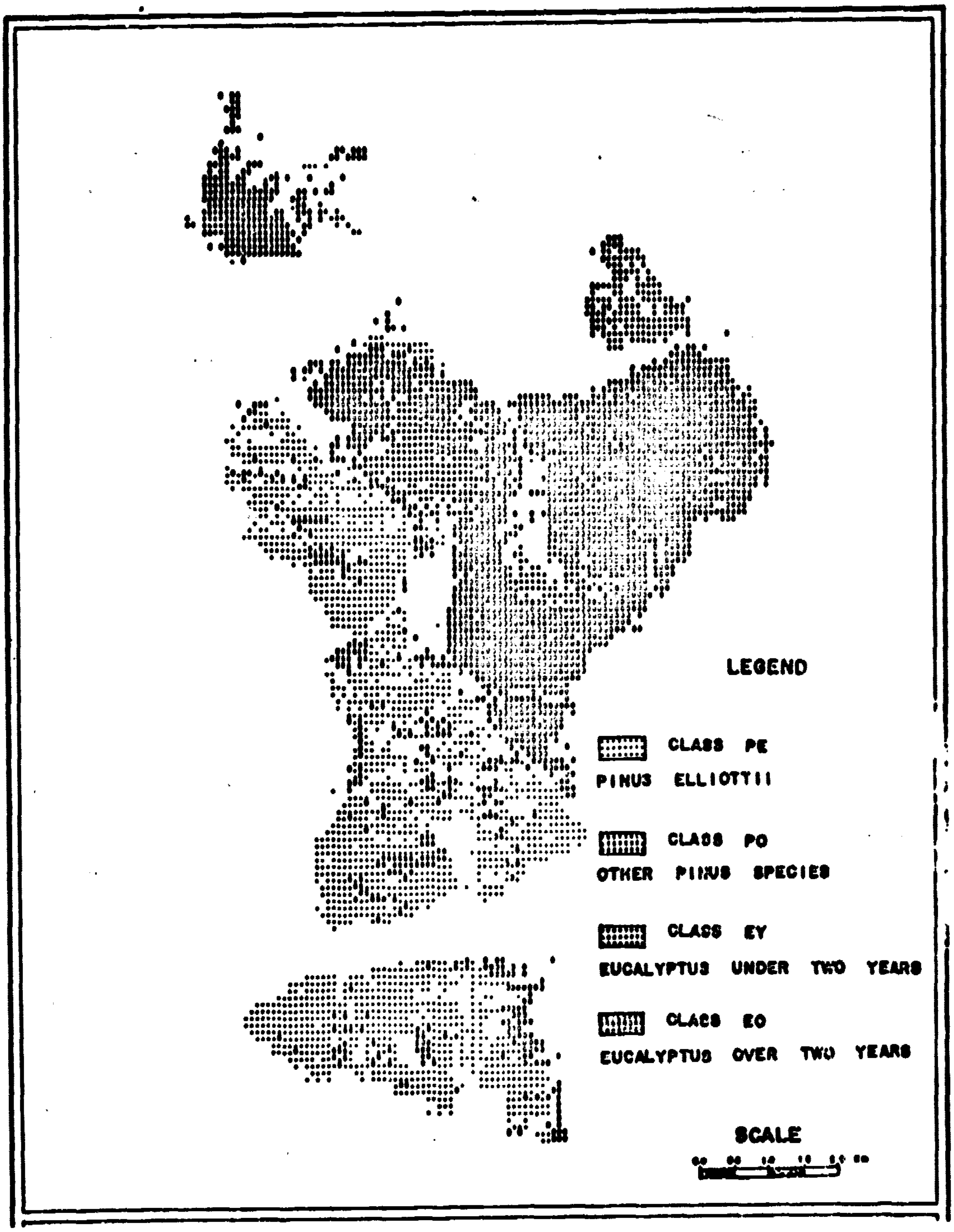

Figure 4. ALPHANUMERIC PRINTOUT OF THE CLASSIFIED MOGI-GUAÇU STUDY AREA

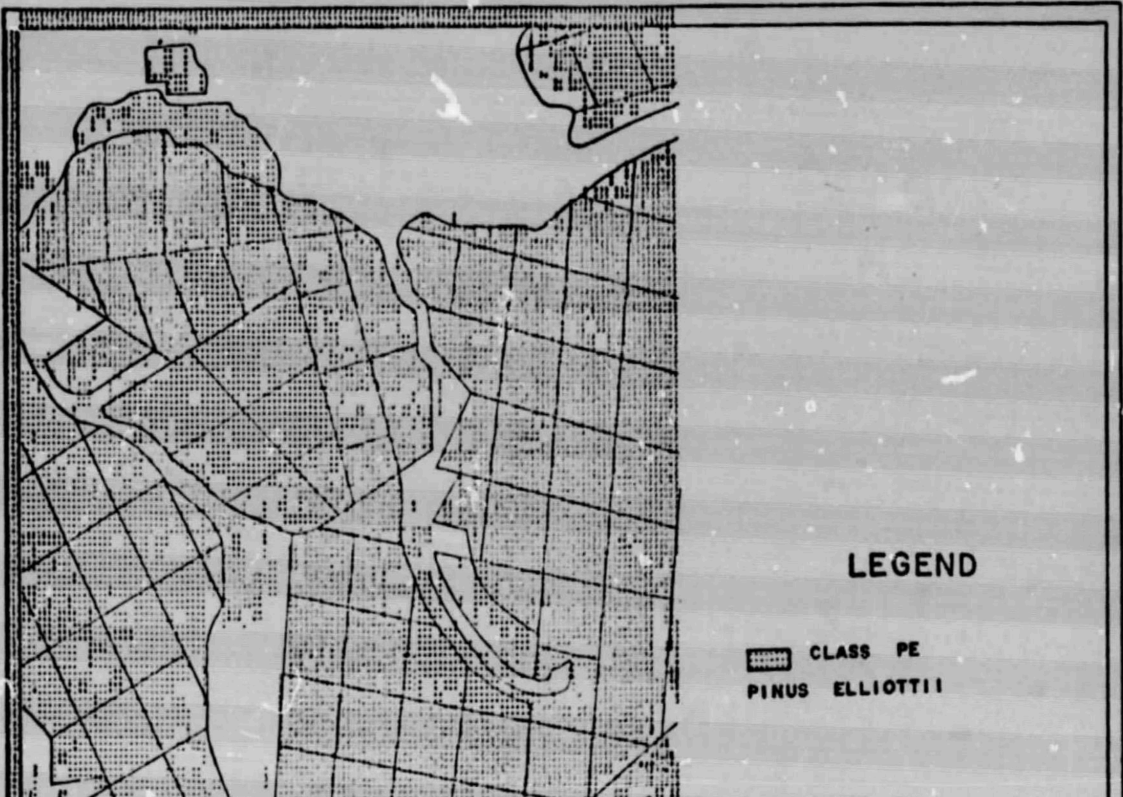

Figure 5. ALPHANUMERIC THEME PRINT USED FOR PIXEL CLASSIFICATION ACCURACY MEASUREMENT

CLASS \ CLASS	PE_b	OP	PT
PE_a	0.22	1.92	1.70
PE_b		1.79	1.37
OP			0.67

TABLE I. B-DISTANCE VALUES OF PRELIMINARY CLASSES OF <u>PINUS</u> IN FOUR SPECTRAL CHANNELS

CLASS \ CLASS	PO	EY	EO
PE	1.78	2.00	1.99
PO		2.00	2.00
EY			1.90

TABLE III. B-DISTANCE VALUES OF CLASSES PE, PO, EY AND EO IN FOUR CHANNELS

CLASS \ CLASS	$E_{2yrs.}$ - 4yrs.	$E_{4yrs.}$ - 7yrs.	$E_{7yrs.}$ - 20yrs.	RE < 2yrs.	$RE_{2yrs.}$ - 9yrs.
$E_{8mo.}$ - 2yrs.	1.99	1.99	1.98	1.94	1.98
$E_{2yrs.}$ - 4yrs.		0.41	0.23	0.51	0.31
$E_{4yrs.}$ - 7yrs.			0.63	0.70	0.49
$E_{7yrs.}$ - 20yrs.				0.56	0.55
RE < 2yrs.					0.23

TABLE II. B-DISTANCE VALUES OF PRELIMINARY CLASSES OF <u>EUCALYPTUS</u> IN FOUR SPECTRAL CHANNELS

CHANNEL	SPECTRAL BOUNDS		DELTA	MEAN	VARIANCE
	LB	UB			
4	17	21	5	19.0	0.8
5	11	15	5	13.3	1.0
6	18	26	9	22.[illegible]	2.5
7	27	32	6	30.1	2.5

TABLE IV. SPECTRAL STATISTICS FROM THE TRAINING SAMPLES OF CLASS PE

CHANNEL	SPECTRAL BOUNDS		DELTA	MEAN	VARIANCE
	LB	UB			
4	18	19	2	18.5	0.2
5	11	15	5	13.3	1.0
6	16	21	6	19.1	2.3
7	24	26	3	25.4	0.7

TABLE V. SPECTRAL STATISTICS FROM THE TRAINING SAMPLES OF CLASS PO

CHANNEL	SPECTRAL BOUNDS		DELTA	MEAN	VARIANCE
	LB	UB			
4	18	21	4	19.6	1.0
5	12	19	8	14.5	2.5
6	35	42	8	36.7	3.8
7	42	56	15	48.1	7.9

TABLE VI. SPECTRAL STATISTICS FROM THE TRAINING SAMPLES OF CLASS EY

CHANNEL	SPECTRAL BOUNDS		DELTA	MEAN	VARIANCE
	LB	UB			
4	18	21	4	19.8	1.0
5	13	18	6	15.2	2.0
6	24	34	11	27.6	4.6
7	34	44	11	37.6	7.7

TABLE VII. SPECTRAL STATISTICS FROM THE TRAINING SAMPLES OF CLASS EO

TRUE CLASS \ CHOSEN CLASS	PE	PO	EY	EO
PE	5730 (92.26%)	30 (5.19%)	0	449 (6.18%)
PO	359 (5.78%)	535 (92.56%)	0	40 (0.55%)
EY	0	0	1713 (91.21%)	195 (2.69%)
EO	24 (0.38%)	0	155 (8.25%)	6272 (86.39%)
NO CLASS	98 (1.58%)	13 (2.25%)	10 (0.53%)	304 (4.19%)
PROPORTION OF MISCLASSIFIED PIXEL	7.74%	7.44%	8.79%	13.61%
PIXEL TOTAL	6211	578	1878	7260

TABLE VIII. CONFUSION MATRIX

CLASS	CORRECT CLASSIFICATION (%)	ERROR OF COMMISSION (%)
PE	89.14	2.11
PO	79.26	0.15
EY	94.12	0.61
EO	70.99	2.08

TABLE IX. SUMMARY OF CORRECT CLASSIFICATIONS AND ERRORS OF COMMISSION

CLASS	AREA (KM^2)		R.V.* (%)	RELATIVE DIFFERENCE (%)
	GROUND TRUTH	IMAGE-100 ESTIMATED		
PE	18.86	17.28	91.62	- 8.38
PO	1.64	1.44	87.80	- 12.20
EY	4.45	4.00	89.89	- 10.11
EO	21.79	22.51	103.30	+ 3.30
OVERALL AREA WEIGHTED DIFFERENCE				6.31

* RECOGNITION VALUE = $\frac{\text{IMAGE-100 ESTIMATED AREA}}{\text{GROUND TRUTH AREA}} \times 100$

TABLE X. COMPARISON OF IMAGE-100 ESTIMATED AREAS WITH "GROUND TRUTH" DATA