

ESTIMAÇÃO DA ÁREA OCUPADA COM CANA-DE-AÇÚCAR NA GRANDE REGIÃO CANA  
VIEIRA DE PIRACICABA, UTILIZANDO-SE IMAGENS LANDSAT E FOTOGRAFIAS  
AÉREAS ATRAVÉS DA SIMULAÇÃO DE UM SISTEMA DE  
AMOSTRAGEM EM DUAS FACES

C. A. Cappelletti, F. J. Mendonça, D. C. L. Lee e Y. E. Shimabukuro  
Instituto de Pesquisas Espaciais  
Conselho Nacional de Desenvolvimento Científico e Tecnológico  
Caixa Postal 515, 12200-São José dos Campos, SP, Brasil

RESUMO

O presente trabalho tem por objetivo realizar a estimativa da área ocupada com cana-de-açúcar na região de Piracicaba, utilizando-se dados combinados do LANDSAT e fotografias aéreas através do método de amostragem em duas fases com estimativa de regressão, variância fixada e custo mínimo. A área ocupada com cana-de-açúcar obtida através da verdade terrestre (interpretação de fotografias aéreas) foi de 278.500 ha. Utilizando-se dados combinados do LANDSAT e fotografias aéreas através do método de amostragem em duas fases da área estimada com cana-de-açúcar, quando se utilizaram dados referentes à safra de 1978/79, foi de 281.038 ha, o que corresponde a um erro de 0,91% em relação aos dados da verdade terrestre. Utilizando-se dados referentes à safra de 1979/80, a área estimada com cana-de-açúcar foi de 0,01% em relação aos dados da verdade terrestre.

ABSTRACT

The objective of this study is to estimate the area of sugar-cane plantation using combined LANDSAT and aircraft data through two-phase sampling method with estimation of regression, fixed variance and minimum cost. The sugar-cane area assessed by aerial photographs (considered as ground truth) was 278,500 ha. Using both LANDSAT data and aerial photographs associated with the two-phase sampling method, the area obtained for sugar-cane was 281,038 ha for the 1978/79 crop year. This result compared with the ground truth presented an error (relative difference) of 0,91%. Using the same technique for the 1979/80 crop year the relative difference was 0,01%.

1. INTRODUÇÃO

A área de estudo selecionada foi a Grande Região Canavieira de Piracicaba, que ocupa uma extensão de 10.813 km<sup>2</sup>, localizada entre os paralelos 22°00' e 23°30' sul e os meridianos 46°30' e 48°00' oeste (Figura 1).

A região escolhida constituiu uma importante área de concentração de cana-de-açúcar do Estado de São Paulo, e é utilizada nas estimativas de safra convencionais executadas pelo IAA/PLANALSUCAR.

Koffler et alii (1979) realizaram um inventário com auxílio de fotografias aéreas, interpretadas através de amostragem sistemática por grade de pontos, de toda a região, onde se verificou que 27% dos solos estão ocupados com a cana-de-açúcar.

Paralelamente, foi realizado por Mendonça et alii (1980) um inventário canavieiro do Estado de São Paulo, que inclui a Grande Região Canavieira de Piracicaba, referente à safra de 1978/79, utilizando dados do LANDSAT. Esses dados foram tratados visualmente, explorando somente as suas características específicas.

Para a safra de 1979/80, Mendonça et alii (1981) repetiram o inventário canavieiro através de imagens LANDSAT, explorando as características espectrais e temporais dos dados orbitais.

Baseado nos dados dos trabalhos já citados, este estudo objetivou realizar para a Grande Região Canavieira de Piracicaba uma estimativa da área de cana-de-açúcar através da implementação de um sistema de amostragem, utilizando-se dados do LANDSAT e fotografias aéreas. Esses dados foram processados através de um sistema de amostragem em duas fases que utiliza estimativa de regressão, fixando a variância da estimação e minimizando o custo do sistema.

2. MATERIAL E MÉTODOS

2.1 - MATERIAL

No delineamento do experimento foram utilizados mapas temáticos de distribuição espacial da cana-de-açúcar na escala de 1:250.000, obtidos nos levantamentos realizados por Koffler et alii (1979) e Mendonça et alii (1980 e 1981). Os primeiros utilizaram fotografias aéreas pancromáticas nas escalas

de 1:45.000, 1:35.000 e 1:25.000, enquanto Mendonça et alii (1980 e 1981) utilizaram imagens LANDSAT na escala de 1:250.000 (canais 5 e 7), tratadas visualmente.

2.2 - TRAÇADO DA GRADE DE AMOSTRAGEM

Na avaliação da área ocupada com cana-de-açúcar, Koffler et alii (1979) superpôs sobre o mapa temático uma grade de pontos na qual cada um deles equivalia a uma área de 100 ha. Nos levantamentos realizados por Mendonça et alii (1980 e 1981), o cálculo da área ocupada com cana-de-açúcar foi feito sobre uma grade de pontos, onde cada um deles equivalia a uma área de 6,25 ha.

Para a seleção da amostra deste experimento (Figura 1), e dado que a escala de trabalho foi de 1:250.000, utilizou-se uma grade de cujo tamanho da quadrícula correspondeu a uma área de 625 ha. O número total dessas quadrículas para cobrir toda a área de estudo foi da ordem de 1868. Essa grade foi superposta aos mapas temáticos gerados a partir de fotografias aéreas e aos mapas temáti-

cos gerados a partir de imagens LANDSAT. Da contagem de pontos em cada quadrícula, obtiveram-se os valores de duas variáveis. A primeira, X, quando a grade estava sobre os mapas temáticos gerados a partir de dados de imagens LANDSAT; e a segunda, Y, quando a grade estava sobre os mapas temáticos obtidos através de dados de fotografias aéreas.

2.3 - DETERMINAÇÃO DO TAMANHO DO SEGMENTO AMOSTRAL

De acordo com Federer (1967), foi delineado um experimento de uniformidade a fim de se determinar o tamanho ótimo do segmento amostral. A escolha desse segmento foi feita utilizando-se o critério da máxima correlação entre as variáveis X (área obtida com LANDSAT) e Y (área obtida com fotografias aéreas). Assim sendo, a partir dos resultados apresentados na Tabela 1 elegeu-se como segmento amostral deste trabalho a quadrícula de 5 km x 7,5 km (Figura 1), a qual corresponde a uma área de 3750 ha.

TABELA 1  
DADOS REFERENTES AO TAMANHO E NÚMERO DOS SEGMENTOS AMOSTRAIS E SEUS PARÂMETROS ESTATÍSTICOS

TAMANHO DOS SEGMENTOS AMOSTRAIS	NÚMERO DE SEGMENTOS AMOSTRAIS NA ÁREA DO EXPERIMENTO	r	ML	DPL	MA	DPA
(km)			(Ha)		(Ha)	
2,5 x 2,5	480	0,15	60	141	143	171
2,5 x 5,0	240	0,21	120	270	286	291
2,5 x 7,5	160	0,31	181	360	429	394
2,5 x 10,0	120	0,28	241	484	573	489
5,0 x 5,0	120	0,24	241	541	573	541
5,0 x 7,5	80	0,54	360	723	859	714
5,0 x 10,0	60	0,35	597	739	1145	904
7,5 x 7,5	48	0,41	541	1088	1346	1020
7,5 x 10,0	36	0,37	722	1466	1794	1287
10,0 x 10,0	30	0,33	963	1960	2290	1653

r: correlação entre as variáveis X (dados do LANDSAT) e Y (dados das fotografias aéreas).

ML: média de área de cana-de-açúcar no segmento amostral obtido a partir da identificação de dados do LANDSAT.

DPL: desvio padrão calculado a partir da média de área de cana-de-açúcar do segmento amostral, utilizando-se dados do LANDSAT.

MA: média de área de cana-de-açúcar no segmento amostral obtido a partir da identificação de dados de fotografias aéreas.

DPA: desvio padrão calculado a partir da média de área de cana-de-açúcar do segmento amostral, utilizando-se dados de fotografias aéreas.

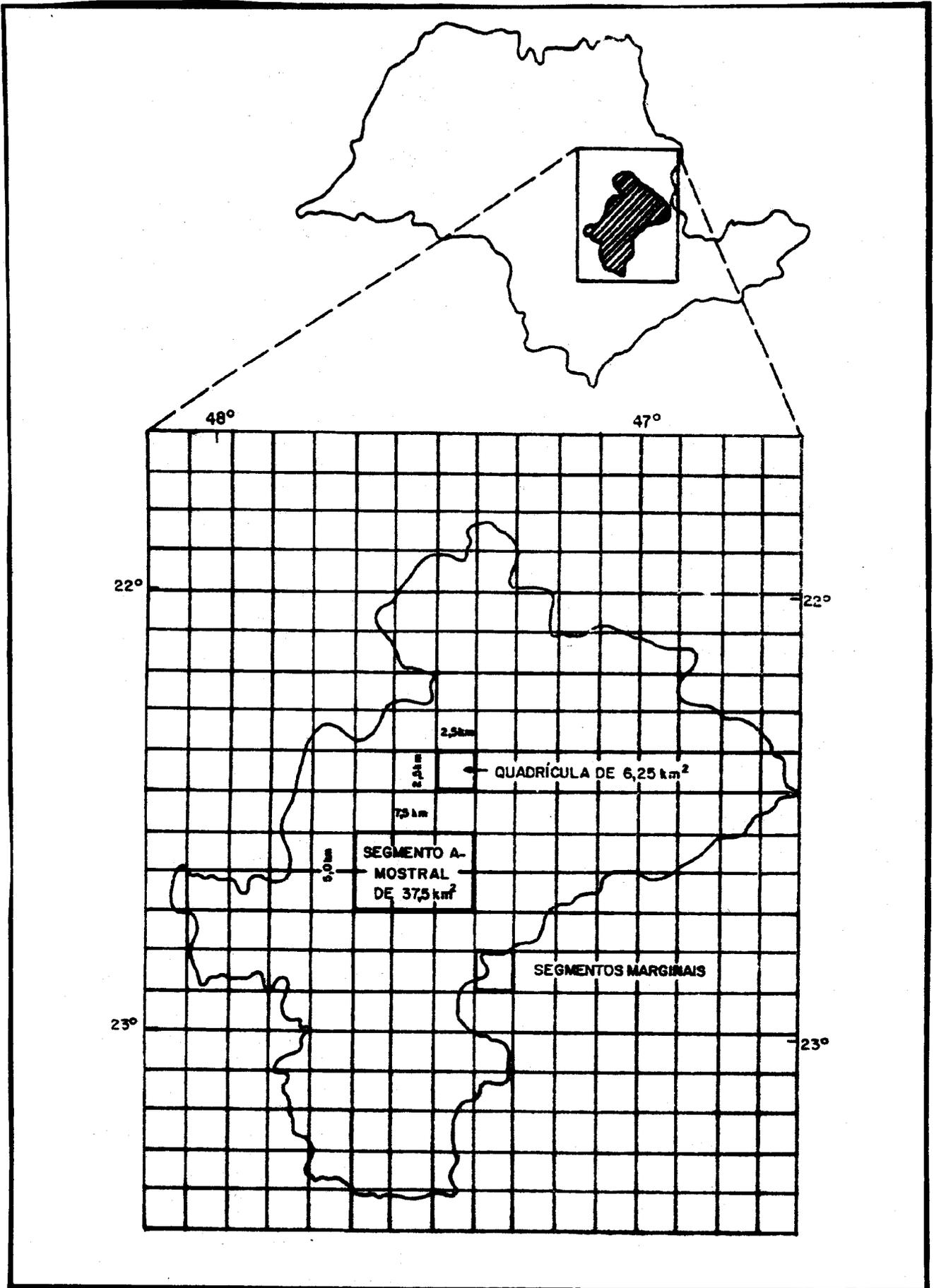


Fig. 1 - Mapa de situação da área de estudo dentro do Estado de São Paulo, demonstrativo do tamanho das quadrículas, dos segmentos amostrais e dos detalhes dos segmentos marginais.

## 2.4 - MÉTODO DE AMOSTRAGEM

O método escolhido foi um sistema de amostragem em duas fases com estimativa de regressão, no qual se combinam as variáveis X e Y.

Esse tipo de amostragem se baseia num critério que leva em conta a correlação entre as variáveis X e Y e a relação de custos da coleta de dados de cada variável. Assim sendo, a variância da estimativa é função da correlação entre X e Y, diminuindo à medida que a correlação aumenta. Por sua vez, a magnitude do coeficiente de correlação influencia na determinação do número de segmento que integram a amostra.

Neste método, a seleção da amostra é feita em duas fases; na primeira fase escolhe-se uma amostra de tamanho grande na variável X (LANDSAT), cujo custo de coleta é menor do que o da variável Y (Avião); a segunda fase é uma subamostra da amostra da primeira fase.

Anteriormente à seleção da amostra, procedeu-se a uma redefinição dos limites da região de estudo, a fim de eliminar zonas cujas áreas não apresentavam cana-de-açúcar, bem como eliminar segmentos marginais (Figura 1) cuja área na região eram inferiores a 1/6 do tamanho padrão dos segmentos amostrais.

Os fundamentos teóricos do método de amostragem aqui utilizados encontram-se em Cochran (1963), Loestsch and Haller (1973) e Jessen (1978).

## 2.5 - ANÁLISE ESTATÍSTICA

Para a determinação do número de amostras em cada fase, adotou-se como critério a minimização do custo total da estimativa e um erro não maior que 5% para um coeficiente de confiança de 95%.

Os valores de k, número de segmentos na primeira fase da amostragem sobre X, e n, número de segmentos na segunda fase da amostragem sobre Y, basearam-se na variância de  $\bar{Y}$  ( $s_y^2$ ) e na correlação (r) de X e Y. O cálculo de k e n e dos parâmetros estatísticos foram obtidos a partir dos valores de toda a população que se tinha disponível.

Obtidos k e n, simulou-se uma sequência de 100 amostras aleatórias irrestritas para cada fase, das quais a amostra da segunda fase foi uma subamostra aleatória da primeira fase.

Com os dados da simulação, calculou-se uma sequência de valores de  $\bar{Y}_R$  (estimador do total de cana-de-açúcar) e de  $V(\bar{Y}_R)$  (variância do estimador), obtendo-se:

$$D = \bar{Y}_R - \alpha$$

onde  $\alpha$  corresponde ao valor de 278.500 ha de cana-de-açúcar de toda a região em estudo, obtido através da interpretação de fotogra-

fias aéreas. O valor  $\alpha$  foi tomado como referência para a avaliação da estimativa de área plantada com cana-de-açúcar, obtida através do sistema de amostragem em duas fases.

## 3. RESULTADOS E DISCUSSÃO

### 3.1 - LEVANTAMENTO DE 1978/79

#### 3.1.1 - OBTENÇÃO DAS AMOSTRAS

Considerando-se uma correlação  $r_{xy} = 0,536$ , um desvio padrão  $s_y = 860$ , uma relação entre os custos dos levantamentos com dados do LANDSAT e de fotografias aéreas (C1/C2) de 1:21 e fixando-se um erro máximo na estimativa da ordem de 5% com uma confiança de 95%, o número de segmentos amostrais para a primeira fase (variável X, área obtida pelo LANDSAT) foi  $k = 233$  e para a segunda fase (variável Y, fotografias aéreas) foi  $n = 80$ . O custo total foi de 1913 unidades monetárias.

#### 3.1.2 - RESULTADOS DA AMOSTRAGEM

Da simulação da amostragem chegou-se aos seguintes resultados:

- O valor médio da sequência de estimativas da área ocupada com cana-de-açúcar na região de estudo foi de 281.038 ha. Tal valor corresponde a 101% do total obtido no levantamento feito através de fotografias aéreas.
- O intervalo de confiança do total da área cultivada, para um coeficiente de confiança de 95%, foi

(275.224 ha; 286.852 ha),

cujos limites, em termos percentuais, correspondem a 98,8% e 102,9% do obtido através das fotografias aéreas, consideradas aqui como verdade terrestre.

- comparando-se o valor obtido da verdade terrestre, 278.500 ha, com a estimativa amostral, pode-se observar que a diferença absoluta é de 2.538 ha, com um erro padrão de 2.907 ha. O cálculo do intervalo de confiança com 95% teve os limites de -3.276 ha e 8.352 ha.
- A área total com cana-de-açúcar obtida através da interpretação visual de imagens LANDSAT foi de 99.994 ha. Comparando-se este resultado com aquele obtido através da amostragem, tem-se uma diferença absoluta de 181.044 ha, que representa 65% daquele obtido através de fotografias aéreas. A diferença diminui para 1% quando se utiliza o sistema de amostragem em duas fases.

- A eficiência do sistema de amostragem em duas fases, comparada com o sistema de fase única, no qual se utilizaram somente dados de fotografias aéreas, foi de 110%, o que significa que para cada uni-

dade monetária gasta na amostragem em duas fases foi necessário investir 1,1 unidades na amostragem simples.

No caso de se utilizar um sistema de amostragem aleatória simples, unicamente com dados do LANDSAT, para o presente experimento, seria necessária uma amostra de  $k = 178$  segmentos para garantir um erro não maior que 5% na estimativa, com um coeficiente de confiança de 95%. Não obstante tal dimensionamento, optou-se, para efeito de comparação, usar o mesmo tamanho da amostra em duas fases, ou seja,  $k = 233$ .

A estimativa da área plantada foi  $\bar{X} = 101.108$  ha, o que corresponde a uma diferença absoluta de 177.392 ha, que representa um erro percentual de 64% em relação ao valor obtido através do levantamento realizado com fotografias aéreas.

### 3.2 - LEVANTAMENTO DE 1979/80

Segundo foi dito, na interpretação dos dados do LANDSAT referentes à safra 1979/80 foi utilizada, além da característica espectral das imagens, a variação temporal da cultura. Desse modo, a relação dos custos C1/C2 passou para 1:13, com relação aos dados do levantamento de 1978/79.

#### 3.2.1 - OBTENÇÃO DAS AMOSTRAS

O desvio padrão  $s_y = 860$  é o mesmo que para 1978/79 (uma vez que os dados de verdade terrestre foram os mesmos); mas a correlação aumentou para  $r_{xy} = 0,721$ . Como consequência, o tamanho da amostra foi de  $k = 229$  segmentos amostrais para a primeira fase e de  $n = 62$  para a segunda. O custo total foi de 1022 unidades monetárias, o que significa uma redução da ordem de 47% em relação ao custo total anterior.

Comparando-se o tamanho da amostra dos dois levantamentos, 1978/79 e 1979/80, pode-se observar que houve uma redução de 23% no tamanho da amostra da segunda fase, redução esta relacionada com o incremento do coeficiente de correlação que passou de 0,536 para 0,721.

#### 3.2.2 - RESULTADOS DA AMOSTRAGEM

Do mesmo modo que para 1978/79, foi simulada uma sequência de 100 amostras, obtendo-se assim os seguintes resultados:

- O valor médio da sequência de estimativas da área ocupada com cana-de-açúcar na região de estudo foi 278.465 ha. Este valor correspondeu a 99,8% do total obtido no levantamento feito através de fotografias aéreas.
- O intervalo de confiança do total da área cultivada, para um coeficiente de confiança de 95%, foi  
(273.149 ha; 283.781 ha),

cujos limites, em termos percentuais, correspondem a 98,1% e 101,9% da verdade terrestre.

- Comparando-se o valor obtido da verdade terrestre, 278.500 ha, com a estimativa amostral, obteve-se uma diferença absoluta de 35 ha, com um erro padrão de 2.658 ha. O intervalo de confiança teve os limites de -5.351 ha e 5.281 ha para um coeficiente de confiança de 95%.
- A área total ocupada com cana-de-açúcar obtida da interpretação das imagens LANDSAT foi de 155.526 ha. Comparando-se este valor e a estimativa amostral (278.465 ha) com o da verdade terrestre (278.500 ha) tem-se diferenças da ordem de 122.974 ha e 35 ha, que correspondem a 44,2% e 0,01%, respectivamente.
- A eficiência do sistema de amostragem em duas fases, comparada com o sistema de amostragem de uma única fase, na qual se utiliza somente dados de fotografias aéreas, foi de 127%.

Comparando-se a eficiência relativa dos dois levantamentos pode-se concluir que o valor de 127%, relativo ao levantamento de 1979/80, em comparação com o valor de 110%, do levantamento de 1978/79, está relacionado com o aumento do coeficiente de correlação daquele levantamento sobre este. E, por sua vez, esse aumento do coeficiente de correlação está relacionado com a melhoria na identificação da cultura da cana-de-açúcar nas imagens LANDSAT, ocorrida mediante o uso conjunto de variação temporal da cultura com suas características espectrais.

Utilizando-se um sistema de amostragem aleatória simples, unicamente com dados do LANDSAT, a estimativa da área total com cana-de-açúcar foi  $\bar{X} = 155.525$  ha, que corresponde a 56% do valor da verdade terrestre e representa um erro percentual de 44%.

### 3.3 - DISCUSSÃO DOS RESULTADOS

A Tabela 2 compara os resultados das estimativas de área de cana-de-açúcar obtidos através dos métodos de amostragem em duas fases e aleatória simples com o valor da área de cana-de-açúcar obtido através de interpretação de fotografias áreas.

Na Tabela 2 pode-se observar que, ao se utilizar o método de amostragem em duas fases, os erros relativos foram muito pequenos, quer se tratasse do levantamento feito em 1978/79 ou 1979/80.

No entanto, ao se utilizar o método de amostragem aleatória simples, os erros relativos se mostraram bastante elevados para os dois levantamentos. A magnitude destes erros parece ter sua justificativa no fato de terem sido utilizados neste método somente da

dos do LANDSAT que apresentam problemas de baixa resolução (tanto espectral como geométrica), o que limita a identificação da cana-de-açúcar em seus diferentes estágios de desenvolvimento. Desse modo, parte da cana nova, que ainda não cobria totalmente o solo, tem sua resposta espectral influenciada pelo solo nu, o que dificulta a sua caracterização espectral. Isto talvez justifique os 44% de erro de omissão cometidos em relação ao

levantamento realizado com fotografias aéreas.

Com relação ao erro de omissão de 64% no levantamento de 1978/79, deve-se acrescentar que, além da dificuldade de caracterizar parte da cana nova, também foi muito difícil a identificação de áreas de cana cortada, justamente pelo fato de se ter utilizado nesse levantamento somente uma data de passagem do LANDSAT.

TABELA 2

RESULTADOS DA ESTIMATIVA DE ÁREA OCUPADA COM CANA-DE-AÇÚCAR UTILIZANDO-SE OS MÉTODOS DE AMOSTRAGEM EM DUAS FASES E AMOSTRAGEM ALEATÓRIA SIMPLES, EM RELAÇÃO AO VALOR DE 278.500 ha OBTIDO ATRAVÉS DE FOTOGRAFIAS AÉREAS

LEVANTAMENTOS	ESTIMATIVA POR AMOSTRAGEM EM DUAS FASES			ESTIMATIVA POR AMOSTRAGEM ALEATÓRIA SIMPLES		
	ÁREA ESTIMADA EM ha	DIFERENÇA EM RELAÇÃO À VERDADE TERRESTRE		ÁREA ESTIMADA EM ha	DIFERENÇA EM RELAÇÃO À VERDADE TERRESTRE	
		ABSOLUTA (ha)	RELATIVA (%)		ABSOLUTA (ha)	RELATIVA (%)
1978/79	281.038	2.538	0,91	101.108	177.392	63,70
1979/80	278.465	35	0,01	155.525	122.975	44,16

#### 4. CONCLUSÕES

Com base nos resultados alcançados no presente trabalho, onde são utilizados dados combinados do LANDSAT e fotografias aéreas, num sistema de amostragem em duas fases para estimativa de área cultivada com cana-de-açúcar, chegou-se às seguintes conclusões:

- 1) A utilização do sistema de amostragem em duas fases mostrou-se adequada e eficiente. As diferenças ocorridas entre os levantamentos de 1978/79 e 1979/80 foram devidas ao aprimoramento do método de tratamento dos dados do LANDSAT no segundo ano, em que se utilizaram as características espectrais e temporais daqueles dados para caracterizar a cultura da cana-de-açúcar.
- 2) O sistema de amostragem em duas fases não pode ser substituído por um sistema de amostragem aleatória simples empregando-se somente dados do LANDSAT, a não ser que haja condições técnicas capazes de caracterizar a cultura da cana-de-açúcar em todos os seus estágios de desenvolvimento.
- 3) A utilização de apenas dados com fotografias aéreas para a realização de estimativa de área ocupada com cana-de-açúcar são adequados e precisos. Todavia, a sua frequente utilização para vastas regiões se

torna impraticável devido ao alto custo de tais levantamentos.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

- COCHRAN, W.G. Sampling Techniques 2<sup>nd</sup> ed., New York, J. Wiley and Sons, 413p, 1963.
- FEDERER, W.T. Experimental Design, Theory and Application, Oxford and IBH, New Delhi, 544p. 1967.
- JESSEN, R.J. Statistical Survey Techniques, New York, J. Wiley and Tom 520p. 1978.
- KOFFLER, N.F.; CAVALLI, A.G.; CHIARINI, J.V.; NOGUEIRA, F. de P. Inventário canavieiro com auxílio de fotografias aéreas. A Grande Região de Piracicaba no ano safra 1978/79. 1(2):3-38, dez. 1979 (série A).
- LOESTSH, F.; HALLER, K.E. Forest Inventory, vol. II, 2<sup>nd</sup> ed., BLV VERLAGS GESELLSCHAFT MUNCHEN BERN WIEN, GERMANY, 433p. 1973.
- MENDONÇA, F.J.; LEE, D.C.L.; SHIMABUKURO, Y. E.; TARDIN, A.T.; NOVAES, R.A.; CHEN, S.C. Utilização de dados do LANDSAT para inventário da cana-de-açúcar do Estado de São Paulo. São José dos Campos, INPE, jan, 1980 (INPE-1668-NTE/157).
- MENDONÇA, F.J.; LEE, D.C.L.; TARDIN, A.T.; SHIMABUKURO, Y.E.; CHEN, S.C.; LUCHT, L.A.M.; MOREIRA, M.A. de; MAIA, F.C.S. Levantamento da área canavieira do Estado de São Pau

lo, utilizando dados do LANDSAT no ano sa  
fra 1979/80. São José dos Campos, INPE,  
mar, 1981. (INPE-2021-RPE/288).

