

## Estimativa da área de cana-de-açúcar no estado de São Paulo por meio de sistema de amostragem de área e imagens de sensoriamento remoto

Mauricio Alves Moreira<sup>1</sup>  
Viviane Gomes Cardoso Rosa<sup>1</sup>  
Rodrigo Rizzi<sup>1</sup>  
Bernardo Friedrich Theodor Rudorff<sup>1</sup>  
Luciana Miura Sugawara Berka<sup>1</sup>  
Valdete Duarte<sup>1</sup>

Instituto Nacional de Pesquisas Espaciais – INPE  
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil  
{mauricio,vivian, rizzi, bernardo, lmiura, bernardo, valdete} @ltid.inpe.br

**Abstract.** The present work has the objective to use a sampling square method to estimate sugarcane crop area in São Paulo State, Brazil. A sugarcane thematic map obtained from multitemporal Landsat images classification was used as reference data and to stratify the State area into eight sugarcane area densities (0-13, 13-21, 21-31, 31-54, 54-74, 74-85, 85-95 and >95%) within 1 x 1 km segments. A probabilistic technique was used to determine four sample rates representing 0.07, 0.1, 0.3 and 0.4% of the study area which were randomly sampled one hundred times. The sugarcane area for each sample was evaluated based on the reference data map. The one hundred estimates for each sample rate were then compared with the reference data for the entire study area. Best results were obtained for the 0.3% sample rate (412 segments) where the relative difference of the area estimation was around  $\pm 2\%$  with a coefficient of variation lower than 2%, indicating that this method is not only suitable to accurately estimate sugarcane at State level but it is also an appropriate alternative for early forecast or when cloud free images are not available.

**Palavras-chave:** remote sensing, sample techniques, area estimation, sensoriamento remoto, técnicas de amostragem, estimativa de área.

### 1. Introdução

Com o lançamento dos satélites de observação da Terra e o avanço na área da informática, em especial nos Sistemas de Informação Geográfica (SIG), houve um grande impulso no desenvolvimento de metodologias que visam fornecer estatísticas agrícolas. No Brasil, vários estudos foram realizados ao longo das últimas três décadas que comprovam o potencial das imagens de sensoriamento remoto e dos SIGs para estimar áreas de culturas agrícolas. Dentre as culturas mais estudadas pode-se citar: cana-de-açúcar (Batista et al. 1990, Rudorff et al., 2004), trigo (Batista et al., 1990), arroz (Moreira e Crossetti, 1988), feijão (Rudorff e Krug, 1986; Formaggio, 1989), milho (Moreira et al. 1998; Derenne et al. 2001), soja (Antunes et al., 1993; Moreira et al., 1998; Derenne et al., 2001), café (Batista et al., 1990; Tardin et al., 1992).

No entanto, a principal limitação do emprego de imagens de satélite é, sem dúvida, a disponibilidade de imagens livre de cobertura de nuvens. Uma situação de contorno para esse problema é aumentar o número de satélites para imagear a superfície terrestre com maior frequência temporal; porém, isto demanda altos investimentos. Uma alternativa é o uso de métodos que consideram o emprego conjunto de imagens de satélites, técnicas de geoprocessamento e métodos estatísticos de amostragem de área tal como é feito em diversos países desenvolvidos. Por exemplo, os Estados Unidos utilizam métodos estatísticos para estimar a área e a produtividade agrícola desde 1938. Exemplos mais recentes do uso experimental de novas tecnologias nos Estados Unidos são os projetos *Crop Identification Technology Assessment*

by Remote Sensing (CITARS), Large Area Crop Inventory Experiment (LACIE) e Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing (AGRISTARS), conforme relatado por Chen (1980).

No Brasil, o INPE em parceria técnico-científica com a UNICA (União da Agroindústria Canavieira de São Paulo) e o CEPEA (Centro de Estudos Avançados em Economia Aplicada) deram início em 2003 a um projeto de mapeamento da área canavieira do estado de São Paulo, por meio de imagens de sensoriamento remoto a fim de estimar a área plantada durante um período de cinco anos safra a partir da safra de 2003/04 para o qual foi possível adquirir imagens do Landsat para todo estado de São Paulo em cinco datas de passagem do satélite, em períodos específicos dentro do ciclo da cultura. Isto permitiu gerar um mapa com a distribuição das áreas canavieiras ano safra 2003/04 (Rudorff et al., 2004). Apesar do êxito alcançado nesse projeto no que se refere ao mapeamento e estimativa da área canavieira no estado de São Paulo em virtude da disponibilidade de imagens livres de nuvens em período favoráveis para a identificação desta cultura foi verificada a necessidade de se dispor de um método alternativo que garanta o fornecimento da estimativa da área em tempo hábil e não dependa unicamente de imagens atuais que eventualmente possam estar com presença de nuvens impedindo a identificação da cana.

Atendendo a esta necessidade o presente trabalho tem como objetivo principal, estabelecer um procedimento de previsão da estimativa de área de cana por meio de técnicas de amostragem probabilística de área através do método estatístico de expansão direta, testando procedimentos computacionais de estratificação e alocação da amostra.

## 2. Material e método

### 2.1 Área de Estudo

A área de estudo abrangeu todos os municípios produtores de cana de açúcar do estado de São Paulo, conforme mostrado na **Figura 1**.

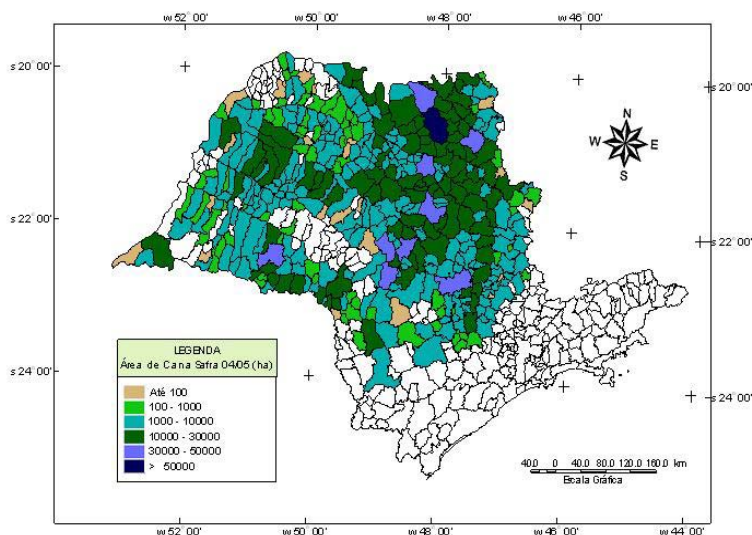


Figura 1 – Municípios do estado de São Paulo produtores de cana-de-açúcar que foram utilizados como área de estudo.

Fonte: Rudorff et al. (2004).

Para o ano safra 2003/04, a produção de cana em São Paulo correspondeu a 61,35% da produção nacional e a 69,52% da produção na Região Centro-Sul (UNICA, 2004). Os Municípios do estado de São Paulo produtores de cana de açúcar são cobertos por 15 cenas do TM/Landsat-5 das quais 4 cenas são completas e 11 parciais.

## 2.2 Estratificação da área

A estratificação tem como finalidade dividir uma região territorial em estratos homogêneos segundo a porcentagem de área do tema de interesse de tal forma que a variância dentro do estrato seja a menor possível e entre estratos seja maximizada (Krug e Yanasse, 1986). Embora existam diferentes abordagens de estratificação, nesse trabalho os estratos foram criados em função da porcentagem de área de cana dentro das unidades amostrais. Empregou-se como unidade amostral, segmentos de 1 km x 1 km, conforme sugerido por Adami (2004). Adotou-se como critério gerar estratos por meio do agrupamento de segmentos com porcentagens de área de cana menor ou igual a um limiar pré-estabelecido. Optou-se por essa abordagem, devido à existência de um mapa temático elaborado a partir da interpretação de imagens TM/Landsat, para o ano safra 2003/04 (Rudorff et al., 2003).

Por outro lado, o emprego do segmento na forma regular, segundo Gonzáles-Alonso et al. (1991), não interfere na precisão da estimativa da área. Além disso, facilita a interpretação do uso e ocupação do solo dentro do segmento. Para tal, gerou-se uma grade regular com dimensão de 1 km x 1 km, na forma matricial, por meio de um algoritmo em LEGAL. A matriz resultante foi posteriormente convertida em um dado cadastral, contendo as unidades amostrais.

Através da ferramenta *Estatística de Imagem por Polígono* calculou-se o percentual de área de cana em cada um desses segmentos de 1km<sup>2</sup>. Os resultados foram exportados para uma planilha eletrônica e gerado um histograma da distribuição de frequência acumulada. Esse histograma serviu como base para estratificar a área em oito estratos, segundo a porcentagem de cana no segmento. Após a estratificação da área, a etapa seguinte consistiu em determinar o tamanho mínimo de amostras, para estimar a área de cana no estado, dentro de uma confiabilidade de 95% de confiança. O tamanho mínimo foi calculado pela Equação 1, proposta por Cochran (1977). O tamanho mínimo da amostra foi de 103 segmentos.

$$n = \frac{(z_{\alpha/2})^2 pq}{E^2} \quad (1)$$

sendo,  $n$  o número total de segmentos a serem amostrados;  $z$  a distribuição normal padronizada;  $\alpha$  o nível de significância;  $p$  o percentual da área agrícola;  $q$  o percentual da área utilizada para outros fins e  $E$  o erro esperado para a estimativa.

Para determinar o número de segmentos por estrato, empregou-se a Equação 2 (Cochran, 1977):

$$n_h = \frac{N_h \sqrt{\frac{p_h q_h}{N_h}} / \sqrt{\frac{1}{p_h}}}{\sum^L (N_h \sqrt{\frac{p_h q_h}{N_h}} / \sqrt{\frac{1}{p_h}})} n \quad (2)$$

em que,  $n_h$  é o número amostral, em cada estrato;  $N_h$  o número total de segmentos no estrato;  $p_h$  o percentual da área do estrato cultivado com cana;  $q_h$  o percentual da área do estrato utilizado para outros fins e  $L$  o número de estratos.

Além do tamanho mínimo da amostra foram testados mais três novos tamanhos de amostras: 206, 412 e 618 unidades amostrais, para analisar o comportamento do coeficiente de variação.

Para testar a eficiência do delineamento amostral, realizou-se, através de um algoritmo no programa Matlab 100 sorteios de segmentos e extraiu de cada um deles os valores de áreas de cana para ser usado no modelo de expansão direta, conforme as Equações 3 a 6 (Krug e Yanasse, 1986; Allen, 1990; Pradhan, 2001):

$$\hat{Z}_c = \sum_{i=1}^m e_i \sum_{k=1}^{n_i} z_{c,ik} \quad (3)$$

sendo  $c$  o índice para representar a cultura de interesse, no caso a cana-de-açúcar;  $i$  o índice para representar o estrato de uso do solo com  $i = 1, \dots, m$ ;  $k$  o índice para representar o segmento amostrado com  $k = 1, \dots, n_i$ ;  $m$  é o número de estratos;  $n_i$  o número de segmentos amostrados no  $i$ -ésimo estrato de uso do solo;  $N_i$  o número total de segmentos no  $i$ -ésimo estrato de uso do solo;  $e_i = (n_i/N_i)^{-1}$  é o fator de expansão ou inverso da probabilidade de que um segmento tem de estar na amostra escolhida no  $i$ -ésimo estrato de uso do solo; e  $z_{c,ik}$  a área da cultura da cana, no  $k$ -ésimo segmento do  $i$ -ésimo estrato de uso do solo.

A variância estimada de  $\hat{Z}_c$  é:

$$v(\hat{Z}_c) = \sum_{i=1}^m N_i (N_i - n_i) \frac{s_{c,i}^2}{n_i} \quad (4)$$

em que  $s_{c,i}^2$  é a variância estimada da cultura  $c$  no estrato  $i$ , dada por:

$$s_{c,i}^2 = \frac{\sum_{k=1}^{n_i} (z_{c,ik} - \bar{z}_{c,i})^2}{(n_i - 1)} \quad (5)$$

e  $\bar{z}_{c,i}$  é a área média da cultura da cana no  $i$ -ésimo estrato de uso do solo:

$$\bar{z}_{c,i} = \frac{\sum_{k=1}^{n_i} z_{c,ik}}{n_i} \quad (6)$$

O Coeficiente de Variação (CV) é estimado por:

$$CV (\%) = \frac{\sqrt{\hat{v}(Z)}}{\hat{Z}} \times 100 \quad (7)$$

As diferenças percentuais entre as estimativas da área obtidas pela expansão direta e pela interpretação das imagens TM/Landsat-5 foram plotadas num gráfico, juntamente com os valores

percentuais do coeficiente de variação, para observar o comportamento dessas duas variáveis nos 100 sorteios realizados, em cada tamanho de amostra.

### 3. Resultado e discussões

Os dados contidos na **Tabela 1** referem-se às características dos delineamentos estabelecidos a partir das Equações 1 e 2 e do histograma de frequência acumulada.

Tabela 1 – Características do painel amostral

Estrato	% de cana	N*	Tamanho da amostra (n)			
			n = 103	n = 216	n = 412	n = 618
1	0 - 13	107.259	11	21	42	63
2	13 – 21	5.472	11	22	44	66
3	21 – 31	6.056	14	27	54	82
4	31 – 54	12.927	22	45	90	134
5	54 – 74	11.323	20	41	81	122
6	74 – 85	6.481	13	26	52	78
7	85 – 95	5.826	9	18	37	55
8	95 - 100	4.664	3	6	12	18
Total	-	160.008	-	-	-	-
Área Amostrada (%)**			0,07	0,1	0,3	0,4

\* número de unidades amostrais de 1km<sup>2</sup>;

\*\* Porcentagem amostrada em relação à área total de cana.

Os gráficos da **Figura 2a-d** mostram as diferenças relativas percentuais, entre a área estimada pela expansão direta e a área determinada por meio da interpretação das imagens TM/Landsat, para os 100 sorteios aleatórios e o coeficiente de variação (CV) para os tamanhos de amostras de 103, 206, 412 e 618 unidades.

Ao observar o comportamento das diferenças relativas e do CV, para a amostra de 103 segmentos (**Figura 2a**), nota-se que em 45 sorteios a área de cana foi subestimada, porém dentro do erro aceitável ( $\alpha=0,05$ ). Observa-se que o CV ficou entorno de 1 a 2%. A partir desse ponto as estimativas da área passaram a apresentar diferenças relativas positivas e o coeficiente de variação cresceu até os 5%.

No caso da amostra de 206 segmentos (**Figura 2b**) observa-se que as diferenças relativas foram menores e que em grande parte dos sorteios a área estimada pela expansão direta foi também subestimada, como aconteceu para a amostra de 103 segmentos. Para esse tamanho de amostra o CV foi menor, em torno de 1 a 3% e o incremento foi quase linear.

Para as amostras de 412 (**Figura 2c**) e 618 (**Figura 2d**) segmentos a estimativa da área de cana e o coeficiente de variação foram bastante semelhantes entre si, indicando que amostras de área com mais de 412 segmentos não acarretam melhoria na precisão da estimativa. Todavia, cabe salientar que o aumento do número de segmentos por amostra de área diminui a diferença relativa da área estimada e o valor do CV em relação à estimativa obtida pelo mapeamento global da área canavieira.

Ao se observar a porcentagem de área amostrada para produzir essas estimativas, nota-se que a amostra de maior tamanho (618 segmentos) não chegou a 0,5% da área canavieira do estado.

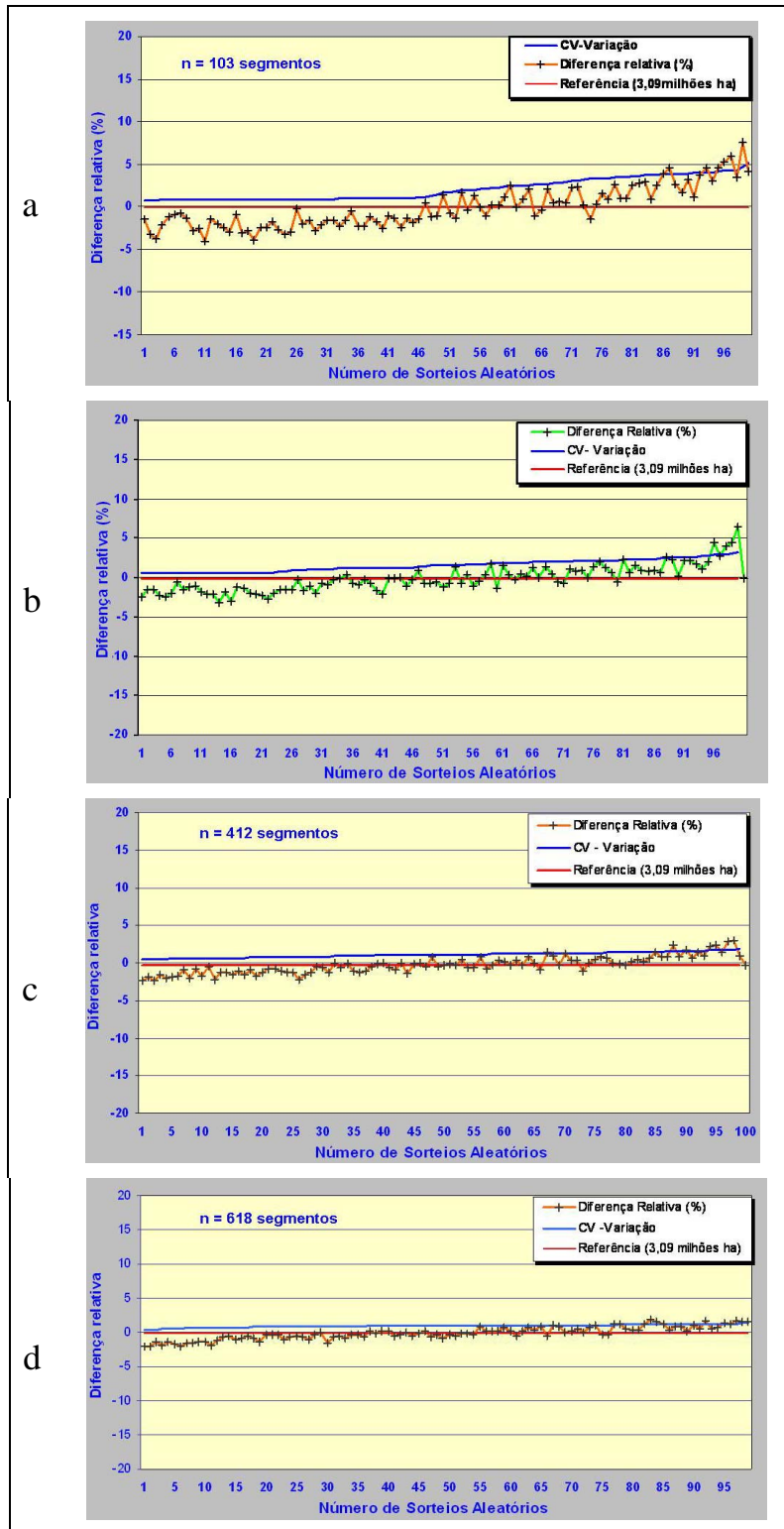


Figura 2 – Estimativas de área para os 100 sorteios aleatórios e 4 tamanhos de amostra.

Em outras palavras, o estabelecimento de um sistema amostral, não só produz uma estatística objetiva, mas também reduz o tempo e o custo do levantamento da estimativa além de poder ser realizado independente da disponibilidade de imagens atuais.

#### **4. Conclusões**

Tendo em vista a existência de um mapeamento recente da área de cana para o estado de São Paulo, algumas conclusões puderam ser tiradas a respeito dessa pesquisa. a) o critério de estratificar a área canavieira, com base na porcentagem de cana nas unidades amostrais, foi adequado porque produziu estimativa da área por meio da expansão direta dentro do intervalo de confiança de 95% e os valores do coeficiente de variação para os 100 sorteios ficaram abaixo de 5%, que foi o erro estabelecido. À medida que aumentou o número de segmentos por amostra a diferença relativa, entre as áreas estimadas pela expansão direta e pela interpretação das imagens do Landsat, diminuiu, porém amostras de área maior que 412 segmentos o incremento na precisão da estimativa da área foi muito pequeno. Para estimar a área de cana do estado, dentro da precisão desejada, foi necessária uma amostragem menor de 0,5% da área canavieira.

#### **Referências**

- Adami, M. **Estimativa de áreas agrícolas por meio de técnica de sensoriamento remoto, geoprocessamento e amostragem**. 2004. 185p. (INPE-10235-TDI/900). Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2004.
- Allen, J. D. A look at the Remote Sensing Applications Program of the National Agricultural Statistics Service. **Journal of Official Statistics**, v. 6, n. 4, p. 393-409, 1990.
- Antunes, M. A. H.; Batista, G. T.; Assad, E. D.; Moreira, M. A. Utilização de medidas espectrais para o monitoramento da cultura de soja (*Glycine max* (L.) Merrill) ao longo do ciclo de crescimento. In: Simpósio Brasileiro de Sensoriamento Remoto, 7., Curitiba, 1993. **Anais**. São José dos Campos: INPE, 1993. v. 4, p. 42-51.
- Batista, G. T.; Tardin, A. T.; Chen, S. C.; Dallemand, J. F. Avaliação de produtos HRV/SPOT e TM/LANDSAT na discriminação de culturas. **Pesquisa Agropecuária Brasileira**, v. 25, n. 3, p. 379-386, 1990.
- Chen, S. C. **Precisão de classificação do mapa de distribuição de culturas obtido a partir de imagens Landsat**. São José dos Campos, 1980. 7 p. (INPE-1889-RPE/232).
- Cochran, W. G. **Técnicas de Amostragem** 2 ed. Rio de Janeiro: Fundo de Cultura, 1977. 555p.
- Derenne, M.; Setzer, A. W.; França, H. O "Cinturão" de Soja e Milho em Mato Grosso nos mosaicos IVDN/AVHRR.[CD-ROM] In: Simpósio Brasileiro de Sensoriamento Remoto, 10., Foz do Iguaçu. 2001. **Anais**. São José dos Campos: INPE, 2001. Seção Poster, p. 41-9.
- Formaggio, A. R. **Características agrônômicas e espectrais para sensoriamento remoto de trigo e de feijão**. Piracicaba, 181 p. (INPE-5061-RPE/621) Tese (Doutorado em Sensoriamento Remoto) – USP/ESALQ - Escola Superior de Agricultura "Luiz de Queiroz", 1989.
- González-Alonso, F.; Soria, S. L.; Gozalo, J. M. C. Comparing two methodologies for crop area estimation in Spain using Landsat TM images and ground-gathered data. **Remote Sensing of Environment**, v.35, n.1, p 29-35, Jan. 1991.
- Krug, T.; Yanasse, C. C. F. **Estimativa de safras agrícolas utilizando dados coletados por satélites de sensoriamento remoto e dados terrestres, através de amostras de substratos geográficos**. São José dos Campos: INPE, 1986. 51 p. (INPE-4102-RPE/534).
- Moreira, M. A.; Crossetti, M. L. Arroz irrigado - identificação e mapeamento através de dados espectrais e temporais do TM- Landsat. In: Simpósio Latino Americano de Sensoriamento Remoto, 4., Gramado, 1986. **Anais**. São José dos Campos: INPE, 1988. v. 1, p. 65.

Moreira, M. A.; Aulicino, L. C. M.; Shimabukuro, Y. E.; Duarte, V.; Rudorff, B. F. T.; Yi, J. L. R.; Souza, I. M. Modelo de mistura espectral para identificar e mapear áreas de soja e milho em quatro municípios do Estado de Minas Gerais.[CD-ROM]. In: Simpósio Brasileiro de Sensoriamento Remoto, 9., Santos, 1998. **Anais**, São Paulo: Fabrica da Imagem Multimídia, 1998. Seção Poster

Pradhan, S., Crop area estimation using GIS, remote sensing and area frame sampling.. **International Journal of Applied Earth Observation and Geoinformation**, v. 3, n. 1, p. 86-92, 2001.

Rudorff, B. F. T.; Krug, T. Sensoriamento remoto na estimativa da área plantada com feijão, milho e mamona no município de Irece-BA. In: Simposio Latino-Americano de Sensoriamento Remoto, 6., Gramado. **Anais**. São José dos Campos: INPE, 1986. v. 1, p. 380-4.

Rudorff, B. F. T.; Berka, L. M. S.; Xavier, A. C.; Moreira, M. A.; Duarte, V.; Rosa, V. G. C.; Shimabukuro, Y. E. **Estimativa de área plantada com cana-de-açúcar em municípios do estado de São Paulo por meio de imagens de satélites e técnicas de geoprocessamento: ano safra 2003/2004**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2004. 47 p. (INPE-10791-RPQ/759).

Tardin, A. T.; Assunção, G. V.; Soares, J. V. Análise preliminar de imagens TM visando a discriminação de café, citrus e cana-de-açúcar na região de Furnas, MG. **Pesquisa Agropecuária Brasileira**, v. 27, n. 9, p. 1355-61, 1992.

UNICA - União da Agroindústria Canavieira de São Paulo. **Informação UNICA**, ano 6, n. 52, março/abril, São Paulo, 2003.