

## Metodologia computacional para criar um painel de amostra de área utilizando imagens de sensoriamento remoto e SIG

Viviane Gomes Cardoso Rosa<sup>1</sup>  
Mauricio Alves Moreira<sup>1</sup>  
Marcos Adami<sup>2</sup>

<sup>1</sup> Instituto Nacional de Pesquisas Espaciais - INPE  
Caixa Postal 515 - 12201-970 - São José dos Campos - SP, Brasil  
{vivian, mauricio}@ltid.inpe.br

<sup>2</sup> Secretaria de Estado da Agricultura e do Abastecimento  
Av. Minas Gerais,1351 – 86300-000 – Cornélio Procópio - Pr, Brasil  
adami@seab.pr.gov.br

**Abstract.** One of the most tedious activities in crop area estimation using probabilistic sample methods is the construction of a sample panel which used to be elaborated through visual image interpretation and some manual activities prior to its implementation in a Geographic Information System (GIS). The objective of this work is to use a much more automated, computerized and straight forward procedure to construct a sample panel in the municipality of Cristais, Minas Gerais State, Brazil, to estimate coffee crop area using the direct expansion method. Results showed that the procedure was adequate to provide fast, objective and low cost coffee crop area estimation at municipality level. A difference of 5,8% in coffee crop area estimation between direct expansion method and personal communication from the institution responsible for official agricultural statistics was found indicating that the proposed method is efficient but can be improved.

**Palavras Chaves:** sample area system, remote sensing image, crop area estimation.

### 1 Introdução

Desde o lançamento do primeiro satélite de recursos naturais, o Landsat-1, um novo sistema de coleta de dados passou a estar disponível para o levantamento de estatísticas agrícolas. As imagens fornecidas pelos satélites de sensoriamento remoto permitem extrair informações de áreas cultivadas em grandes extensões territoriais e, ao mesmo tempo, gerar mapas temáticos com a distribuição especial das áreas agrícolas a custo relativamente baixo, quando comparados às fotografias aéreas. Além disso, os satélites obtêm dados sistematicamente, o que possibilita monitorar as culturas ao longo do seu ciclo de crescimento e desenvolvimento.

Entretanto, o uso das imagens de satélites, nas regiões tropicais, é limitado pela alta incidência de cobertura de nuvens, principalmente, para as culturas de verão, época em que a nebulosidade é mais intensa. Nesse caso é importante que o método utilizado para geração de estatísticas agrícolas não seja totalmente dependente de imagens de satélite a fim de que possa ser aplicado mesmo não havendo disponibilidade de imagens livres de cobertura de nuvens.

Diante dessa limitação surgiram métodos estatísticos por amostragem de área, em que o painel amostral (estratificação, alocação dos segmentos da amostra, etc) é determinado a partir da interpretação de imagens de satélite que definem o uso e ocupação do solo com imagens disponíveis fora do período crítico de disponibilidade de imagens (FAO, 1998). A introdução de dados orbitais para construir o painel amostral, representou um avanço no aperfeiçoamento da metodologia de amostragem, uma vez que os estratos puderam ser melhor definidos, isto é, maior homogeneidade intra-estrato e maior heterogeneidade entre estratos. Por outro lado, a disponibilidade de imagens em diferentes épocas do ano, permite avaliar

adequadamente o uso e ocupação do solo além de fazer ajuste periódicos nos estratos o que seria muito oneroso se fosse realizado por meio de fotografias aéreas.

A construção do painel de amostra de área, ou seja, estratificação e alocação dos segmentos da amostra de área é uma fase demorada no processo de amostragem de área quando feito manualmente a partir de mosaicos de fotografias aéreas sobre os quais os limites dos estratos são transcritos para posterior alocação dos segmentos e realização de coleta de dados no campo. Um aspecto importante nesta etapa é que os segmentos, devem ser definidos de tal forma que os seus limites sejam facilmente identificados no campo. Com a disponibilidade dos GPS (Global Positioning System) os segmentos podem ser facilmente localizados no campo e não dependem mais de limites físicos podendo ser delimitados de forma regular.

Uma das principais vantagens do uso de imagens de satélites em relação às fotografias é a atualidade da informação dentro do segmento facilitando e aprimorando o levantamento da informação dentro dos segmentos por ocasião da coleta de dados no campo conforme relatam Collares et al. (1993).

Nos Estados Unidos foi implementado um sistema automático de estratificação e amostragem de área, denominado *Computer Aided Stratification and Sampling* (CASS) (Cotter e Tomczack, 1994; FAO, 1998). Este sistema foi utilizado durante muitos anos e recentemente foi substituído por um novo método que emprega os *softwares* ARCVIEW e ERDAS para a estratificação e amostragem das áreas.

No Brasil, um sistema automatizado para construção de um painel de amostra foi realizado por Adami (2004) visando estimar áreas de café, soja e milho na região de Cornélio Procópio, no Paraná, utilizado o aplicativo SPRING. Posteriormente essa metodologia foi empregada por Rizzi (2004) na estimativa da área de soja no Estado do Rio Grande do Sul.

O objetivo deste trabalho é descrever essa metodologia computacional de construção do painel de amostra além de mostrar uma aplicação na estimativa de área de café em nível municipal.

## **2. Metodologia**

Para aplicar e avaliar essa metodologia foi selecionado o município de Cristais, localizado no estado de Minas Gerais, região sudeste do Brasil, entre as coordenadas geográficas - 45° 24' 00''W e - 45° 42' 00''W e as latitudes - 20° 40' 40''S e - 20° 50' 57''S. Segundos dados do IBGE de 2002, o município de Cristais possuía uma área de 16.366 ha utilizados para agricultura, dos quais 8.804 ha com café equivalendo a aproximadamente 50% da área agrícola do município.

### **2.1 Mapeamento**

A primeira etapa do trabalho consistiu em gerar um mapa de uso do solo com as seguintes classes: água, pastagem, cidade, área agrícola e solo. O *software* utilizado foi o SPRING (Sistema de Processamento de Informações Georreferenciadas), versão 4.1 (INPE, 2003). Sobre as imagens inseridas no banco de dados, aplicaram-se os procedimentos de pré-processamento, ou seja, realce de imagens, registro e recorte da área de estudo. O georreferenciamento foi feito imagem-imagem, através do mosaico de imagens Landsat disponível na Internet (NASA, 2000). Posteriormente, gerou-se o mapa temático por meio de uma classificação não-supervisionada, em que numa fase inicial é feita uma segmentação da área. Para segmentar aplicou-se o segmentador por regiões, com limiares de similaridade e área de 5 e 20 pixels respectivamente. A segmentação foi realizada nas imagens das bandas 3, 4 e 5 do TM/Landsat-5 da data de 22/07/2004.

Em seguida aplicou-se a extração de regiões cuja finalidade foi rotular cada polígono obtido na segmentação. A partir de então, foi feita uma classificação não supervisionada, através do classificador ISOSEG, a fim de agrupar as regiões em classes espectralmente homogêneas. Essas classes foram superpostas às imagens e analisadas individualmente, sendo então associadas às classes temáticas de uso do solo. O último passo da classificação foi a interpretação visual, ou seja, interpretação na tela do computador, realizada por meio da ferramenta *Edição Matricial* existente no SPRING. A Edição Matricial é uma ferramenta interativa entre analista e o computador, que permite melhorar os resultados da classificação digital. O mapa temático final foi a base para o procedimento de criação do painel de amostra de área.

## 2.2 Criação do painel amostral

Essa fase consiste em estratificar a área e alocar a amostra de segmentos para obter os dados de uso e ocupação do solo. A metodologia empregada é descrita a seguir:

### 2.2.1 Estratificação da área

Seguindo a metodologia proposta por Adami (2004), para estratificar a área em estratos homogêneos, a primeira etapa foi criar um estrato com as classes de uso do solo onde não se esperava encontrar lavouras de café. Desta forma foram incluídas no estrato não amostral as classes temáticas: água e cidade. Posteriormente, dividiu-se o restante do mapa temático classificado em segmentos regulares de 1km<sup>2</sup>. Para tal, foi necessário gerar uma grade regular com dimensão de 1x1km, na forma matricial, por meio de um algoritmo em LEGAL (**Figura 1**). A matriz resultante foi posteriormente convertida em um dado cadastral, totalizando 531 unidades (segmentos) amostrais.

```
{
Tematico tem ("Xadrez");
tem = Novo (Nome="Xadrez_1x1", ResX=1000, ResY=1000, Escala=50000);
tem = Atribua {
    "al" : (tem [-1,0]!="al") && (tem [0,-1]!="al"),
    "ma" : Outros
};
}
```

Figura 1 – Geração de Grade Regular de 1 x 1km.

Através da ferramenta *Estatística de Imagem por Polígono* calculou-se o percentual de área agrícola em cada um dos segmentos. De acordo com os resultados obtidos, adotou-se como critério para estratificação, o percentual de área agrícola no segmento, agrupando a área em 5 estratos, de tal forma que eles apresentassem maior homogeneidade intra-estrato e maior heterogeneidade entre estratos.

### 2.2.2 Amostragem da Área

Após a estratificação da área, a etapa seguinte consistiu em determinar o número de amostras necessário em cada estrato. Foi estabelecido que a estimativa deveria apresentar um intervalo de confiança de 95%, com um erro amostral de 5%. A amostra de área foi determinada segundo a **Equação 1**, proposta por Cochran (1977).

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

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.

Através da **Equação 2** (Cochran, 1977) determinou-se o número de segmentos (unidades amostrais) para compor a amostra em cada estrato (subamostras):

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

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 café;  $q_h$  o percentual da área do estrato utilizado para outros fins e  $L$  o número de estratos.

Com intuito de verificar a eficiência do esquema de amostragem, determinado a partir da **Equação 2**, isto é, avaliar o comportamento da estimativa de área agrícola, foram realizados 100 sorteios aleatórios, para a amostra estabelecida que foi de 85 segmentos. Para a realização desses sorteios aleatórios, foi gerada no SPRING, uma consulta espacial em cada um dos estratos, cujos dados (ID e área agrícola) foram exportados para um arquivo ASCII. Em seguida, utilizou-se um algoritmo no programa Matlab para que os 100 sorteios fossem realizados automaticamente. Os dados de áreas agrícolas para cada um dos 100 sorteios e por subamostras foram empregados no modelo de expansão direta, seguindo as **Equações 3 a 5**, propostas por Cochran (1977), para estimar área agrícola no município. Os resultados das estimativas de área foram plotados num gráfico e comparados com a informação subjetiva obtida a campo. Além das estimativas da área plotou-se também os valores dos coeficientes de variação para cada um dos 100 sorteios, cuja finalidade foi observar o comportamento do erro de amostragem e selecionar o melhor esquema de amostragem para determinar a área de café.

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

Cálculo da variância estimada:

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

em que:  $\hat{Z}$  é a área total estimada da cultura;  $s_i^2$  é a variância estimada no estrato  $i$ ;  $k$  é o índice para representar o segmento escolhido;  $n_i$  é o número de segmentos escolhidos no  $i$ -ésimo estrato;  $N_i$  é o número total de segmentos no  $i$ -ésimo estrato;  $e_i = (n_i/N_i)^{-1}$ , fator de expansão ou inverso da probabilidade de que um segmento tem de estar na amostra escolhida no  $i$ -ésimo estrato;  $z_{i,k}$  é o valor da característica desejada, no  $k$ -ésimo segmento, no  $i$ -ésimo estrato.

Foi estimado também, o Coeficiente de Variação (CV) através da **Equação 5**:

$$CV = \frac{\sqrt{\hat{v}(Z)}}{\hat{Z}} \quad (5)$$

Sendo,  $\hat{v}(Z)$  a variância estimada da cultura, (**Equação 4**) e  $\hat{Z}$  a área total estimada da cultura, (**Equação 3**).

Uma vez estabelecidos as subamostras de área, a etapa seguinte foi realizar a importação dos 85 segmentos selecionados para o banco de dados. Para tanto, foi utilizado uma consulta espacial na coluna ID de cada um dos estratos. Através da *Edição Matricial* e novamente com o auxílio das imagens multitemporais, foi realizado a interpretação detalhada das classes existentes em cada um dos segmentos amostrados, identificando principalmente a cultura de café. Por meio da estatística de imagem por polígono foi possível calcular o percentual de área de café em cada um desses segmentos. Mais uma vez, por meio do método de expansão direta (**Equação 3**), estimou-se a área de café no município.

### 3 Resultados e Discussões

A **Figura 1** apresenta os resultados da classificação da área de estudo feita, inicialmente, para servir de base para estratificar a área em regiões homogêneas quanto ao uso do solo e para alocar as amostras de área para estimativa do café. A partir dessa classificação foi possível mapear a área de estudo em cinco classes de uso do solo: área agrícola (incluindo culturas de verão e áreas de café); mata (reserva e de galeria); água; área urbana e pastagem. A área agrícola mapeada correspondeu a 15.939 ha.

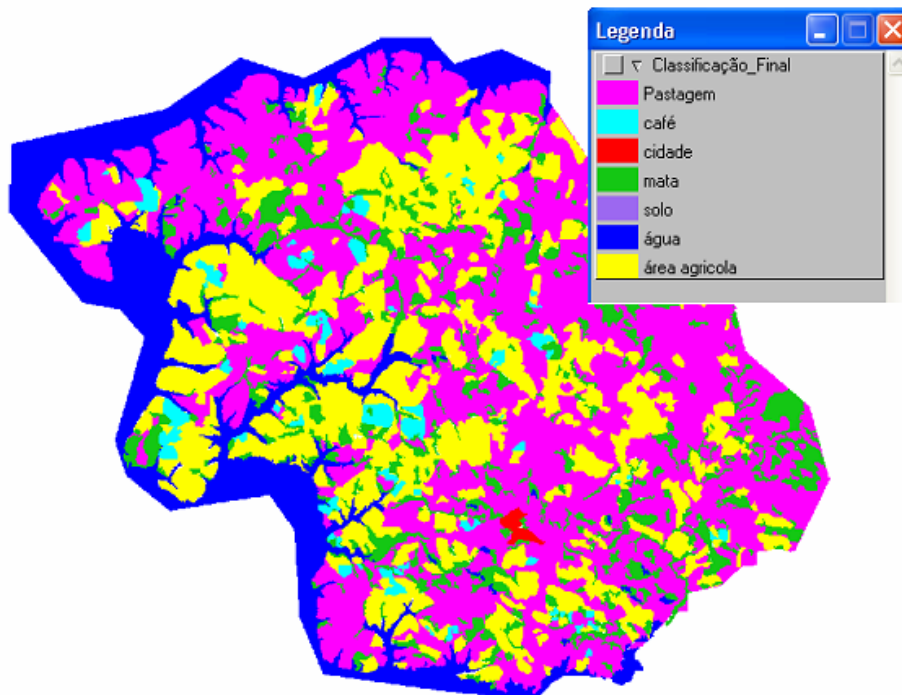


Figura 1 – Mapa Temático do município de Cristais.

A **Tabela 1** apresenta os limites adotados para cada um dos estratos, segundo a porcentagem de área agrícola no segmento.

Tabela 1 – Características dos estratos na área de estudo

ESTRATO	PERCENTUAL DE ÁREA AGRÍCOLA	Nº DE SEGMENTOS
1	0% a 5% de área agrícola nos segmentos	155
2	5% a 20% de área agrícola nos segmentos	89
3	20% a 40% de área agrícola nos segmentos	98
4	40% a 60% de área agrícola nos segmentos	82
5	60% a 100% de área agrícola nos segmentos	107
<b>TOTAL</b>		<b>531</b>

O número de unidades amostradas por substratos para compor a amostra de área agrícola pode ser visto na Tabela 2.

Tabela 2 – Número de unidades (segmentos) selecionados nos estratos

ESTRATO	PERCENTUAL DE ÁREA AGRÍCOLA (%)	Nº DE UNIDADE AMOSTRAL
Estrato 1	0 - 5	5
Estrato 2	5 – 20	15
Estrato 3	20 – 40	22
Estrato 4 %	40 – 60	22
Estrato 5	60 - 100	21
<b>TOTAL</b>		<b>85</b>
% de Seg. Amostrados		16%

O gráfico da **Figura 2** apresenta os resultados comparativos da estimativa da área agrícola e o comportamento do coeficiente de variação obtido em cada sorteio para as 100 repetições.

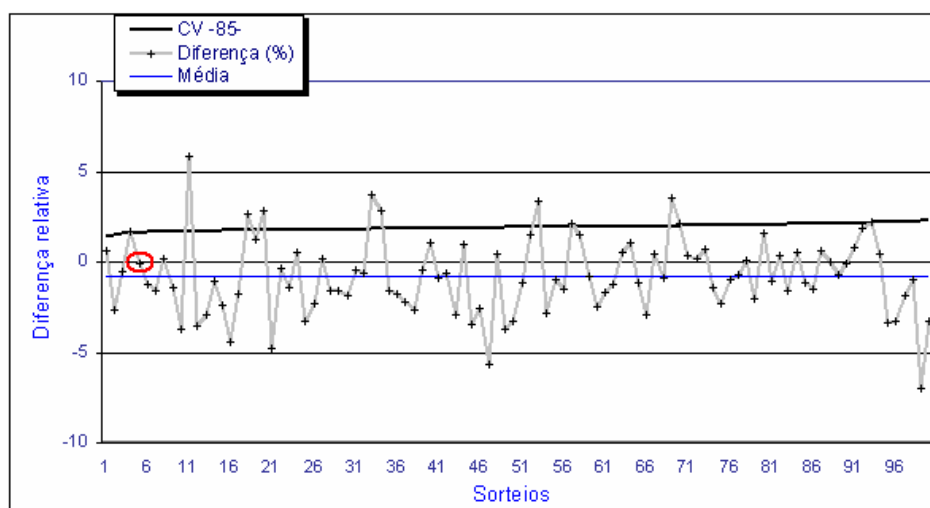


Figura 2 – Comportamento da estimativa da área agrícola para os 100 sorteios aleatórios com uma amostra de 85 segmentos.

A partir do gráfico determinou-se a melhor amostra de área, isto é, aquela cujos resultados apresentaram menor diferença em relação ao dado fornecido pela Empresa de Assistência Técnica e Extensão Rural (EMATER-MG), para ser novamente interpretada, a fim de extrair a área de café nos segmentos. Sendo assim, foram importados para o banco de

dados os segmentos que compreendiam o sorteio de nº 5. O resultado dessa importação pode ser visto na **Figura 3**.

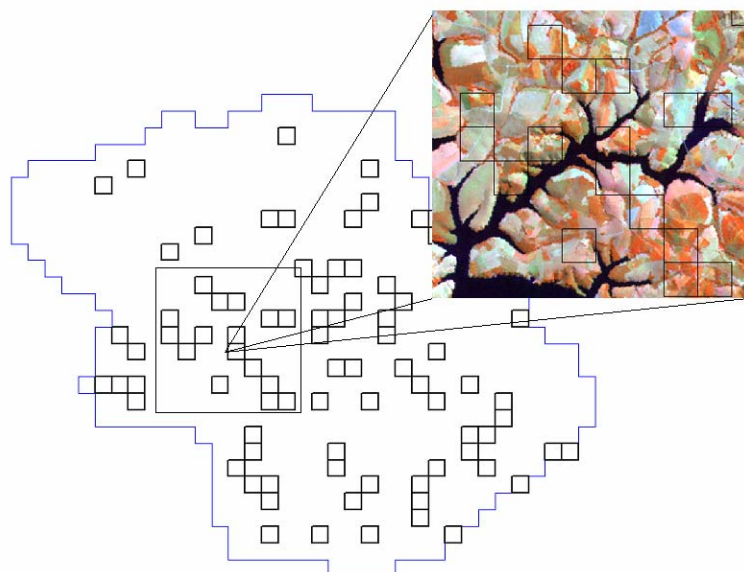


Figura 3 – Visualização e interpretação dos segmentos sorteados.

Realizou-se então, a interpretação visual na tela do computador dos segmentos amostrados, cujos valores de área de café foram aplicados no modelo de expansão direta (**Equação 3**). Os resultados da estimativa da área para essa amostra estão contidos na **Tabela 3**, juntamente com os dados subjetivos fornecidos pelo técnico da EMATER que atua no município.

Conforme pode-se verificar nos resultados da **Tabela 3**, o CV foi de 11,6 % que pode ser considerado bom quando comparado com os resultados obtidos por Adami (2004) que ao estimar a área de café em nível municipal encontrou um CV de 26%. Entretanto, acredita-se que esse resultado pode ser aprimorado através de um melhor ajuste do mapa de áreas agrícolas utilizado como base para estratificar a área.

Tabela 3 – Resultado da estimativa de área de café no município de Cristais, MG

CAFÉ (2004)	ÁREA
EMATER - 2004 *	7.043,20 ha
Estimado	7.450,94 ha
Diferença (ha)	407,74 ha
Diferença (%)	5,8%
CV (%)	11,6%

\* Comunicação pessoal

Por outro lado, nota-se que a diferença de 5,8 % na área estimada pelo método de expansão, em relação à informação pessoal do técnico da EMATER-MG é um resultado encorajador considerando que apenas 16% da área foi interpretada.

#### 4 Conclusões

De acordo com a abordagem metodológica e dos resultados obtidos pode-se concluir que o desenvolvimento de técnicas computacionais, para criar o painel amostral, é eficiente, pois eliminou uma série de procedimentos anteriormente realizados de forma manual como a

estratificação e a alocação do segmento. Além disso, a criação do painel amostral por métodos computacionais elimina a fase de partição da área de estudo em unidades primárias de amostragem (UPA), procedimento usado na metodologia visual. Outra vantagem dessa metodologia foi a possibilidade de se realizar a interpretação do uso do solo diretamente nas imagens, conferindo no campo apenas os pontos de dúvida.

Com relação aos resultados da pesquisa, pode-se dizer que os resultados foram satisfatórios apesar do coeficiente de variação de 11%. A diferença de 5,8% entre a área estimada por expansão e área de café estimada pela EMATER (dado de referência) pode ser considerada pequena em função da inexistência de um mapeamento de café para esta região uma vez que isto inviabiliza o agrupamento correto dos estratos segundo sua área. O resultado pode ser melhorado através de um ajuste da estratificação. Recomenda-se que seja feito um melhor ajuste do mapa de áreas agrícolas utilizado como base para estratificar a área.

Finalmente, conclui-se que a automatização da criação do painel de amostras e da alocação das amostras de área por meio do aplicativo SPRING é viável.

## Referências Bibliográficas

Adami, M. **Estimativa de áreas agrícolas por meio de técnica de sensoriamento remoto, geoprocessamento e amostragem**. 2004. 184p. (INPE-10235-TDI/900). Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2003.

Cochran, W.G. **Técnicas de Amostragem**. 2 ed. Rio de Janeiro: Fundo de Cultura, 1977. 555p.

Collares, J. E. R.; Lauria, C. A.; Carrilho, M. M. Pesquisa de Previsão e Acompanhamento de Safras baseada em Painéis de Amostras de Áreas. In: Simpósio Brasileiro de Sensoriamento Remoto, 7.. **Anais**. Curitiba: INPE, 1993. v.4, p.450-3.

Cotter, J.J. Tomezack, C. M. An Image Analysis System to Develop Área Sampling Frames for Agricultural Surveys. **Photogrammetric Engineering and Remote Sensing**, v. 60, n. 3, p. 229-306, Mar.1994.

Food and Agriculture Organization of the United Nations (FAO). **Multiple frame agricultural surveys: agricultural survey programmes based on area frame or dual frame (area and list) sample designs**. Roma, 1998. v. 2.

Instituto Nacional de Pesquisas Espaciais (INPE). **Sistema de Processamento de Informações Georreferenciadas (SPRING)**. São José dos Campos. Disponível em: <[www.dpi.inpe.br/spring](http://www.dpi.inpe.br/spring)>. Acesso em: 10 maio 2003.

National Aeronautics and Space Administration (NASA). **GeoCover** [on line]. Disponível em: <<https://zulu.ssc.nasa.gov/mrsid/>>. Acesso em 10 de Set, 2004.

Rizzi, R. **Geotecnologias em um sistema de estimativa da produção de soja: estudo de caso no Rio Grande do Sul**. 2003. 204p. Tese (Doutorado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2004.

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).