

Estimativa de área plantada com soja no Rio Grande do Sul através de amostragem por segmentos quadrados

Rodrigo Rizzi ¹
Bernardo Friedrich Theodor Rudorff ¹
Marcos Adami ²

¹ Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12201-970 - São José dos Campos - SP, Brasil
{rizzi, bernardo}@dsr.inpe.br

² Secretaria de Estado da Agricultura e do Abastecimento do Paraná - SEAB
Av. Minas Gerais, 1351, 86.300-000 - Cornélio Procópio - PR, Brasil
adami@seab.pr.gov.br

Abstract. This paper evaluates a sampling square method to estimate soybean crop area in Rio Grande do Sul State, Brazil. A soybean thematic map obtained from multitemporal Landsat images classification was used as reference data. The State area was divided into cells of 1 x 1 km of size and stratified into three soybean area densities (0-20, 20-40 and >40%) at municipality level. A probabilistic technique was used to determine four sample rates representing 0.06, 0.12, 0.24 and 0.48% of the study area which were randomly sampled one hundred times. The soybean 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 highest sample rate with low Coefficient of Variation (5.2), indicating that this method is not only suitable to accurate estimate soybean crop area at State level but it is also an appropriate alternative for early forecast or when cloud free images are not available.

Key words: sampling square method, crop area estimate, soybean, amostragem, estimativa de área plantada, soja.

1.Introdução

Ações governamentais para o controle eficiente das importações e exportações dos produtos agrícolas, tanto em respeito à balança comercial, quanto ao adequado abastecimento do mercado interno exigem informações constantes sobre a produção final desses produtos. No Brasil, as previsões de safras oficiais são realizadas por meio de informações municipais subjetivas, baseadas em opiniões de agentes técnicos e econômicos relacionados ao setor (Sano et al., 1998). Desta forma, o Levantamento Sistemático da Produção Agrícola (LSPA), fornece informações mensais referentes à estimativa de área plantada, produtividade média e produção de diversas culturas, a partir da fase de intenção de plantio até o final da colheita (IBGE, 2003). Entretanto, em função do seu caráter subjetivo, tais informações não permitem uma análise quantitativa dos erros envolvidos, além de serem passíveis de manipulação (Pino, 2001).

Dada a grande extensão territorial brasileira, aliada à dinâmica espaço-temporal da atividade agrícola, a maneira mais fácil de minimizar a subjetividade inerente às estatísticas oficiais é a introdução das chamadas *geotecnologias*, que envolvem imagens de sensoriamento remoto, Sistemas de Informações Geográficas (SIGs) e Sistemas de Posicionamento Global (GPS), às técnicas tradicionalmente utilizadas.

Uma das premissas necessárias em um sistema de levantamento de safras agrícolas é a estimativa da área plantada com as diferentes culturas presentes em uma determinada região (Pax-Lenney e Woodcock, 1997). A possibilidade da estimativa de área plantada de grandes culturas através de imagens de sensoriamento remoto já é pauta em pesquisas relacionadas ao tema (Fontana et al., 2001; Motta et al., 2001; Berka e Rudorff, 2003; Ippoliti-Ramilo et al.,

2003; Rizzi e Rudorff, 2003). A identificação de áreas ocupadas pelas culturas através da classificação de imagens de sensoriamento remoto parte do princípio que as culturas, em função de suas características intrínsecas, possuem comportamento espectral distintos, os quais são manifestados em uma imagem digital. Entretanto, tal método não possui caráter preditivo, visto que as imagens a serem utilizadas só serão adquiridas cerca de dois meses antes da colheita, sendo necessária, ainda, a realização dos procedimentos referentes à classificação propriamente dita, o que retarda a disponibilidade e reduz consideravelmente o valor das informações obtidas, para fins de previsão. Por conseguinte, faz-se necessária a utilização de uma técnica que forneça informações prévias a respeito da área plantada.

Adami (2003) sugere um método para estimativa de área plantada baseado na técnica de amostragem por segmentos quadrados. Nesse método, a área de estudo é dividida em segmentos regulares de 1 x 1 km. Alguns desses segmentos são aleatoriamente escolhidos, nos quais as culturas de interesse podem ter sua área quantificada por meio de visitas em campo por ocasião da implantação da cultura, dando à estimativa um caráter preditivo. Através de um método denominado *expansão direta*, esta informação é utilizada para se inferir a respeito da área plantada no restante da região em estudo. O autor utilizou tal método objetivando estimar a área plantada com café, milho e soja em uma região do norte do Paraná. Os resultados alcançados mostraram diferenças entre a estimativa de área plantada obtida pelo método empregado em comparação à gerada subjetivamente pelo Departamento de Economia Rural do Estado (DERAL). Para as culturas de café e milho, as diferenças relativas foram de 14,2 e -78,6%, com coeficientes de variação (CV) de 24,53 e 26,05%, respectivamente. Isto pode ser justificado pelo fato de essas culturas terem pouca expressão na região em estudo (2% para o café e 3% para o milho), o que diminui a precisão das estimativas. No caso da soja, que representa cerca de 30% da área, o método subestimou a área plantada em apenas 5,9% em relação ao DERAL, com o CV ficando em 6,59%. Apesar dos resultados encontrados, é difícil inferir sobre a exatidão do método proposto, pois não se pode assegurar que a estimativa subjetiva do DERAL corresponda à realidade do campo, em virtude da ausência de um dado de referência confiável.

Sendo assim, este trabalho objetivou avaliar o método de amostragem por segmentos proposto por Adami (2003), para a estimativa de área plantada com soja no Rio Grande do Sul, em nível estadual.

2. Materiais e Método

A área de estudo está localizada entre as latitudes S 27° 03' e S 30° 13' e as longitudes W 50° 40' e W 56° 20', abrangendo 322 municípios (117.628 km²), os quais correspondem a aproximadamente 90% da área plantada com soja no Rio Grande do Sul. O dado de referência utilizado neste trabalho foi um mapa temático produzido por Rizzi (2004), onde as áreas de soja foram mapeadas através da classificação automática e visual de imagens adquiridas pelos sensores TM e ETM+, no ano-safra 2000/01 (**Figura 1**). Como o método proposto sugere uma estratificação visando minimizar o número de segmentos a serem amostrados para otimizar o trabalho em campo, o mapa temático foi associado à divisão política municipal e estratificado considerando-se a porcentagem de área plantada com soja em cada município, em relação à sua área total (estratificação em nível municipal). Para tanto, os limiares estabelecidos para a definição dos estratos foram: 0-20% (*estrato 1*), 20-40% (*estrato 2*) e 40-67% (sendo 67% o limite superior) (*estrato 3*). Posteriormente, o mapa temático foi dividido em segmentos regulares de 1 x 1 km (unidades amostrais), os quais respeitaram os limites geodésicos dos estratos (**Figura 2**). Sendo assim, obtiveram-se 111.628 segmentos, distribuídos em: 60.917 no *estrato 1*; 22.542 no *estrato 2*; e 28.169 no *estrato 3*.

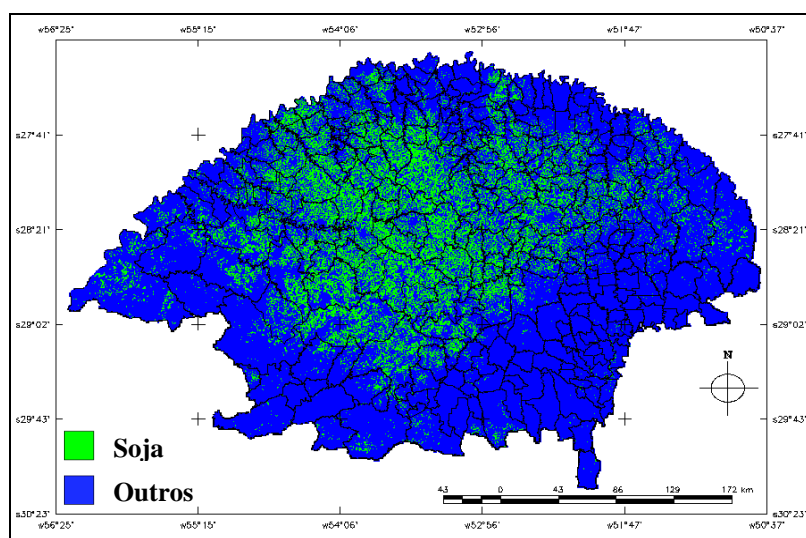


Figura 1. Mapa temático referente ao mapeamento das áreas com soja nas imagens Landsat (ano-safra 2000/01) e a divisão política municipal.

A **Figura 2** ilustra uma porção da área de estudo representada pelos três estratos, dividida em segmentos regulares de 1 x 1 km. As áreas vermelhas com formato irregular representam os talhões de soja.

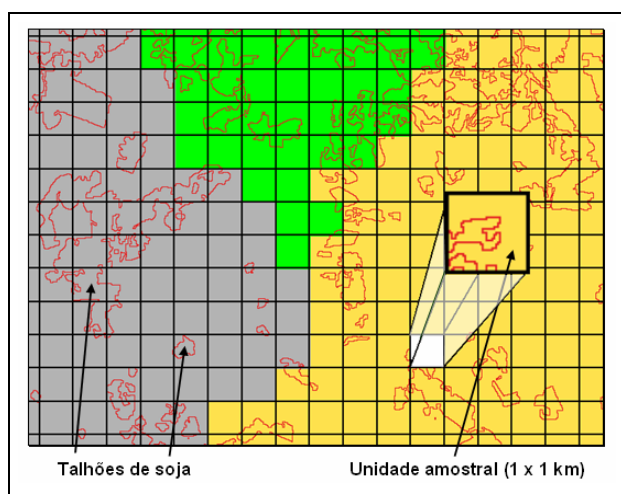


Figura 2. Divisão de parte da área de estudo em segmentos regulares de 1 x 1 km (unidades amostrais).

Em Seguida, determinou-se o tamanho amostral necessário para gerar uma estimativa com intervalo de confiança de 95% e um erro amostral de 5%, de acordo com a Equação 1, baseada em Cochran (1977):

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

onde,

n = número total de segmentos a serem amostrados;

z = distribuição normal padronizada;

α = nível de significância;

p = percentual da área cultivada com soja;

q = percentual da área utilizada para outros fins;

E = erro esperado para a estimativa.

O número amostral em cada estrato (subamostras), determinado pela Equação 2 (Cochran, 1977), é dado por:

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

onde,

n_h = número amostral, em cada estrato;

N_h = número total de segmentos no estrato;

p_h = percentual da área do estrato cultivado com soja;

q_h = percentual da área do estrato utilizado para outros fins;

L = número de estratos;

n = número total de segmentos a serem amostrados (Equação 1).

Deste modo, foram determinadas 26 amostras para o *estrato 1*, 74 amostras para o *estrato 2* e 165 amostras para o *estrato 3*, totalizando 265 amostras. Como o presente trabalho visa testar o método proposto por Adami (2003) no desenvolvimento de uma metodologia operacional para estimativa de área plantada, onde se busca minimizar o trabalho em campo sem, entretanto, comprometer a precisão da estimativa, optou-se por testar um número menor de amostras em cada estrato, o qual foi reduzido em 50 e 75%. Decidiu-se, também, duplicar o número amostral, obtendo-se, assim, 4 *conjuntos amostrais*, de acordo com a **Tabela 1**.

Tabela 1. Número de amostras distribuídas por estrato em cada conjunto amostral.

Conjunto amostral	1	2	3	4
Estratos (nº de municípios)	Número de amostras (amostras/município)			
Estrato 1 - 0-20% (176)	7 (0,04)	13 (0,07)	26 (0,15)	52 (0,30)
Estrato 2 - 20-40% (75)	18 (0,24)	37 (0,49)	74 (0,99)	148 (1,97)
Estrato 3 - 40-67% (71)	42 (0,59)	83 (1,17)	165 (2,32)	330 (4,65)
Total	67 (0,21)	133 (0,41)	265 (0,82)	530 (1,65)
Quantidade de segmentos amostrados	0,06%	0,12%	0,24%	0,48%

Para efeito de verificação do método, realizaram-se 100 sorteios aleatórios, para cada conjunto amostral, visando obter a estimativa da área plantada com soja em cada sorteio, totalizando 400 sorteios. Por exemplo, para o *conjunto amostral 1* foram escolhidos, aleatoriamente, 7 segmentos no *estrato 1*, 18 segmentos no *estrato 2* e 42 segmentos no *estrato 3*, perfazendo 67 segmentos de um total de 111.628 (procedimento efetuado 100

vezes). A **Figura 3** mostra o resultado da estratificação da área de estudo em 3 regiões e um exemplo de um sorteio com 530 amostras, representado pelos pontos escuros.

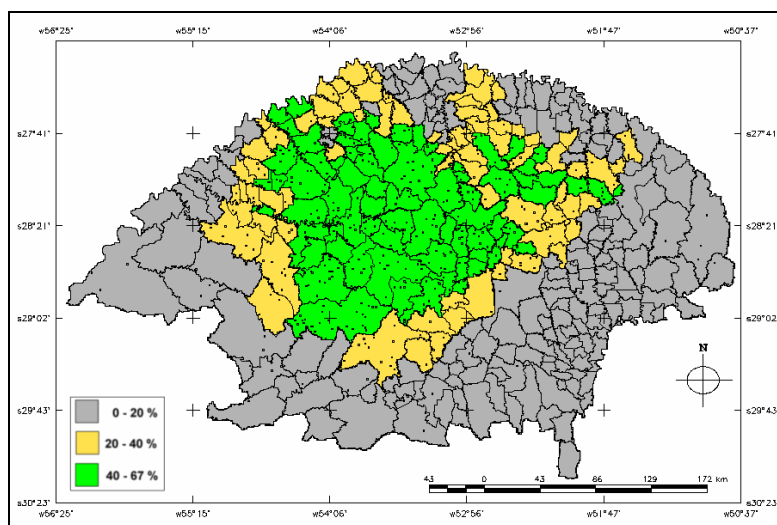


Figura 3. Estratos correspondentes à porcentagem de área plantada com soja em relação à área do município e exemplo de um sorteio com 530 amostras.

A estimativa da área plantada foi realizada através do método de expansão direta, na qual a área total estimada da cultura (\hat{Z}) é obtida pela Equação 3 (Cochran, 1977):

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

A variância estimada é:

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

onde,

s_i^2 = variância estimada no estrato i ;

i = índice para representar o estrato; $i = 1, \dots, m$;

k = índice para representar o segmento escolhido; $k = 1, \dots, n_i$;

m = número de estratos;

n_i = número de segmentos escolhidos no i -ésimo estrato;

N_i = 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;

$z_{i,k}$ = valor da característica desejada, no k -ésimo segmento, no i -ésimo estrato;

O Coeficiente de Variação (CV) estimado é dado pela Equação 5:

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

onde,

$\hat{v}(Z)$ = variância estimada da cultura, (Equação 4);

\hat{Z} = área total estimada da cultura, (Equação 3);

3. Resultados e discussão

A **Figura 4** (entre a e d) apresenta os resultados das estimativas de área plantada pelo método de amostragem, em ordem crescente de coeficiente de variação (CV), e da diferença relativa em relação ao mapa temático (dado de referência), além da média das estimativas, obtidos para cada um dos quatro conjuntos amostrais (67, 133, 265 e 530 amostras) nos 100 sorteios realizados.

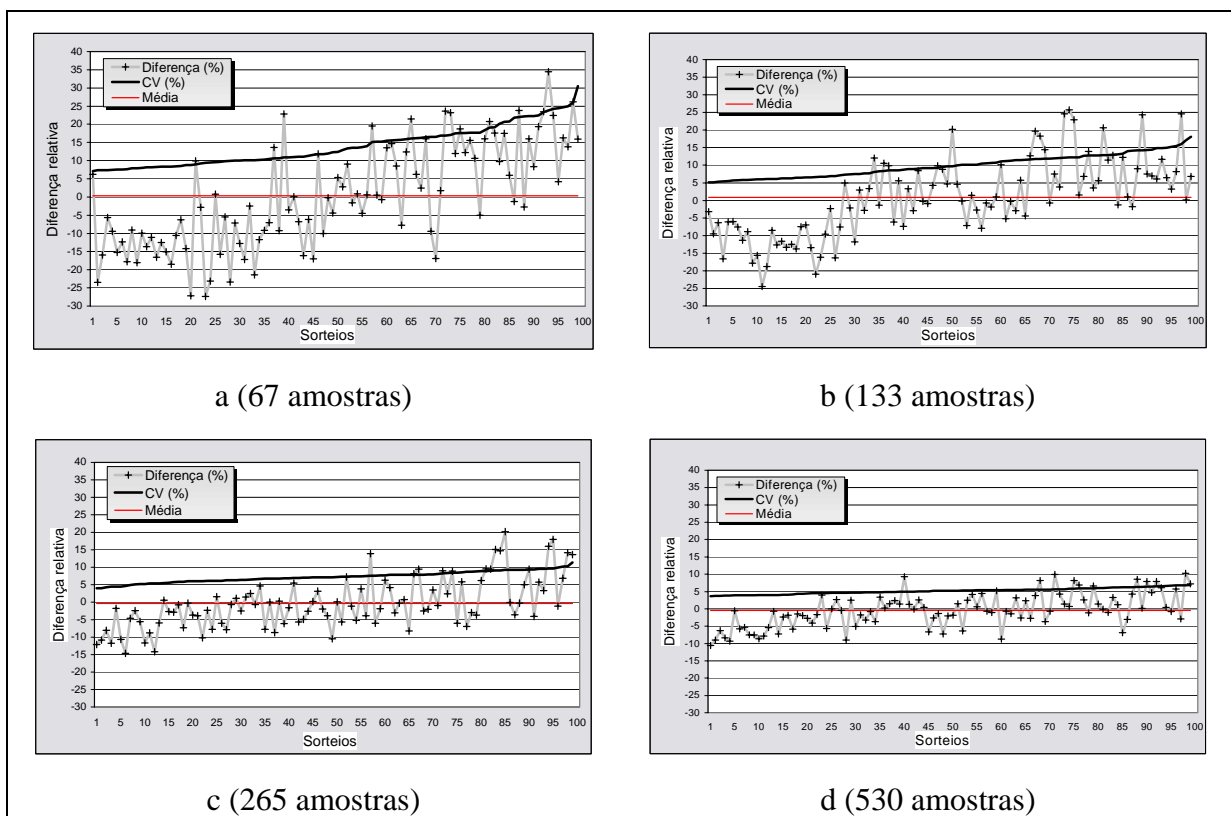


Figura 4. Diferenças percentuais da estimativa amostral em relação ao mapeamento das imagens landsat e coeficientes de variação correspondentes aos 100 sorteios realizados para os *conjuntos amostrais* de 1 a 4 (67 a 530 amostras).

A análise da **Figura 4** permite concluir que à medida que aumenta o número amostral diminui a amplitude de variação e a magnitude das diferenças relativas das estimativas em relação ao mapa temático, o mesmo ocorrendo em relação aos valores e, conseqüentemente, da média dos CVs. Observa-se ainda a grande variação entre as estimativas dos diferentes

sorteios em um mesmo conjunto amostral, muito embora a média das estimativas dos 100 sorteios tenha se aproximado muito do mapa temático, em todos os conjuntos amostrais (67, 133, 265 e 530 segmentos).

Ainda assim, não se constatou efeito do CV sobre a precisão das estimativas para um mesmo conjunto amostral, já que as maiores diferenças (em módulo) estiveram associadas tanto a altos, quanto a baixos valores de CV, o mesmo ocorrendo com as menores diferenças. O que se verificou, de fato, foi uma tendência dos menores valores de CV corresponderem às diferenças negativas (subestimativas), enquanto que os maiores CVs estarem relacionados às diferenças positivas (superestimativas). Sendo assim, é fácil compreender porque as estimativas mais próximas ao dado de referência estão associadas aos valores intermediários de CV.

Ainda em relação ao CV, cabe mencionar que o ângulo de inclinação da sua reta de distribuição diminui do *conjunto amostral 1* para o *conjunto amostral 4* (**Figura 4**). Isto significa que a relação entre o valor máximo e o mínimo diminui consideravelmente, o que provoca certa estabilização dos valores à medida que se aumenta o número amostral, conforme visualizado na **Tabela 2**. Isto converge para o fato que o número amostral ideal para a estimativa em questão pode estar em torno de 530.

A despeito das assertivas anteriores, ao se adotar um sistema de amostragem desta natureza visando à determinação da área plantada com determinada cultura em uma região, a principal referência disponível para se inferir sobre a precisão da estimativa é o CV, o qual está intimamente relacionado ao número amostral utilizado. Contudo, a escolha do número amostral deve considerar, primordialmente, a viabilidade da execução do método (ou seja, da avaliação da área plantada nos segmentos escolhidos) de uma forma economicamente viável e em tempo oportuno, o qual na maioria dos casos será realizado através de visitas a campo, auxiliadas ou não, por imagens de sensoriamento remoto.

No presente estudo, para o conjunto amostral contendo 530 amostras, as diferenças relativas da estimativa via amostragem frente ao mapa temático variaram entre -10,53 e 10,28% (64% dos CVs ficaram entre -5 e 5%), sendo que os CVs oscilaram entre 3,63 e 6,88% (43% dos CVs foram menores que 5%) (**Tabela 2**). Além disso, os 530 segmentos representam tão somente 0,46% da área estudada, abrangida por 322 municípios, totalizando menos de 2 segmentos por município (**Tabela 1**), o que deve tornar viável a execução prática do método. Para os demais conjuntos amostrais, a magnitude das incertezas encontradas, tanto em termos de diferença relativa, quanto no que se refere ao CV (principalmente para 67 e 133 amostras), torna sua aplicação pouco indicada.

Por fim, o que fica evidenciado no sistema proposto ante os resultados encontrados, particularmente no que se refere ao *conjunto amostral 4*, é que o mesmo, além de proporcionar uma idéia da real área plantada, fornece o erro associado à estimativa, o que não ocorre em se tratando dos dados oficiais fornecidos pelo LSPA.

Tabela 2. Variação das diferenças e dos CVs e média das diferenças e dos CVs para os 100 sorteios nos diferentes conjuntos amostrais das estimativas de área plantada através do método de amostragem.

Conjunto amostral	Variação das diferenças (%) (% entre - 5 e 5%)	Média das diferenças (%)	Variação dos CVs (%) (max/min)	Média dos CVs (%)
1 (67 amostras)	-27,35 a 34,47 (20)	0,36	7,10 a 30,48 (4,3)	13,96
2 (133 amostras)	-24,49 a 25,72 (33)	0,87	5,13 a 18,12 (3,5)	9,88
3 (265 amostras)	-14,65 a 20,19 (51)	-0,34	3,96 a 11,36 (2,9)	7,27
4 (530 amostras)	-10,53 a 10,28 (64)	-0,41	3,36 a 6,88 (1,9)	5,19

4. Conclusão

A avaliação do método de amostragem por segmentos quadrados proposto por Adami (2003) indicou que o mesmo pode fornecer informações a respeito da área plantada com soja no Rio Grande do Sul com um baixo CV desde que seja amostrado um número suficiente de segmentos por estrato. No caso específico do presente estudo, foi observado que o número amostral ideal para a estimativa em nível estadual pode estar em torno de 530.

5. Referências bibliográficas

- Adami, M. **Estimativa de áreas agrícolas por meio de técnica de sensoriamento remoto, geoprocessamento e amostragem**. 2003. 183 p. (INPE-10235-TDI/900). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2003.
- Berka, L.M.S.; Rudorff, B.F.T. Estimativa de área plantada com soja através de imagens Landsat em municípios do norte do Paraná. In: Simpósio Brasileiro de Sensoriamento Remoto, 11., Belo Horizonte. 2003. **Anais...** São José dos Campos: INPE, 2003. p. 27-31. Disponível na biblioteca digital URLib <lagavulin.ltid.inpe.br:1905/col/ltid.inpe.br/sbsr/2003/03.27.11.57/doc/01_469.PDF> Acesso em: 3 nov. 2004.
- Cochran, W. G. **Técnicas de amostragem**. 2 ed. Rio de Janeiro: fundo de Cultura, 1977. 555p.
- Fontana, D.C.; Ducati, J.R.; Weber, E.; Berlato, M.A.; Bergamaschi, H.; Figueiredo, D.C. **Monitoramento e Previsão da Safra de Soja 1999/2000 no Brasil**. Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, CEPARM/UFRRGS. 2001, 116 p. (Série D: Relatório Técnico - N° 005/01).
- IBGE - Instituto Brasileiro de Geografia e Estatística. 2003. Disponível em: <www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/lspa/>. Acesso em: 7 nov. 2003.
- Ippoliti-Ramilo, G.A.; Epiphanyo, J.C.N.; Shimabukuro, Y.E. Landsat-5 Thematic Mapper data for pre-planting crop area evaluation in tropical countries. **International Journal of Remote Sensing**, v. 24, n. 7, p. 1521-1534, 2003.
- Motta, J.L.G.; Fontana, D.C.; Weber, E. Verificação da acurácia da estimativa de área cultivada com soja através de classificação digital em imagens Landsat. In: Simpósio Brasileiro de Sensoriamento Remoto, 10., Foz do Iguaçu. 2001. **Anais...** São José dos Campos: INPE, 2001. p. 123-129. Disponível na biblioteca digital URLib <150.163.34.2:1908/rep/dpi.inpe.br/lise/2001/09.13.11.01> Acesso em: 4 set. 2004.
- Sano, E.E.; Assad, E.D.; Orioli, A.L. Monitoramento da ocupação agrícola. In: Assad, E. D.; Sano, E. E. **Sistemas de informação geográfica: Aplicações na agricultura**. 2 ed. Brasília. Embrapa-CPAC, 1998, Cap. 10, p. 179-190.
- Pax-Lenney, M.; Woodcock, C.E. The effect of spatial resolution on the ability to monitor the status of agricultural lands. **Remote Sensing of Environment**, v.61, n. 2, p. 210-220, 1997.
- Pino, F.A. Estimativa subjetiva de safras agrícolas. **Informações Econômicas**, v. 31, n. 6, São Paulo, 2001.
- Rizzi, R. **Geotecnologias em um sistema de estimativa da produção de soja**: estudo de caso no Rio Grande do Sul. 2004-04-30. (INPE--TDI). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2004.
- Rizzi, R.; Rudorff, B.F.T. imagens Landsat na estimativa de área plantada com soja em municípios do Rio Grande do sul. In: Simpósio Brasileiro de Sensoriamento Remoto, 11., Belo Horizonte. 2003. **Anais...** São José dos Campos: INPE, 2003. p. 231-238. Disponível na biblioteca digital URLib <http://lagavulin.ltid.inpe.br:1905/col/ltid.inpe.br/sbsr/2003/03.27.11.57/doc/01_411.PDF> Acesso em: 8 out. 2004.