

## **Estimativa da área de soja no Rio Grande do Sul por meio de amostragem aleatória estratificada de pontos**

Marcos Adami <sup>1</sup>

Rodrigo Rizzi <sup>1</sup>

Bernardo Friedrich Theodor Rudorff <sup>1</sup>

Mauricio Alves Moreira <sup>1</sup>

<sup>1</sup> Instituto Nacional de Pesquisas Espaciais - INPE  
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil  
{adami, rizzi, bernardo, mauricio}@dsr.inpe.br

**Abstract:** This work has the objective to identify the best sample size based on points to estimate soybean planted area in Rio Grande do Sul State. Samples were randomly selected according to a stratification based on percentage of planted area within each municipality. A comparative evaluation for the estimates from six different sample sizes with three values of  $\alpha/2$  (90, 95 e 99%) and two values of E (5% e 2.5%) was carried out. In order to verify the accuracy of the method 400 random drawing were performed for each sample size. Results indicated that best estimates were obtained for sample size greater or equal to 1,152 points.

**Palavras-chave:** remote sensing, sampling, agricultural statistics, sensoriamento remoto, amostragem, estatísticas agrícolas.

### **1. Introdução**

O estabelecimento de um sistema de previsão de safra capaz de estimar, em tempo hábil e de forma confiável, a produção das principais culturas agrícolas é importante para a correta tomada de decisões sobre a importação e exportação de produtos agrícolas (Adami et al., 2005; Rizzi et al., 2006). O Instituto Brasileiro de Geografia e Estatística (IBGE) e a Companhia Nacional de Abastecimento (CONAB) são os órgãos oficiais responsáveis pela estimativa da previsão da safra a qual é determinada de forma subjetiva com base nas informações e opiniões de agentes técnicos e econômicos do setor agrícola. Segundo Pino (2001), devido ao caráter subjetivo tais estimativas não permitem uma análise quantitativa dos erros envolvidos, além de serem passíveis de manipulação.

Uma das maneiras de reduzir a subjetividade do processo de estimativa de área plantada de grandes culturas é integrar imagens de Sensoriamento Remoto (SR) e Sistemas de Informações Geográficas (SIGs) aos atuais métodos de estimativa de safra (Fontana et al., 2001; Berka & Rudorff, 2003; Rizzi & Rudorff, 2003). Entretanto, devido à grande extensão territorial brasileira, à dinâmica espaço-temporal da atividade agrícola e ao problema relacionado com a obtenção de imagens livres de nuvens, a abordagem sugerida pelos autores não fornece estimativas com caráter prognóstico, uma vez que a identificação e o mapeamento das culturas nas imagens de satélite, na maioria dos casos, somente poderá ser feito em fase avançada do ciclo fenológico das culturas, o que retarda a disponibilidade da informação e reduz, consideravelmente, o seu valor estratégico para fins de negociação no mercado globalizado. Por conseguinte, faz-se necessária a utilização de uma abordagem que forneça informações prognósticas a respeito da área plantada. Neste sentido, o emprego de técnicas de amostragem para gerar estimativas de populações é amplamente utilizado, principalmente em função de questões que envolvem tempo de obtenção de dados (Adami et al., 2005; Rizzi et al., 2005). Exemplos de métodos de amostragem que utilizam sensoriamento remoto e Sistema de Informação Geográfica (SIG) para estimativas de safras, em diversos países, foram relatados por FAO (1998) e Gallego (2004).

Neste sentido, o presente trabalho tem por objetivo avaliar um método de amostragem aleatória estratificada, por pontos, para estimar a área plantada com soja no Estado do Rio Grande do Sul, para o ano safra 2000/01.

## 2. Material e Métodos

A área de estudo situa-se entre as latitudes S 27° 03' e S 30° 13' e as longitudes W 50° 40' e W 56° 20', abrangendo 322 municípios (111.628 km<sup>2</sup>). Estes municípios representam aproximadamente 90% da área plantada com soja no Estado do Rio Grande do Sul no ano-safra 2000/01.

A área de estudo foi estratificada com o intuito de minimizar o número de pontos amostrados sendo calculado o percentual da área ocupada pela soja em cada município em relação à sua área total (estratificação em nível municipal). Com base no percentual de área de soja, foi feita a estratificação, cujos limiares estabelecidos foram: 0-20% (estrato 1), 20-40% (estrato 2) e 40-67% (sendo 67% o limite superior; estrato 3). Para determinar o tamanho da amostra foi empregada a Equação 1 proposta por Thompson (2002):

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

em que  $n$  é o número total de segmentos amostrados para estimar a área com certo percentual de confiança ( $\alpha/2$ ) e um erro de amostragem ( $E$ );  $p$  é o percentual de área de soja dentro do estrato;  $q$  é a área utilizada com outras ocupações do solo, dentro do estrato, representado por  $(1-p)$ , e  $Z$  é o valor crítico tabelado da função normal padrão, para o intervalo de confiança ( $\alpha/2$ ).

Foram calculados seis tamanhos de amostras que combinaram três valores de  $\alpha/2$ , 90, 95 e 99% e 2 valores de  $E$ , 5% e 2,5%. Para verificar a exatidão do método foram realizados 400 sorteios aleatórios para cada tamanho amostral, visando obter a estimativa da área plantada com soja em cada sorteio. Este procedimento permite escolher o conjunto amostral que fornece estimativas compatíveis com a exatidão requerida com o menor número de amostras a fim de minimizar o custo do trabalho a campo sem contudo diminuir a exatidão requerida. A divisão de elementos por estrato ( $nh$ ) foi determinada pela Equação 2:

$$nh = \frac{Nh \cdot Sh}{\sum_{i=1}^m \frac{Nh \cdot Sh}{\sqrt{Ch}}} \cdot n \quad (2)$$

em que  $Nh$  é o número total de elementos do estrato;  $Sh$  é o desvio padrão do estrato;  $Ch$  é o custo da amostragem, calculado em função de  $p$ , definido por  $Ch=1/p$ ,  $n$  é o número total de segmentos que deve ser amostrado na área e  $i$  é o índice para representar o estrato de uso do solo com  $i = 1, \dots, m$ .

Para a estimativa da área ocupada com soja foi utilizado o modelo de expansão direta (Equação 3):

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

em que  $\hat{Z}$  é o estimador da área;  $i$  é o índice para representar o estrato de uso do solo com  $i = 1, \dots, m$ ;  $m$  é o número de estratos;  $k$  é o índice para representar o ponto amostrado com  $k =$

$1, \dots, n_i$ ;  $n_i$  é o número de segmentos amostrados no  $i$ -ésimo estrato de uso do solo;  $e_i$  é o fator de expansão ou inverso da probabilidade que um ponto tem de estar na amostra escolhida no  $i$ -ésimo estrato de uso do solo, sendo definido por  $e_i = (n_i/N_i)^{-1}$ , em que  $N_i$  é o número total de elementos no  $i$ -ésimo estrato de uso do solo; e  $z_{i,k}$  é a área de soja, no  $k$ -ésimo ponto do  $i$ -ésimo estrato de uso do solo.

Para aplicar a expansão direta na estimativa da área de soja, convencionou-se que cada ponto representa um pixel, com as dimensões de 900 m<sup>2</sup>, obtendo-se desta forma a relação entre cada ponto e a área representada por ele.

A variância foi estimada pela Equação 4:

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

em que  $s_i^2$  é a variância estimada da cultura no estrato  $i$ , definida pela Equação 5:

$$s_i^2 = \frac{\sum_{k=1}^{n_i} (z_{ik} - \bar{z}_i)^2}{(n_i - 1)} \quad (5)$$

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

$$\bar{z}_i = \frac{\sum_{k=1}^{n_i} z_{ik}}{n_i} \quad (6)$$

O coeficiente de variação (C.V.) é calculado pela Equação 7:

$$C.V.(%) = \frac{\sqrt{\hat{v}(\hat{Z})}}{\hat{Z}_c} * 100 \quad (7)$$

As estimativas resultantes desta amostragem foram comparadas com o dado de referência que consistiu de um mapa temático da soja referente à safra 2000/01 produzido por Rizzi (2004). Neste mapa as áreas de soja foram identificadas e mapeadas por meio da classificação digital e visual de imagens adquiridas pelos sensores Thematic Mapper (TM) e Enhanced TM plus (ETM+) a bordo dos satélites Landsat -5 e -7, respectivamente.

### 3. Resultados e Discussão

A **Tabela 1** apresenta o número de pontos por estrato e o total de pontos em cada amostra em função do percentual de confiança ( $\alpha/2$ ) e do erro de amostragem ( $E$ ) para estimar a área cultivada de soja. O número de pontos por amostra variou de 203 a 1.990. Apesar do estrato 1 possuir a maior área a quantidade de pontos sorteados para este estrato é menor em função do percentual de área ocupado com a cultura de soja (0 a 20%). Pode-se observar que a maior quantidade de pontos sorteadas encontra-se no estrato 3 que tem a maior concentração da área ocupada com soja (40-67%).

Tabela 1. Tamanhos amostrais utilizados em função do E e de  $\alpha$  e a subdivisão por estrato.

E		5,0%			2,5%			Área (km <sup>2</sup> )
$\alpha$		90%	95%	99%	90%	95%	99%	
		Número de pontos por amostra						
Estrato	1	44	63	109	177	251	434	60.917
	2	49	70	121	196	279	482	22.543
	3	110	155	269	438	622	1.074	28.169
Total		203	288	499	811	1.152	1.990	111.629

A **Figura 1** (entre a e f) apresenta o gráfico de dispersão com as diferenças relativas das estimativas de área plantada pelo método de amostragem em relação ao dado de referência (mapa temático) pelo seu C.V., para cada um dos seis conjuntos amostrais (203, 288, 499, 811, 1152 e 1990 amostras) nos 400 sorteios realizados, para o ano-safra 2000/01.

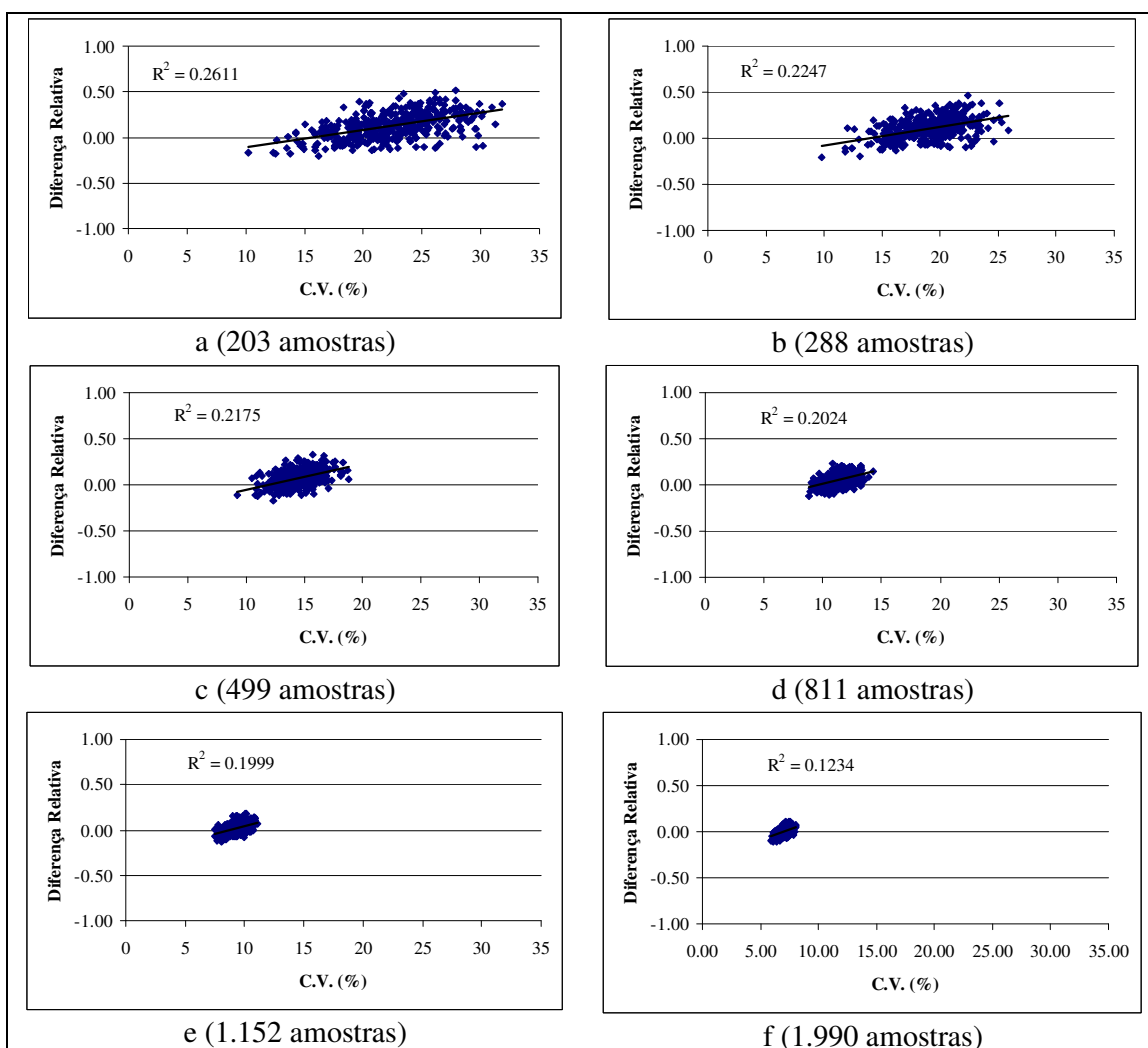


Figura 1. Gráfico de dispersão entre a diferença relativa (%) e o C.V.(%) obtidos para os conjuntos amostrais: a) 203 amostras; b) 288 amostras; c) 499 amostras; d) 811 amostras; e) 1.152 amostras e; f) 1.990 amostras.

Na **Figura 1** observa-se que na medida em que aumenta o tamanho amostral a variabilidade das diferenças relativas diminui e a diferença relativa tende para zero. Tal fato também ocorre com os C.V. cuja variabilidade e valores médios diminuem com o aumento do número de amostras.

Em média, os tamanhos amostrais menores ou iguais a 811 (**Figura 1a, b, c, d.**) tendem a superestimar a área cultivada, enquanto que os demais sorteios (**Figura 1e, f.**) apresentam estimativas em torno do valor esperado.

Observa-se também que os menores valores de C.V. estão associados às subestimativas enquanto que os maiores valores de C.V. estão associados às superestimativas. Essa correlação diminui em função do tamanho amostral (**Figura 1 e Tabela 2**). A **Tabela 2** apresenta os resultados dos C.V., as diferenças relativas e o coeficiente de correlação entre os C.V. e as diferenças relativas para cada conjunto amostral. Nesta tabela pode-se observar que na medida em que o tamanho amostral aumenta de 203 para 1.990 pontos a correlação entre o C.V. e a diferença relativa diminui de 0,51 para 0,35. Resultados semelhantes a estes foram encontrados por Rizzi et al. (2006), para sistema de amostragem por segmentos.

Tabela 2. Amplitude e média dos C.V. e da diferença relativa, e o coeficiente de correlação entre os C.V. e a diferença relativa, por sorteio.

Tamanho amostral	C.V.(%)			Diferença relativa			Coeficiente de correlação
	mínimo	máximo	média	mínimo	máximo	média	
203	10	32	22	-0,20	0,51	0,12	0,51
288	10	26	19	-0,20	0,46	0,11	0,47
499	9	19	15	-0,17	0,33	0,08	0,47
811	9	14	11	-0,13	0,23	0,05	0,45
1.152	7	11	9	-0,13	0,18	0,03	0,45
1.990	6	8	7	-0,11	0,11	0,00	0,35

Na **Tabela 2** pode-se verificar ainda que a partir do tamanho amostral 1.152 as estimativas tendem a aproximar-se dos valores médios e com C.V. que varia de 0,07 a 0,11 (para o tamanho amostral 1.152 pontos) e de 0,06 a 0,08 (para o tamanho amostral de 1.990 pontos). Desta forma, os tamanhos amostrais mais indicados para realizar a estimativa de área de soja no estado do Rio Grande do Sul, por pontos amostrais, estratificados, são os tamanhos amostrais que contem 1.152 ou 1.990 pontos. Esses tamanhos amostrais foram determinados com o uso de E igual a 2,5% e com  $\alpha$  igual a 95% ou 99%.

#### 4. Conclusões

O resultado desta pesquisa mostrou que a área de soja do estado do Rio Grande do Sul pode ser estimada por amostragem, sendo que os tamanhos amostrais com maior precisão foram os determinados para E igual a 2,5% e pelo  $\alpha$  igual a 95 ou 99%. Estes tamanhos amostrais diminuíram a correlação entre a diferença relativa e o C.V. e não apresentaram tendência a super ou subestimar a área cultivada e o C.V. ficou em média em torno de 7 a 9%. Com base nos resultados encontrados, o tamanho amostral recomendado é o de 1.152 pontos, pois tem boa precisão na estimativa e reduz o trabalho de campo para obter a estimativa.

#### Referências

Adami, M.; Moreira, M.A.; Rudorff, B.F.T.; Freitas, C.C.; Faria, R.T. Expansão direta na estimativa de culturas agrícolas por meio de segmentos regulares. **Revista Brasileira de Cartografia**, v.1, p. 22 – 27, 2005.

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., 2003, Belo Horizonte. **Anais...** São José dos Campos: INPE, 2003. p.27-31.

Fontana, D.C.; Ducati, J.R.; Weber, E.; Berlato, M. A.; Bergamaschi, H.; Figueiredo, D.C. **Monitoramento e Previsão de Safra de Soja no Brasil**. Porto Alegre : Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia. Universidade Federal do Rio Grande do Sul, 1999. 115p. il. (Série D. Relatório Técnico, 005/01).

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**. v.2, Roma, 1998. 242p.

Gallego, F.J. Remote sensing and land cover area estimation. **International Journal of Remote Sensing**, v.25, p.3019-47, 2004.

Pino, F.A. Estimativa subjetiva de safras agrícolas. **Informações Econômicas**, v.31, p.55-58, 2001.

Rizzi, R. **Geotecnologias em um sistema de estimativa da produção de soja: estudo de caso no Rio Grande do Sul**. 2004. 212p. Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

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., 2003, Belo Horizonte. **Anais...** São José dos Campos: INPE, 2003. p.231-238. CD-ROM.

Rizzi, R.; Rudorff, B.F.T; Adami, M. Estimativa da área de soja no Estado do Rio Grande do Sul por um método de amostragem, **Ciência Rural**, v.36, p. 30-35, 2006.

Thompson, S.K. **Sampling** 2 ed. New York: John Wiley & Sons, Inc., 2002. 343 p.