

# Estimativa amostral objetiva de área plantada regional, apoiada em imagens de sensoriamento remoto

Alfredo José Barreto Luiz<sup>(1)</sup>, Antônio Roberto Formaggio<sup>(2)</sup>, José Carlos Neves Epiphani<sup>(2)</sup>, John Mauricio Arenas-Toledo<sup>(3)</sup>, Elizabeth Goltz<sup>(2)</sup> e Daniela Brandão<sup>(2)</sup>

<sup>(1)</sup>Embrapa Meio Ambiente, Caixa Postal 69, CEP 13820-000 Jaguariúna, SP. E-mail: alfredo@cnpma.embrapa.br <sup>(2)</sup>Instituto Nacional de Pesquisas Espaciais, Avenida dos Astronautas, nº 1.758, Jardim Granja, CEP 12227-010 São José dos Campos, SP. E-mail: formag@dsr.inpe.br, epiphani@dsr.inpe.br, goltz@dsr.inpe.br, dany@dsr.inpe.br <sup>(3)</sup>Louis Dreyfus Commodities Brasil S.A., Avenida Brigadeiro Faria Lima, nº 1.355, 13º andar, CEP 01452-919 São Paulo, SP. E-mail: jonmauro@gmail.com

Resumo – O objetivo deste trabalho foi desenvolver e avaliar um método para estimar a área plantada de soja em escala regional e calcular o erro estatístico associado à estimação. O método (Geosafra), que associa técnicas de amostragem estatística com características das imagens obtidas por sensoriamento remoto orbital, foi aplicado para obter estimativa amostral objetiva da área cultivada com soja, na safra de 2005/2006, no Estado do Rio Grande do Sul. Os municípios produtores de soja, no RS, foram distribuídos em dez estratos, com base em dados pré-existentes de área cultivada com a cultura. O número de municípios selecionados, em cada estrato, seguiu a regra de alocação de Neyman. Em cada município selecionado, foram aleatorizados pontos correspondentes aos pixels das imagens, classificados como “soja” ou “não soja” após visita a campo. A partir dos dados de 3.000 pontos distribuídos nos 30 municípios selecionados, nos dez estratos, foi estimada a área cultivada com soja no RS, que totalizou 4.069.887 ha, com coeficiente de variação (CV) de 3,4%. Esta estimativa foi consistente com os dados oficiais. O método amostral objetivo estratificado, auxiliado por sensoriamento remoto, permite estimar a área cultivada com soja no Rio Grande do Sul e é capaz de quantificar o erro associado à estimativa realizada.

Termos para indexação: *Glycine max*, erro amostral, estatística agrícola, estratificação, imagem de satélite, previsão de safras.

## Objective sampling estimation of regional crop area supported by remotely sensed images

Abstract – The objective of this work was to develop and evaluate a method for estimating soybean crop area on a regional scale and to calculate the statistical error associated with the estimate. The method (Geosafra), which combines statistical sampling techniques with characteristics of images obtained by orbital remote sensing, was applied to obtain an objective sampling estimation for soybean crop area in the 2005/2006 harvest season in the state of Rio Grande do Sul (RS), Brazil. Soybean-producing municipalities in RS were distributed into ten strata according to preexisting data on the area cultivated with the crop. The number of municipalities selected in each stratum followed Neyman's allocation rule. In each selected municipality, points corresponding to the pixels of images were randomized and classified as “soybean” or “non-soybean” after site visitation. From the data of 3,000 points in the 30 selected municipalities across the ten strata, soybean crop area in RS was estimated, totaling 4,069,887 ha, with a coefficient of variation (CV) of 3.4%. This estimate was consistent with official data. The stratified objective sampling method, supported by remote sensing, allows for the estimation of the area cultivated with soybean in the state of Rio Grande do Sul and is able to quantify the error associated with the calculated estimate.

Index terms: *Glycine max*, sampling error, agricultural statistics, stratification, satellite image, crop forecasting.

### Introdução

As imagens obtidas por sensoriamento remoto são formas de captura indireta de informação espacial da superfície terrestre. Essas imagens são armazenadas como matrizes, nas quais cada elemento, denominado

pixel, é representado por um valor relacionado às características do alvo da observação, localizado na superfície imageada. Dessa forma, nas imagens digitais, cada porção do espaço, imageada pelos sensores remotos, corresponde a uma célula matricial

ou pixel, que deve estar geograficamente identificada (Jensen, 2000).

Métodos de levantamento agrícola, que considerem o crescente aumento da disponibilidade de imagens de sensoriamento remoto e outros avanços tecnológicos, podem vir a ser as formas mais práticas de um país produzir dados sobre as principais mercadorias agrícolas (Food and Agriculture Organization of the United Nations, 1996; Gallego, 2004). Para Gao (2002), a tecnologia sistema de posicionamento global (GPS) facilitou a aquisição de dados espaciais e diversificou enfoques, ao permitir a integração com o sensoriamento remoto e com os sistemas de informações geográficas (SIGs).

Enquanto os SIGs estão presentes em diversas aplicações agrícolas, o sensoriamento remoto, apesar do seu potencial, estabeleceu-se nos levantamentos agrícolas apenas na construção de painéis amostrais por área (Adami et al., 2007). Somente nos EUA, foi incorporado de forma operacional ao programa de estatísticas agrícolas (Food and Agriculture Organization of the United Nations, 1998).

No Brasil, o Instituto Brasileiro de Geografia e Estatística (IBGE) realiza, anualmente, levantamento sistemático da produção agrícola por município, com dados obtidos de modo subjetivo, baseado em consultas a especialistas; e, em periodicidade variável, censo agropecuário, com informações coletadas por entrevistas em cada estabelecimento rural. No entanto, foram observadas deficiências relacionadas à subjetividade e à magnitude dos erros nestes levantamentos (Batista et al., 1978; Pino, 1999). Artigos recentes ressaltam a importância de se apresentar os erros associados às estimativas de área plantada (Wu & Li, 2012).

Portanto, uma das demandas atuais é promover a incorporação de avanços tecnológicos modernos aos procedimentos utilizados nos levantamentos das estatísticas agrícolas, para aumentar a objetividade, a rapidez e a confiabilidade das estimativas de safras de alguns produtos agrícolas cultivados no Brasil. Assim, para associar a tecnologia do sensoriamento remoto orbital aos procedimentos utilizados nos levantamentos das estatísticas agrícolas no Brasil, foi desenvolvido um método de amostragem simples para estimativa de áreas de culturas agrícolas de maior expressão (Luiz et al., 2011). Esse método também permite a obtenção do coeficiente de variação associado à estimativa

realizada, o que só é possível pelo uso da amostragem probabilística.

O objetivo deste trabalho foi avaliar um método (Geosafra) para estimar a área cultivada de soja em escala regional e calcular o erro estatístico associado à estimação.

## Material e Métodos

A área de estudo compreendeu o Estado do Rio Grande do Sul, pioneiro no cultivo da soja [*Glycine max* (L.) Merrill] em grande escala no Brasil. Os principais passos metodológicos para a estimativa da área agrícola de soja, na safra de 2006, foram: estratificação estadual; composição do painel amostral de municípios intraestadual; definição do painel amostral intramunicípios; definição, processamento e geração do material de sensoriamento remoto auxiliar aos trabalhos de campo; trabalho de campo, em que os técnicos visitaram todos os pontos amostrais, em cada município de cada estrato; e cálculo das estimativas de área agrícola em escala estadual para soja.

Como a área cultivada com soja, em cada um dos 497 municípios do RS, constituiu população heterogênea, a estratificação foi recomendada para obtenção de subpopulações internamente homogêneas. Se cada estrato fosse homogêneo, de modo que as medidas variassem pouco entre seus componentes, uma estimativa precisa da média do estrato poderia ser obtida de uma pequena amostra. Posteriormente, as estimativas dos estratos foram combinadas em uma única estimativa precisa de toda a população (Cochran, 1977).

Foi feita a divisão da população (total de municípios que cultivam soja no estado) em  $L$  estratos, cada um com  $N_h$  unidades, em que o sufixo  $h$  definiu o  $h$ -ésimo estrato da população. Os estratos não se sobrepuseram e, juntos, compreenderam toda a população, conforme a equação:

$$N = N_1 + N_2 + \dots + N_h + \dots + N_L = \sum_{h=1}^L N_h.$$

Para a caracterização da amostragem aleatória estratificada (AAE), foi necessária a obtenção de uma amostra aleatória simples (AAS), de forma independente, em cada um dos estratos. Os tamanhos das amostras dentro dos estratos foram chamados de  $n_1, n_2, \dots, n_h, \dots, n_L$ . O tamanho total da amostra ( $n$ )

foi dado pela soma dos  $L$  valores de  $n_h$ . Para seguir a notação adotada no presente trabalho, tendo-se considerado que a classe de interesse ( $U$ ) era formada pelos pixels classificados como “soja” e a classe complementar ( $\bar{U}$ ) pelos “não soja”, assumiu-se que:  $A_h$  é o número de elementos da classe  $U$  no estrato  $h$ ;  $a_h$  é o número de elementos da classe  $U$  na amostra do estrato  $h$ ;  $P_h = A_h/N_h$  é a proporção de elementos da classe  $U$  no estrato  $h$ ;  $p_h = a_h/n_h$  é a proporção de elementos da classe  $U$  na amostra do estrato  $h$ ;  $Q_h = 1 - P_h$  é a proporção de elementos da classe  $\bar{U}$  no estrato  $h$ ;  $q_h = 1 - p_h$  é a proporção de elementos da classe  $\bar{U}$  na amostra do estrato  $h$ ;  $f_h = n_h/N_h$  é a fração amostral no estrato e  $1/f_h = N_h/n_h$  é o fator de expansão; e  $W_h = N_h/N$  é o peso do estrato  $h$ .

Para determinar o número de pontos amostrais  $n_h$  (ou o tamanho da amostra) em cada estrato, supondo função de custo linear e dada a restrição em termos de número total de pontos amostrais de 3.000, foi feita a melhor alocação de pontos para obter a variância total mínima.

Segundo Cochran (1977), quando a AAE é aplicada a proporções, os estratos devem ser construídos de tal forma que a proporção na classe  $U$  varie tanto quanto possível de um estrato para outro. Para essa situação, a estimativa da proporção  $P$ , apropriada para a AAE, foi chamada de  $\hat{P}_{est}$  e dada pela equação:

$$\hat{P}_{est} = \sum_{h=1}^L \frac{N_h \times P_h}{N} = \sum_{h=1}^L (W_h \times p_h).$$

Já a estimativa da variância de  $P_{est}$ , foi calculada com o auxílio da equação:

$$\hat{S}_{P_{est}}^2 = \sum_{h=1}^L W_h^2 \frac{p_h \times q_h}{n_h}.$$

Na alocação ótima de Neyman, a intensidade de amostragem calculada foi distribuída proporcionalmente à variância da área em cada estrato, conforme apresentado a seguir:

$$n_h = \frac{W_h S_h}{\sum_{h=1}^L (W_h S_h)} \times n.$$

A intensidade de amostragem foi obtida da mesma maneira que na amostragem aleatória simples; no entanto, neste caso, a estimativa da variância

corresponde à variância ponderada dos estratos para a população finita:

$$n' = \frac{t^2 \times \left( \sum_{h=1}^L W_h \times S_h \right)^2}{E_r^2 + t^2 \times \sum_{h=1}^L \frac{W_h \times S_h^2}{N}},$$

em que:  $n'$  é a intensidade amostral ideal;  $t$  é o valor tabelado de  $t$  para um nível de significância  $\alpha$  definido pelo usuário; e  $E_r$  é o erro amostral relativo.

Adotou-se o conceito do pixel da imagem como unidade amostral básica, o que permitiu e facilitou a realização de estimação objetiva (Luiz et al., 2002). A possibilidade de estabelecer a precisão das estimativas é garantida pela natureza das variáveis medidas e pelo estabelecimento do plano amostral adequado (Epiphany et al., 2002).

Para extrair uma AAS, de  $n$  pixels, de uma imagem digital, enumeraram-se, de 1 até  $N$ , os elementos de cena que compuseram a imagem e selecionaram-se os  $n$  elementos da amostra. Como a unidade amostral é o pixel, que tem área conhecida e constante em toda a região investigada, a proporção de ocorrência da cultura de interesse, na amostra de pixels, permitiu, por expansão, estimar a proporção da cultura no município. O método para realizar essa estimativa foi descrito em Luiz et al. (2011).

Para calcular o tamanho de amostra adequado à estimativa de área ocupada por determinada cultura agrícola, em uma região delimitada, é necessário supor, a priori, a proporção  $p$  da área total ocupada pela cultura. Ou seja, dados os municípios do RS, a partir de sua área e das dimensões do elemento de cena, tem-se o valor de  $N$ , e, com base em informações obtidas de levantamentos agrícolas anteriores (Instituto Brasileiro de Geografia e Estatística, 2012), obtém-se um valor  $p$  para cada um. Para o cálculo do coeficiente de variação ( $CV$ ) e do percentual da estimativa do número total de pixels com a cultura ( $Np$ ), utilizou-se a equação:

$$CV_{Np} = \frac{100}{p} \times \sqrt{\frac{pq}{n-1}},$$

em que  $q = 1 - p$ .

Dessa forma, foi possível, para uma determinada proporção, a priori, do território do município, ocupada

pela soja, estimar o tamanho da amostra ( $n$ ) necessário para a obtenção do CV desejado.

Para garantir a aleatoriedade, assegurou-se que todos os componentes da população tivessem chance não nula e conhecida de pertencer à amostra. Para isso, é necessário conhecer todos os elementos de uma população finita ou a distribuição de probabilidade no caso de populações infinitas.

A estimativa de área plantada foi caracterizada como pertencente ao grupo das populações finitas. Por exemplo, se a unidade amostral básica escolhida for a propriedade rural, para poder realizar amostragem aleatória, será necessário uma lista completa de todas as propriedades existentes na população considerada, o que geralmente não está disponível ou não existe. Para contornar este obstáculo, uma opção é o uso dos chamados “painéis amostrais por área”. Embora, em tese, eles possam contemplar elementos com limites definidos artificialmente, essa não era a regra até recentemente e os segmentos normalmente eram definidos pelos limites físicos do terreno.

O uso do pixel como unidade amostral básica permitiu, em levantamentos agrícolas, a utilização de imagens de sensoriamento remoto, que fornecem, naturalmente, elementos com limites artificiais – os pixels, na obtenção de amostras aleatórias para estimativa da área plantada em levantamentos agrícolas. Isso facilita o planejamento amostral e a análise estatística dos dados coletados. Entretanto, é preciso garantir que o procedimento prático de obtenção da amostra realmente resulte em processo aleatório, condição fundamental para a aplicação de todos os demais conceitos estatísticos usados nos cálculos das estimativas.

Ao se considerar que cada cena de imagem digital, obtida por sensoriamento remoto orbital, é representada por uma matriz  $n \times m$ , em que  $n$  é o número de colunas e  $m$  é o número de linhas, e que cada célula dessa matriz compõe um elemento de cena (pixel), que pode ter sua posição no espaço representada por coordenadas planas  $X$  e  $Y$ , pode-se afirmar que: se, por meio de um processo de geração de números aleatórios, forem produzidos  $k$  pares ordenados  $(x_i, y_j)$ , que determinem a posição de  $k$  pixels na imagem de interesse, esses pixels constituirão amostra aleatória (AA) representativa da imagem; e se, ao se sobrepor um polígono irregular qualquer (por exemplo, os limites de um município) sobre a imagem, forem conservados apenas os  $k'$  pixels ( $k' \leq k$ ), cujos

centros estejam localizados no interior do referido polígono, esse novo conjunto de pixels continuará a compor uma AA representativa da região da imagem correspondente à área do município.

A partir da exportação das tabelas de dados amostrais para uma planilha eletrônica, os dados foram trabalhados para gerar as estimativas. O procedimento consistiu em contagem direta de pontos da classe de interesse, ou seja, a soja.

O cálculo da área foi feito pela divisão da frequência observada para a classe de interesse pelo tamanho da amostra, seguida da multiplicação pela área total do município. A razão entre a frequência observada e o tamanho da amostra nada mais foi do que a proporção  $p$  de ocorrência da classe na amostra, que é um estimador não tendencioso da proporção  $P$  da classe na população. Assim, multiplicou-se o  $p$  da classe soja pela área total do município ( $AM$ ), para se obter a área estimada para aquela classe:  $\hat{A} = p \times AM$ .

Em todo o Estado do Rio Grande do Sul, foram considerados 279 municípios representativos no cultivo de soja, em 2004, com, pelo menos, 1 ha plantado (Instituto Brasileiro de Geografia e Estatística, 2012). Esses municípios foram estratificados com uso do programa SAS, com a aplicação de um algoritmo de agrupamento denominado Fastclus.

Em seguida, foram analisados os municípios que ficariam sozinhos nos estratos por apresentarem áreas significativamente discrepantes das dos demais. Estes municípios foram considerados pertencentes ao estrato certo e, no caso da soja, no RS, totalizaram três municípios, cujas áreas plantadas, em 2004, foram superiores a 70.000 ha, de acordo com dados do Sistema IBGE de Recuperação Automática – Sidra (Instituto Brasileiro de Geografia e Estatística, 2012). Os municípios que compuseram o estrato certo foram: Tupanciretã (116.780 ha com soja), Palmeira das Missões (92.000 ha) e Cruz Alta (82.000 ha). Portanto, foram estratificados no SAS os 276 municípios restantes. Entretanto, estabeleceu-se que o número máximo de municípios a serem incluídos na amostra seria 30. Como três deles foram definidos como estrato certo, restavam 27 a serem escolhidos para distribuição dos pontos amostrais.

Para determinar o número de amostras por estrato, foram realizadas duas análises: a alocação ótima de Neyman (Neyman, 1934) e a distribuição proporcional (Cochran, 1977). A alocação de Neyman, de modo



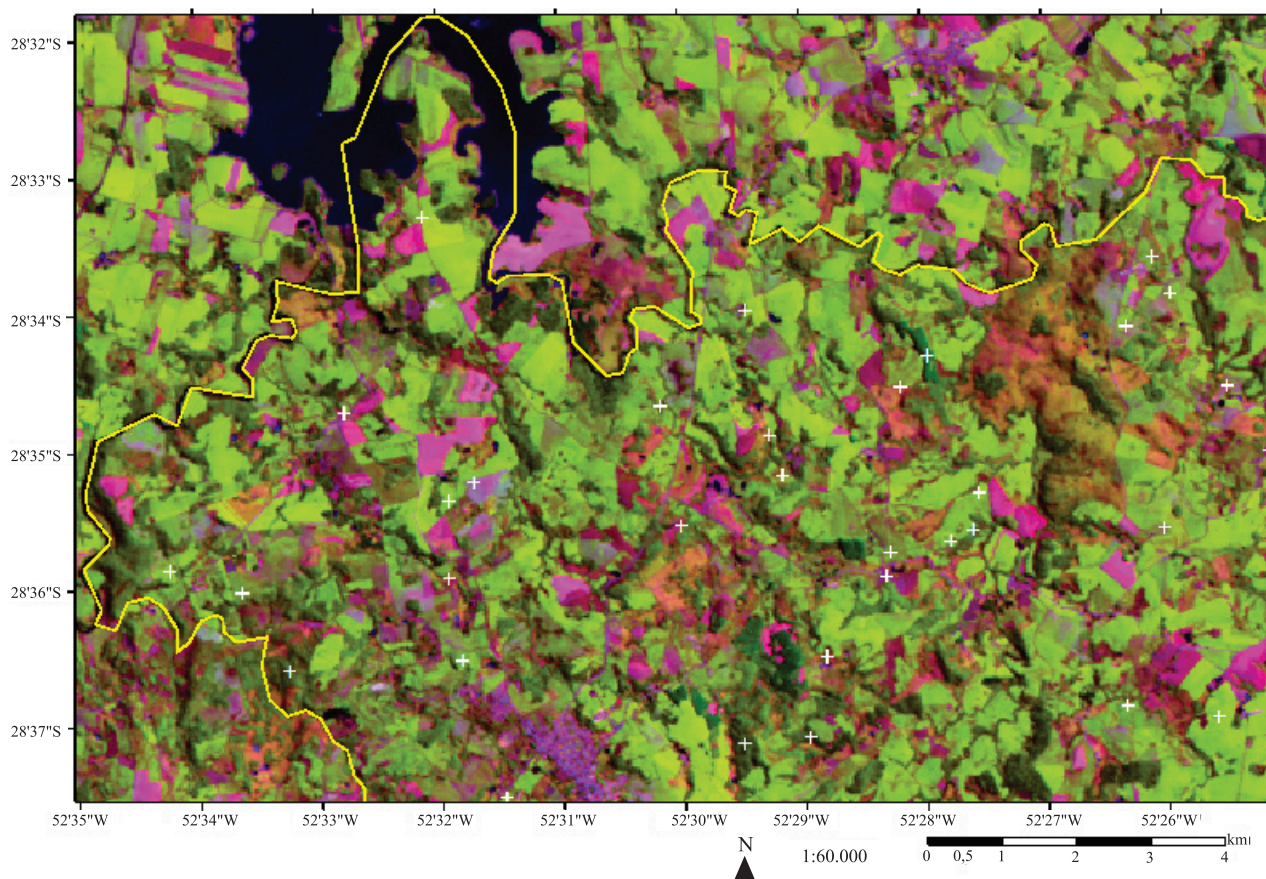
geral, apresentou melhor distribuição das amostras dentro dos estratos do que a distribuição proporcional. Seguindo os cálculos da alocação de Neyman, para os casos de estratificação da soja no Estado do Rio Grande do Sul, que variaram de cinco a oito estratos, os resultados produziram alguns grupos nos quais as amostras seriam constituídas por município ou fração próxima a zero (nos grupos de seis, sete e oito estratos), o que afetaria as estimativas destes grupos. Portanto, fixou-se um tamanho amostral mínimo de três municípios por grupo e executou-se, novamente, a alocação de Neyman para os demais grupos.

A escolha dos municípios amostrados, em cada estrato, foi realizada de maneira aleatória, conforme recomendado por Cochran (1977). A distribuição de pontos, nos municípios, foi feita tendo-se levado em conta as áreas municipais fornecidas pelo Sidra (Instituto Brasileiro de Geografia e Estatística, 2012). O número de pontos, em cada município, foi determinado,

proporcionalmente, em função do parâmetro  $p$ , fixado o número de pontos no estrato e redistribuídos os pontos pela área territorial do estrato, a partir de dados do Sidra, o que totalizou 3.000 pontos.

Todos os dados (imagens de satélite registradas, painel amostral municipal intraestadual, limites municipais e estadual, lista de pontos amostrais intramunicipais com coordenadas identificadas) foram organizados em banco de dados georeferenciados, desenvolvido no programa Spring (Câmara et al., 1996).

As pranchas ou imagetes de campo, em escala 1:60.000, e as planilhas de campo foram confeccionadas e enviadas para técnicos de campo, que tinham a função de visitar cada ponto e identificar quais culturas estavam presentes na ocasião da visita (Figura 1). O trabalho de campo foi realizado pelos técnicos de campo entre 31 de março e 14 de junho de 2006. Cada técnico de campo estava equipado com aparelho GPS, com as fichas de campo e com as pranchas de campo



**Figura 1.** Exemplo de imagete originado de uma imagem TM/Landsat-5 de 11/3/2006 de parte de um município (Ibirapuitã, RS), com marcas (+) dos pontos amostrais visitados.

feitas com imagens de satélite. As datas de referência adotadas para cada município foram sempre fevereiro de 2006, ou seja, mesmo que o trabalho de campo tenha sido realizado em março, abril, maio ou junho, o técnico de campo deveria especificar qual tinha sido, em fevereiro, a cultura presente em cada ponto visitado.

Após a realização dos levantamentos de campo, em que cada ponto ou unidade amostral teve seu uso identificado, os dados brutos foram colocados na forma de planilha eletrônica. Os municípios Capão do Leão, Encruzilhada do Sul, Rio Pardo, São Borja e São Gabriel tiveram alguns pontos pré-interpretados para reduzir os pontos a serem visitados em campo e, conseqüentemente, os custos. De posse da proporção de pontos classificados como soja em cada município, foi calculada a área da cultura pela multiplicação pela área total do município. De acordo com os valores calculados, cada município foi realocado no estrato correspondente, segundo intervalos pré-determinados, e foram realizadas as estimativas de área por estrato e para o Estado do Rio Grande do Sul, com os respectivos coeficientes de variação.

## Resultados e Discussão

Entre os valores testados, a estratificação em sete grupos foi escolhida porque foi aquela que otimizou a redução da variância em relação ao número de estratos. Isso foi observado ao se comparar a soma, em cada estratificação, do desvio-padrão multiplicado pelo número de municípios (Tabela 1). Ao se aumentar o número de estratos de cinco para seis, de seis para sete e de sete para oito, esse total foi reduzido em 7,4% (de 749.598 para 694.426), 20,3% (de 694.426 para 553.248) e 3,2% (de 553.248 para 535.504),

respectivamente. Com base nos resultados anteriores para a soja no Rio Grande do Sul, o tamanho da amostra de municípios, para a estratificação escolhida, ficou como apresentado na Tabela 2, onde constam, ainda, os limites de área cultivada com soja que definem cada estrato. A Tabela 3 mostra os municípios escolhidos em cada estrato, inclusive o estrato certo, com as áreas municipais fornecidas pelo Sidra (Instituto Brasileiro de Geografia e Estatística, 2012). O número de pontos classificados em cada município pode ser visto na Tabela 4, na qual também podem ser observadas as datas de realização do trabalho de campo.

A área de cultura da soja foi estimada por amostragem estratificada, na safra de 2006, para os estratos considerados e para o Estado do Rio Grande do Sul, tendo totalizado 4.069.887 ha (Tabela 5). Os coeficientes de variação, para os estratos, variaram entre 5 e 15%, e, para o estado, o CV ficou em 3,4%.

Ao se comparar estas estimativas com as do IBGE, verificou-se que, pelo Censo Agropecuário 2006, a área colhida com soja, no RS, foi de 3.390.591 ha, o que difere, de forma significativa, do resultado obtido pela amostragem; entretanto, o IBGE estimou em 3.863.726 ha a área colhida de soja naquele estado, em 2006, pelo levantamento Produção Agrícola Municipal (PAM). O valor estimado pela PAM do IBGE está dentro do intervalo de confiança de 95% da média populacional estimada por amostragem probabilística, que variou de 3.798.670 a 4.341.104 ha. Os dados do censo exigem esforço substancial para sua obtenção, pois é necessário visitar cada unidade de produção agropecuária do país em um curto período de tempo e coletar dezenas de informações socioeconômicas, que diferem, em muito, das informações coletadas quando da realização dos levantamentos da PAM. Os informantes também são diferentes, ou seja, produtores, no caso do censo,

**Tabela 1.** Resultados das análises de Neyman para sete estratos.

Estrato	Média	Variância	Desvio (A)	CV (%)	Nº de municípios (B)	A x B	Amostra de Neyman	
							Inicial	Ajustada
1	3.887	2.535.275	1.592	41	115	183.109	8,9 $\cong$ 9	→ 7
2	9.592	3.576.838	1.891	20	63	119.149	5,8 $\cong$ 6	→ 5
3	15.887	3.792.737	1.948	12	44	85.690	4,2 $\cong$ 4	→ 3
4	26.088	9.000.890	3.000	12	26	78.004	3,8 $\cong$ 4	→ 3
5	36.431	9.416.217	3.069	8	18	55.234	2,7 $\cong$ 3	→ 3
6	48.000	2.000.000	1.414	3	4	5.657	0,3 $\cong$ 0	→ 3
7	62.833	19.366.667	4.401	7	6	26.405	1,3 $\cong$ 1	→ 3
Total					276	553.248	27	27

e atores importantes da cadeia produtiva, no caso da PAM. Embora essas características não sirvam para definir se os dados de área cultivada com soja, provenientes do censo, são melhores ou piores que os

obtidos pela PAM, explicam, em parte, as diferenças encontradas entre as estimativas.

Assim, um dos pontos mais relevantes a ser considerado na estimativa objetiva amostral, utilizada

**Tabela 2.** Limites inferiores e superiores dos sete estratos e dos estratos certos, em valores da área cultivada com soja, segundo dados de 2004 (IBGE, 2012).

Limite inferior (ha)	Limite superior (ha)	Nº de municípios	Área total dos municípios (ha)	Estrato
1.400	6.699	115	4.929.657	1
6.700	12.824	63	3.835.387	2
12.825	20.949	44	2.303.335	3
20.950	31.499	26	4.030.138	4
31.500	44.499	18	1.622.619	5
44.500	53.499	4	1.200.410	6
53.500	74.999	6	1.008.600	7
75.000	86.999	1	243.260	certo – 8
87.000	104.389	1	154.740	certo – 9
104.390	∞	1	225.010	certo – 10

**Tabela 3.** Painel amostral da soja no Rio Grande do Sul.

Municípios	Área de soja em 2004 (ha)	Área do município (ha)	p <sup>(1)</sup>	Estrato
Planalto	1.800	23.042	0,08	1
Santo Antônio do Palma	2.200	12.610	0,17	1
Novo Cabrais	2.400	19.229	0,12	1
São Valentim	2.400	15.419	0,16	1
Capão do Leão	4.800	78.538	0,06	1
Encruzilhada do Sul	6.000	334.833	0,02	1
Santa Cecília do Sul	6.600	19.940	0,33	1
São Valério do Sul	6.800	10.797	0,63	2
Ciríaco	7.200	27.387	0,26	2
Alegria	10.500	17.269	0,61	2
Coronel Barros	11.200	16.295	0,69	2
Ibirapuitã	11.430	30.703	0,37	2
Redentora	13.000	30.268	0,43	3
Almirante Tamandaré do Sul	16.400	26.537	0,62	3
Victor Graeff	16.700	23.827	0,70	3
Independência	22.500	35.744	0,63	4
Rio Pardo	23.000	205.060	0,11	4
Condor	29.740	46.519	0,64	4
Passo Fundo	37.400	78.342	0,48	5
Espumoso	40.000	78.307	0,51	5
Santo Antônio das Missões	42.000	171.088	0,25	5
Cachoeira do Sul	47.000	373.518	0,13	6
São Miguel das Missões	48.000	122.985	0,39	6
São Gabriel	50.000	502.384	0,10	6
São Borja	60.000	361.604	0,17	7
Jóia	66.000	123.589	0,53	7
Júlio de Castilhos	68.000	192.939	0,35	7
Cruz Alta	82.000	136.038	0,60	certo - 8
Palmeira das Missões	92.000	141.944	0,65	certo - 9
Tupanciretã	116.780	225.187	0,52	certo - 10

<sup>(1)</sup>Proporção estimada da área do município ocupada pela cultura da soja, com base nos dados mais recentes disponíveis à época da realização do trabalho.

no presente trabalho, é a sua capacidade de gerar um erro associado à estimativa. Dessa forma, a metodologia proposta traz a possibilidade de gerar números estimativos das safras agrícolas em nível de estado, de forma rápida, barata e com precisão estatisticamente garantida, o que pode ser utilizado para avaliar os

**Tabela 4.** Número total de pontos aleatorizados, número de pontos com soja, valor da proporção p e área estimada da cultura em cada município amostrado, no Estado do Rio Grande do Sul, em 2006.

Municípios	Nº total de pontos	Data do trabalho de campo (2006)	Nº de pontos com soja	p	Área estimada com soja (ha) e novo estrato
São Valentim	31	3 a 30/4	3	0,10	1.492-1
Planalto	48	3 a 27/4	4	0,08	1.920-1
Capão do Leão	160	30/3 a 28/4	8	0,05	3.927-1
Santo Antônio do Palma	26	1 a 20/4	10	0,38	4.850-1
São Valério do Sul	37	3/4 a 12/5	17	0,46	4.961-1
Encruzilhada do Sul	698	3/4 a 2/6	15	0,02	7.196-2
Novo Cabrais	40	3 a 24/4	15	0,38	7.211-2
Redentora	64	31/3 a 14/6	16	0,25	7.567-2
Alegria	55	3 a 30/4	25	0,45	7.849-2
Santa Cecília do Sul	40	31/3 a 19/4	17	0,43	8.474-2
Ciríaco	88	5 a 29/4	29	0,33	9.025-2
Coronel Barros	51	4 a 27/4	30	0,59	9.585-2
Victor Graeff	55	1 a 29/4	28	0,51	12.130-2
Ibirapuitã	118	7/4 a 16/5	58	0,49	15.091-3
Almirante Tamandaré do Sul	55	3 a 28/4	32	0,58	15.440-3
Independência	25	10 a 13/4	12	0,48	17.157-3
Rio Pardo	158	28/4 a 12/5	18	0,11	23.361-4
Condor	34	24 a 27/4	21	0,62	28.732-4
Santo Antônio das Missões	138	2/4 a 13/6	24	0,17	29.754-4
Passo Fundo	62	27/4 a 16/5	32	0,52	40.435-5
Espumoso	73	3/4 a 12/5	40	0,55	42.908-5
São Miguel das Missões	51	25/4 a 17/5	23	0,45	55.464-7
Cachoeira do Sul	138	3/4 a 11/5	23	0,17	62.253-7
Jóia	54	31/3 a 14/6	29	0,54	66.372-7
São Borja	157	24/4 a 8/5	31	0,20	71.399-7
São Gabriel	223	3/5 a 14/6	38	0,17	85.608-8
Palmeira das Missões	60	3 a 27/4	38	0,63	89.898-9
Júlio de Castilhos	80	3/4 a 15/5	43	0,54	103.705-9
Cruz Alta	94	10/4 a 26/5	72	0,77	104.199-9
Tupanciretã	87	5/4 a 10/5	51	0,59	132.006-10

**Tabela 5.** Área estimada por amostragem estratificada, para a cultura da soja, na safra de 2005/2006, no Estado do Rio Grande do Sul, com respectivo coeficiente de variação da estimativa.

Estratos (h)	N <sub>h</sub>	n <sub>h</sub>	Área média de soja nos municípios da amostra	Área estimada de soja no estrato	CV (%)
1	113	5	3.430	387.590	14
2	66	8	8.630	560.926	6
3	44	3	15.896	699.424	9
4	26	3	27.282	709.341	11
5	18	2	41.672	646.960	8
6	0	0	-	-	-
7	7	4	63.872	447.104	8
8	1	1	85.608	85.608	15
9	3	3	99.267	297.802	5
10	1	1	132.006	132.006	9
Total para o estado				4.069.887	3,4



métodos convencionais de estatísticas agrícolas em uso no país.

Outros trabalhos que utilizam imagens para estimar a área cultivada, mas optam por fazer o mapeamento do território estudado e não utilizam a amostragem probabilística, geralmente apresentam medidas de acurácia da classificação realizada (Howard et al., 2012; Pan et al., 2012), o que não é equivalente à estimativa do erro amostral representado pelo CV calculado.

### Conclusão

O método amostral estratificado, auxiliado por sensoriamento remoto, permite estimar a área cultivada com soja no Rio Grande do Sul e é capaz de quantificar o erro associado à estimativa realizada.

### Agradecimentos

À Companhia Nacional de Abastecimento, pelo apoio financeiro.

### Referências

- ADAMI, M.; MOREIRA, M.A.; RUDORFF, B.F.T.; FREITAS, C. da C.; FARIA, R.T. de; DEPPE, F. Painel amostral para estimativa de áreas agrícolas. **Pesquisa Agropecuária Brasileira**, v.42, n.1, p.81-88, 2007.
- BATISTA, G.T.; MENDONÇA, F.J.; LEE, D.C.L.; TARDIN, A.T.; CHEN, S.C.; NOVAES, R.A. **Uso de sensores remotos a bordo de satélite e aeronave na identificação e avaliação de áreas de culturas para fins de previsão de safras**. São José dos Campos: INPE, 1978. 41p.
- CÂMARA, G.; SOUZA, R.C.M.; FREITAS, U.M.; GARRIDO, J. SPRING: integrating remote sensing and GIS by object-oriented data modelling. **Computers and Graphics**, v.20, p.395-403, 1996.
- COCHRAN, W.G. **Sampling techniques**. 3<sup>rd</sup> ed. New York: John Wiley, 1977. 428p.
- EPIPHANIO, J.C.N.; LUIZ, A.J.B.; FORMAGGIO, A.R. Estimativa de áreas agrícolas municipais, utilizando sistema de amostragem simples sobre imagens de satélite. **Bragantia**, v.61, p.187-197, 2002.
- FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **Multiple frame agricultural surveys: agricultural survey programmes based on area frame or dual frame (area and list) sample designs**. Rome: FAO, 1998. v.2, 242p. (FAO. Statistical development series, 10).
- FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **Multiple frame agricultural surveys: current surveys based on area and list sampling methods**. Rome: FAO, 1996. v.1, 119p. (FAO. Statistical development series, 7).
- GALLEGO, F.J. Remote sensing and land cover area estimation. **International Journal of Remote Sensing**, v.25, p.3019-3047, 2004.
- GAO, J. Integration of GPS with remote sensing and GIS: reality and prospect. **Photogrammetric Engineering and Remote Sensing**, v.68, p.447-453, 2002.
- HOWARD, D.M.; WYLIE, B.K.; TIESZEN, L.L. Crop classification modelling using remote sensing and environmental data in the Greater Platte River Basin, USA. **International Journal of Remote Sensing**, v.33, p.6094-6108, 2012.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Sistema IBGE de Recuperação Automática – SIDRA**. Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em: 31 ago. 2012.
- JENSEN, J.R. **Remote sensing of the environment: an earth resource perspective**. Upper Saddle River: Prentice Hall, 2000. 544p.
- LUIZ, A.J.B.; FORMAGGIO, A.R.; EPIPHANIO, J.C.N. Objective sampling estimation of crop area based on remote sensing images. In: PRADO, H.A. do; LUIZ, A.J.B.; CHAIB FILHO, H. (Ed.). **Computational methods for agricultural research: advances and applications**. Hershey: Information Science Reference, 2011. p.73-95.
- LUIZ, A.J.B.; OLIVEIRA, J.C.; EPIPHANIO, J.C.N.; FORMAGGIO, A.R. Auxílio das imagens de satélite aos levantamentos por amostragem em agricultura. **Agricultura em São Paulo**, v.49, p.41-54, 2002.
- NEYMAN, J. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. **Journal of the Royal Statistical Society**, v.97, p.558-625, 1934.
- PAN, Y.Z.; LI, L.; ZHANG, J.S.; LIANG, S.L.; ZHU, X.F.; SULLA-MENAS, D. Winter wheat area estimation from MODIS-EVI time series data using the Crop Proportion Phenology Index. **Remote Sensing of Environment**, v.119, p.232-242, 2012.
- PINO, F.A. Estatísticas agrícolas para o século XXI. **Agricultura em São Paulo**, v.46, p.71-105, 1999.
- WU, B.F.; LI, Q.Z. Crop planting and type proportion method for crop acreage estimation of complex agricultural landscapes. **International Journal of Applied Earth Observation and Geoinformation**, v.16, p.101-112, 2012.