

Aprimoramento do método de amostragem simples utilizado pelo Projeto Geosafras para estimativa municipal de área plantada com soja

Pablo Santana Santos¹
José Carlos Neves Epiphanyo¹

¹ Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
{santos, epiphanyo}@dsr.inpe.br

Abstract. This paper presents a methodology that improves the estimates of soybean at municipality level. It is based on the analysis of remote sensing products – TM/Landsat-5 and 7 images in this case – and the elimination of non-crop areas. Over this resulting products a sampling design is applied, in which the pixel is the key element to be visited and identified in the field or interpreted based on remote sensing analysis. In addition, it is analyzed the temporal evolution of soybean for the last five years. In comparison to the official statistics, the results that came up from this methodology were very valuable, based on statistical t-test.

Palavras-chave: remote sensing, crop area, sampling, sensoriamento remoto, área plantada, amostragem.

1. Introdução

O setor agrícola é de grande importância para a economia mundial. No Brasil, este setor representa aproximadamente um Produto Interno Bruto (PIB) de R\$ 524,5 bilhões, que correspondente a 30,2% de todo o PIB da economia nacional (CNA, 2005). Atualmente as estatísticas agrícolas oficiais no Brasil são realizadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE), por meio do Levantamento Sistemático da Produção Agropecuária (LSPA), com um método subjetivo. Segundo Collares *et al.* (1993), este método baseia-se em informações colhidas de produtores, cooperativas e instituições de crédito agrícola.

Segundo Luiz (2005), um dos grandes desafios para melhorar as estatísticas agrícolas no Brasil é estabelecer um método que permita associar de maneira definitiva a tecnologia do sensoriamento remoto orbital aos procedimentos rotineiramente utilizados nos levantamentos das estatísticas agrícolas. Portanto, o uso de técnicas de sensoriamento remoto pode ser uma alternativa para tornar mais objetiva a identificação de panoramas agrícolas no Brasil. Imagens de sensores de resolução espacial intermediária como TM e ETM/Landsat ou CCD/CBERS são importantes na representação dos alvos agrícolas em escalas compatíveis para análise ou fins de estimativas de áreas agrícolas.

Para um país com tanto potencial agrícola, a previsão de safras se torna imprescindível, o que justifica que pesquisas sejam continuamente geradas para o desenvolvimento e aprimoramento de metodologias que permitam a aquisição de dados mais confiáveis de produtividade e área cultivada para a estimativa de safras (Sanchez, 2004). Com vista a uma estimativa agrícola com maior objetividade, o Projeto Geosafras, coordenado pela CONAB (Companhia Nacional de Abastecimento), vem desenvolvendo para alguns Estados brasileiros, com base em amostragem por município, um método de amostragem simples para estimativa de área agrícola de alguns cultivos. Este (Epiphanyo et al., 2002; Luiz e Epiphanyo, 2005) ainda é executado experimentalmente. Diferencia-se do método realizado pelo IBGE por utilizar imagens de satélite para auxiliar na determinação de áreas agrícolas e por basear-se num procedimento estatístico-amostral.

Tem havido constantes esforços no Projeto Geosafras para tornar as estimativas agrícolas mais eficientes. Neste contexto, o presente trabalho tem o objetivo de auxiliar no aperfeiçoamento do método de amostragem simples utilizado pelo Projeto Geosafras,

utilizando como estudo de caso a estimativa de área plantada com a cultura da soja nas últimas seis safras do município de Luís Eduardo Magalhães (LEM), Bahia. Tal aperfeiçoamento, para nível de distinção, será denominado como Geosafras Otimizado.

2. Área de Estudo

A área de estudo localiza-se na região oeste do Estado da Bahia, e corresponde ao município de Luís Eduardo Magalhães, como mostra a **Figura 1**:



Figura 1 – Localização do município de Luís Eduardo Magalhães (LEM), BA.

O município de LEM tem uma população estimada em 22.081 habitantes, com uma extensão territorial de 4.015 km² (IBGE, 2005). Segundo a SEI (2005), este município apresenta solos variando de arenosos a argilo-arenosos, com relevo plano a suave ondulado, e tipo climático B1wA' (critério de *Thornthwaite*). O clima é úmido e apresenta índice hídrico de 40% a 60% com excedente superior a 600 mm/ano, com maior concentração do regime de chuvas entre os meses de outubro e março. A agricultura é tecnificada, e há muitas áreas de cultivos com sistema de irrigação por “*pivô central*”, principalmente na porção leste do município, onde a concentração da agricultura é menor.

A principal atividade econômica é a agricultura, sendo a soja, o milho e o algodão os principais cultivos em ordem econômica.

O setor do agronegócio foi um fator primordial para o desenvolvimento do município. Segundo Francisco (2004), Luís Eduardo Magalhães vem crescendo em um ritmo acelerado, a partir da sua emancipação, em 2000. Este rápido desenvolvimento do setor agrícola nos últimos anos, aliado à falta de conhecimento sobre a dinâmica agrícola local, foram fatores primordiais que influenciaram na escolha da área de estudo.

3. Procedimentos Metodológicos

Como descrito anteriormente, o Projeto GeoSafras utiliza um sistema de amostragem simples sobre imagens de satélite, com o objetivo de estimar áreas agrícolas municipais, como base para estimativas estaduais.

Nos fundamentos básicos de estatística, como mencionado por Meyer (1983), o aumento do número de amostras é uma das alternativas utilizadas para obter uma melhor estimativa e diminuir, assim, os erros inerentes ao sistema de amostragem. Entretanto, em algumas situações, este princípio não pode ser aplicado, principalmente quando o acréscimo no número de amostras leva a uma elevação dos custos de execução do levantamento.

Este trabalho objetiva a otimização do sistema de amostragem proposto pelo Projeto GeoSafras, mantendo-se inalterado o número de amostras. A fim de melhorar a precisão do levantamento, propõe-se a “diminuição” da área municipal, com a eliminação de locais

impróprios para o estabelecimento agrícola, tais como: lâminas d'água, brejo, unidades de conservação, entre outros. Para melhor entendimento do funcionamento deste método, o exemplo abaixo demonstra detalhadamente como se dará a otimização.

O método GeoSafras atual parte do princípio que a quantidade de pixels distribuídos na área do município representa a população. Portanto, cada ponto amostrado, fruto da distribuição aleatória, tem a mesma probabilidade de ocorrência em qualquer pixel (**Figura 2**).

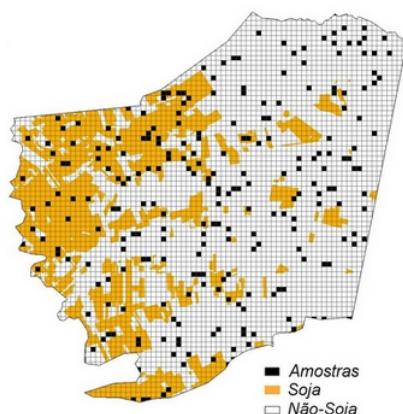


Figura 2 – Sorteio aleatório das amostras.

Inicialmente, a imagem é numerada de 1 até “N”, onde “N” representa a quantidade total de pixels presentes no município (população), a qual depende da área municipal e da resolução espacial utilizada. Posteriormente são sorteadas amostras (n) de forma aleatória, e é feita a verificação de campo, que indica o valor de acertos ou ocorrências para uma determinada cultura. Com isso, é possível calcular a probabilidade (p) de ocorrência desta cultura no município. Visto o valor de probabilidade, é possível calcular o valor “q” (1-p) pela diferença, uma vez que ambos são complementares. Sabido os valores da área que representa um pixel, a probabilidade de ocorrência de uma determinada cultura e o número populacional, é possível estimar o valor de área plantada para esta cultura no município, compondo assim, os procedimentos necessários para um sistema de amostragem simples.

$$N = \frac{A}{a} ; p = \frac{ta}{n} ; q = 1 - p ; n = \frac{(Z_{\alpha/2}) \cdot pq}{e^2} ; EAP = a \cdot p \cdot N$$

N = População total;

A = Área municipal;

a = Área do pixel;

ta = Acertos, conforme verificação de campo;

p = Probabilidade;

q = Complementar à probabilidade;

n = Número de amostras;

$Z_{\alpha/2}$ = Valor tabelado (segundo o grau de significância adotado);

e = Erro admitido no sistema de amostragem;

EAP = Estimativa de área plantada.

O método aqui proposto, denominado GeoSafras Otimizado, utiliza o mesmo procedimento descrito acima, porém busca uma redução da área municipal com base na exclusão de áreas impróprias para o estabelecimento agrícola (lâminas d'água, unidades de

conservação, entre outras). Consumada a redução da área municipal o valor de “N” diminui. Considerando o mesmo número de amostras, conseqüentemente o valor de probabilidade tende a aumentar, diminuindo assim, o coeficiente de variação do sistema. Tal estratégia reflete em uma estimativa de área com uma menor magnitude de variação nos resultados estimados, mantendo-se o mesmo número de amostras utilizado no método inicial.

Para a avaliação do método foi construída uma base de dados de área plantada com o “cultivo da soja” nas últimas seis safras (2001 a 2006), exceto para o ano de 2005, devido à impossibilidade de obtenção de imagens livres de nuvens no período do calendário agrícola deste cultivo. Tal base foi obtida por meio de interpretação visual (**Figura 3**) em imagens orbitais dos Sensores TM e ETM+ a bordo dos satélites Landsat-5 e 7, respectivamente. A interpretação foi realizada com base no reconhecimento de padrões utilizando uma composição colorida falsa cor 4R 5G 3B, apoiada em trabalho de campo realizado em janeiro de 2006, com índice Kappa de 0,96 para esta safra, entre as possibilidades: Soja e Não-Soja. Presume-se que esta interpretação forneça uma base de dados confiável para a aferição dos métodos aqui expostos, além de possibilitar uma análise comparativa entre os resultados oficiais de estimativa de área plantada, fornecidos pelo IBGE.



Figura 3 – Padrões obtidos por meio de interpretação visual, composição - 4R 5G 3B.

Com o objetivo de adotar um nível de significância de 5% foi calculado um número de 250 amostras para o município LEM, as quais foram sorteadas aleatoriamente para os dois métodos: Geosafras e Geosafras Otimizado. Na estimativa da área plantada do cultivo da soja para os dois métodos, o município LEM foi dividido em células de 900m² (30m x 30m). Com a eliminação de áreas impróprias para o cultivo agrícola, a área útil municipal passou de 401.515ha para 361.364ha, o que corresponde a uma redução de aproximadamente 10%, compondo uma população de 4.461.277 e 4.015.155 pixels, respectivamente.

4. Resultados e Discussão

O plantio da soja no município de LEM apresenta uma única safra durante o ano, que se inicia em novembro, com colheita prevista para fevereiro. Em algumas situações, representando uma minoria, os talhões são plantados entre os meses de dezembro e janeiro. A **Figura 4** apresenta a distribuição espacial e quantitativa da área plantada com o cultivo da soja de 2001 até 2006, exceto para o ano de 2005. Em 2001, o cultivo da soja ocupava uma área de 127.279ha, o que corresponde a uma porção bem significativa da área total do município (31,7%). Com o aumento de área plantada com esse cultivo nas últimas seis safras houve um incremento de 13.432ha, e esta porção passou para 35,0%, o que corresponde a uma projeção de avanço da soja em torno de 3,3% em relação à área municipal.

Até o ano de 2004 a tendência de crescimento da área plantada apresenta-se mais acentuada com um incremento médio de 4.166ha/ano, enquanto que a partir de 2004 o incremento foi de apenas 934ha/ano. É evidente que vários fatores podem influenciar no valor de área plantada de uma determinada cultura, porém a linha de tendência aponta para uma certa estabilidade nos últimos anos (**Figura 5**).

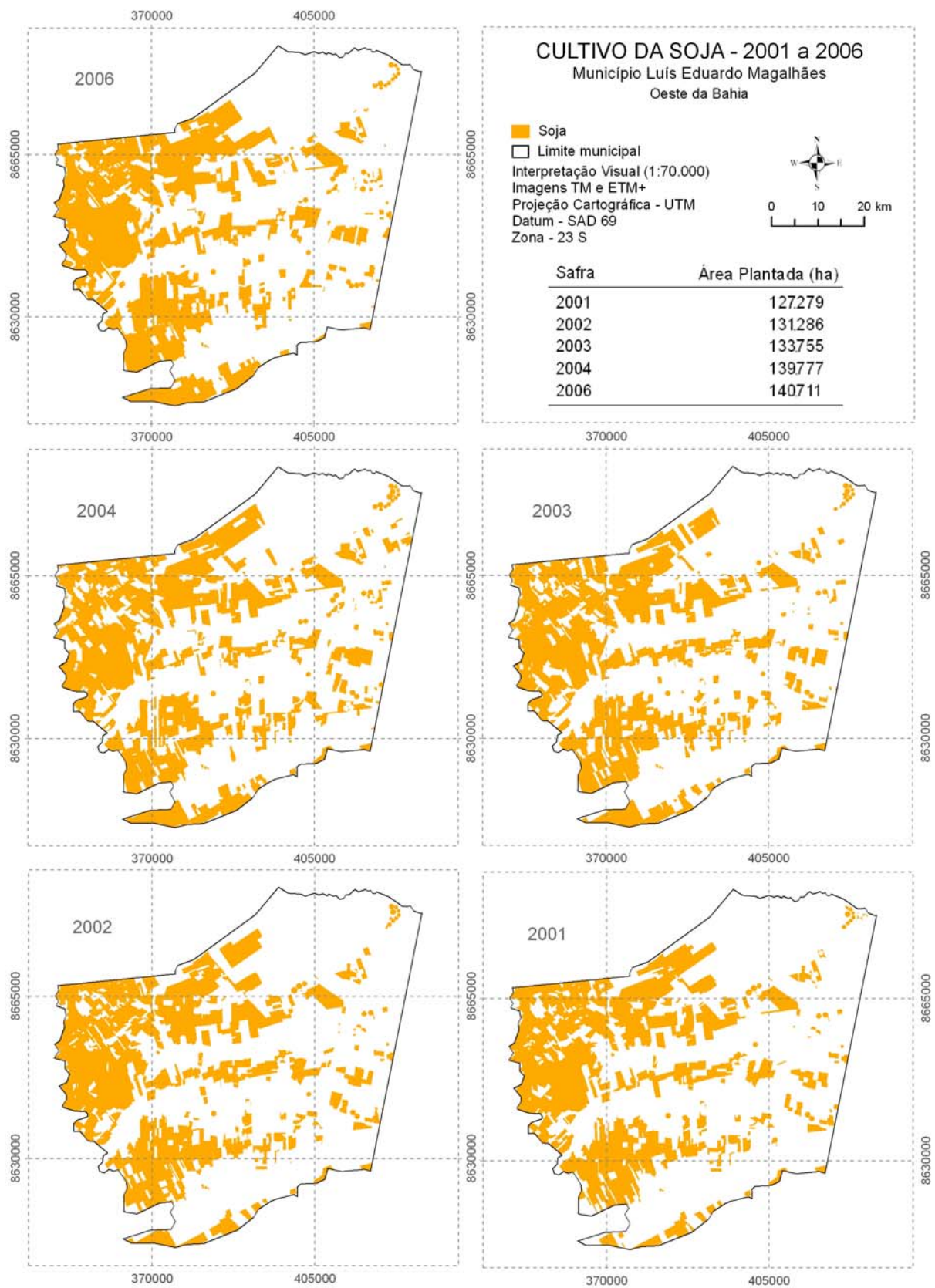


Figura 4 – Distribuição espacial do cultivo da soja em LEM de 2001 a 2006.

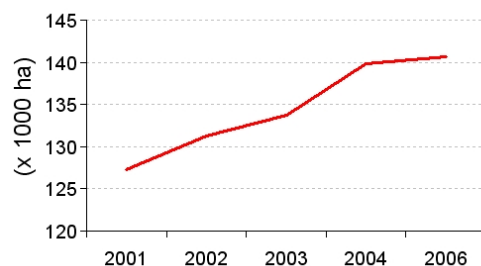


Figura 5 – Linha de tendência do incremento de área plantada com a cultura da soja no município LEM com base em interpretação visual.

No que diz respeito à comparação entre os dois métodos: Geosafas e Geosafas Otimizado, tema principal deste trabalho, fica evidente que o método Geosafas Otimizado foi mais eficiente na determinação de área plantada com o cultivo da soja no município de Luís Eduardo Magalhães (**Figura 6**). Tal comparação foi realizada com base nos resultados da interpretação visual (usada como referência).

A melhoria do método proposto ocorreu em função da redução na área municipal, devido à eliminação de áreas impróprias para o cultivo agrícola, visto que houve um aumento no valor de probabilidade e conseqüentemente uma redução no coeficiente de variação associado ao sistema.

Sabido da necessidade de aumentar o valor de probabilidade, e diante da impossibilidade de acréscimo no número de amostras por conta dos custos adicionais atribuídos ao sistema, a redução populacional (redução da área municipal) foi fator primordial para a melhoria dos valores estimados no método Geosafas Otimizado, podendo assim, manter constante o número de amostras. Para as cinco safras em análise, o aumento médio no valor de probabilidade foi de 2,68%, o que significa uma fonte média de erro em torno de 5.203ha/safra.

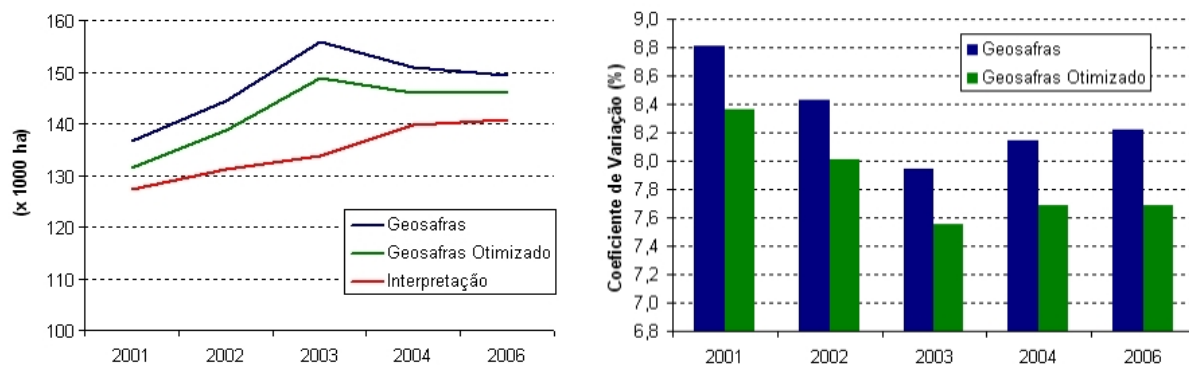


Figura 6 – Comparação entre os dois métodos na estimativa de área plantada e na amplitude dos coeficientes de variação.

Em todas as safras avaliadas houve uma redução significativa das fontes de erros na estimativa de área plantada pelo método Geosafas Otimizado, principalmente para safra de 2006 onde o coeficiente de variação passou de 8,21% para 7,68%. Tal redução reflete em uma estimativa com uma menor magnitude de variação nos resultados estimados de área plantada com o cultivo da soja neste município.

Com base nas estatísticas oficiais de área plantada municipal, realizadas pelo IBGE por meio do Levantamento Sistemático da Produção Agrícola – LSPA, fez-se uma análise comparativa dessas estatísticas com os valores de área plantada com a cultura da soja em LEM obtidos a partir da interpretação visual, com base no reconhecimento de padrões

estabelecidos em campo, realizado em janeiro de 2006. Em linhas gerais, os resultados mostram que os valores estimados são distintos e não apresentam um cenário linear em relação às duas curvas de tendência (**Figura 7**). A safra de 2002 foi a que apresentou valores mais discrepantes, o que corresponde a uma diferença de 42.406ha.

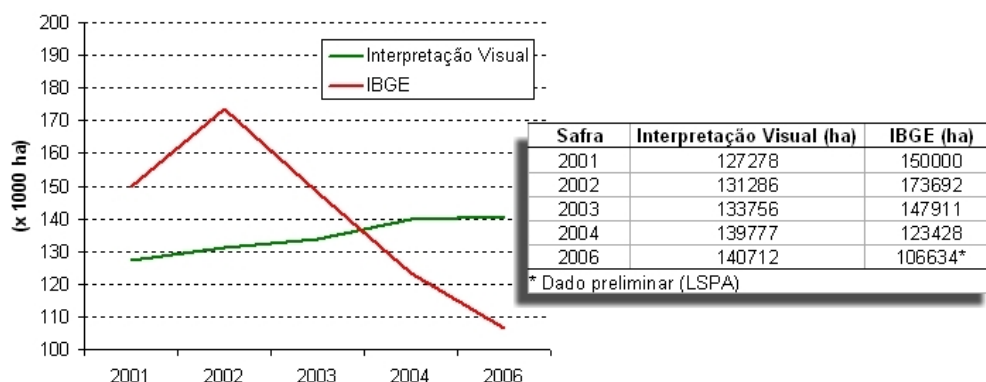


Figura 7 – Comparativo entre as estimativas oficiais e os resultados obtidos por meio de interpretação visual.

Tomando como referência a interpretação visual, optou-se em comparar os erros obtidos pelo método Geosafra Otimizado e as estatísticas oficiais realizadas pelo IBGE, para as cinco safras em análise. Para isso, partiu-se do princípio que tais variações, quantificadas em hectares, são correspondentes às médias de duas populações quaisquer.

As **Figuras 6 e 7** mostram que existiram erros de estimativa tanto no método Geosafra Otimizado quanto no método oficial (IBGE). Entretanto, é pertinente verificar se existe diferença entre os dois métodos. Tal avaliação foi possível pela aplicação do *Teste de Student-T* (**Tabela 1**). Inicialmente, para a escolha do *Teste T*, foi realizado uma análise de homogeneidade das variâncias, por meio do *Teste Fisher-F* (**Tabela 2**). Ambos foram realizados a 5% de significância.

Tabela 1 - Análise das variâncias

Teste F	(Geosafra Otimizado-Interp. Visual) - ha	(IBGE-Interp. Visual) - ha
Média	7.671,1	25.942,0
Variância	18.775.782,2	144.646.302,5
Observações	5	5
gl	4	4
F	0,129804785	
P(F<=f) uni-caudal	0,036566924	
F Crítico uni-caudal	0,156537812	

As variâncias foram consideradas heterocedásticas, ou seja, apresentam diferença estatística a 5% de significância, visto que o valor “p” foi menor que o nível de significância adotado (0,05). Este fato indica que o Teste T deve assumir variâncias distintas.

Tabela 2 – Comparação da significância estatística entre os dois métodos.

Teste T	(Geosafra Otimizado-Interp. Visual) - ha	(IBGE-Interp. Visual) - ha
Média	7.671,0	25.942,0
Variância	18.775.782,2	144.646.302,5
Observações	5	5
Hipótese da diferença de média	0	
gl	5	
t Stat	-3,195882039	
P(T<=t) uni-caudal	0,012052936	
t Crítico uni-caudal	2,015048372	

Segundo o Teste T unilateral, com 5% de significância, houve diferença estatística entre os erros de estimativa de área plantada com o cultivo da soja em LEM, tomando como base os valores obtidos por interpretação visual (referência). Tal fato pode ser verificado pela análise do valor “p”, o qual foi menor que o nível de significância adotado. Os erros oriundos das estimativas oficiais foram superiores quando comparados aos erros obtidos pelo método Geosafra Otimizado.

4. Conclusões

O método Geosafra Otimizado foi mais eficiente na estimativa de área plantada com o cultivo da soja em LEM em todas as safras analisadas. A redução da área municipal aumentou o valor de probabilidade, conseqüentemente, houve uma redução do coeficiente de variação, refletindo em uma estimativa com menor magnitude de erros. Isso leva a uma estimativa com menores custos, os quais são na maioria vinculados aos levantamentos de campo. Para a redução de custos, aconselha-se manter o mesmo coeficiente de variação para os dois métodos, podendo assim, reduzir o número de amostras do método Geosafra Otimizado, a ponto de obter resultados equiparados de estimativa de área plantada.

Houve diferença estatística a 5% de significância entre os erros apresentados pelas estimativas oficiais do IBGE, e os erros obtidos pelo método Geosafra Otimizado, quando comparados à interpretação visual.

Referências

- Agropecuária Brasileira. **Confederação da Agricultura e Pecuária do Brasil – CNA**, 2005. Disponível em: <<http://www.cna.org.br/cna/publicacao/noticia.wsp?tmp.noticia=4256>>. Acesso em 20 jan. 2006.
- 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., Curitiba, 1993. **Anais...** São José dos Campos: INPE, 1993. p. 450-453.
- Epiphânio, 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, n.2, p.187-197. 2002.
- Estatísticas Municipais. **Instituto Brasileiro de Geografia e Estatística – IBGE**, 2005. Disponível em: <<http://www.ibge.gov.br/cidadesat/default.php>>. Acesso em 19 jan. 2006.
- Francisco, 2004. **Folha de São Paulo**. Disponível em: <<http://www.agr.feis.unesp.br/fsp14032004.php>>. Acesso em 29 jan. 2006.
- Informações Geoambientais. **Superintendência de Estudos Econômicos e Sociais da Bahia SEI**, 2005. Disponível em: <http://www.sei.ba.gov.br/geoambientais/index_geoamb_carto.php>. Acesso em 18 jan. 2006.
- Luiz, A. J. B. Estatísticas agrícolas por amostragem auxiliadas por sensoriamento remoto. In: Simpósio Brasileiro de Sensoriamento Remoto, 10., 2005, Goiânia. **Anais...** São José dos Campos: INPE, 2005. p. 16-21. Disponível na biblioteca digital URLib: <<http://marte.dpi.inpe.br/col/ltid.inpe.br/sbsr/2004/11.19.22.11/doc/181.pdf>>. Acesso em 16 jan. 2006
- Luiz, A. J. B.; Epiphânio, J. C. N. Amostragem por pontos em imagens de sensoriamento remoto para estimativa de área plantada por município. In: Simpósio Brasileiro de Sensoriamento Remoto, 10., 2001, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2001. p. 111-118. Disponível na biblioteca digital URLib: <<http://marte.dpi.inpe.br/col/dpi.inpe.br/lise/2001/09.13.10.55/doc/0111.118.116.pdf>>. Acesso em 20 jan. 2006.
- Meyer, P. L. **Probabilidade: aplicações à estatística**; tradução do Prof. Ruy de C. B. Lourenço Filho. Rio de Janeiro: LTC, 1983, 426p.
- Sanches, I. D’A.; Epiphânio, J. C. N.; Formaggio, A. R. Culturas agrícolas em imagens multitemporais do Satélite Landsat. **Agricultura em São Paulo**, v. 52, n. 1, p. 83-96, 2005.