

# ESTIMATIVA DE ÁREA DA CULTURA DE SOJA COM IMAGENS LANDSAT 5/TM POR MEIO DE MATRIZ DE ERROS

ERIVELTO MERCANTE<sup>[1]</sup>, JOÃO F. G. ANTUNES<sup>[2]</sup>, RUBENS A. C. LAMPARELLI<sup>[3]</sup>, JANSLE V. ROCHA<sup>[4]</sup>

[1] Engº Agrícola, Doutorando Engenharia Agrícola e Bolsista CNPq, FEAGRI/UNICAMP, Campinas – SP, (0xx19) 3788-1060, erivelto.mercante@agr.unicamp.br

[2] Mestre em Engenharia Agrícola, Pesquisador Embrapa Informática Agropecuária, Campinas – SP.

[3] Engº Agrícola, Pesquisador Doutor, CEPAGRI/UNICAMP, Campinas – SP.

[4] Engº Agrícola, Professor Doutor Engenharia Agrícola, FEAGRI/UNICAMP, Campinas – SP.

Escrito para apresentação no  
XXXV Congresso Brasileiro de Engenharia Agrícola  
31 de julho a 4 de agosto de 2006 - João Pessoa - PB

**RESUMO:** A área plantada é de fundamental importância para a previsão de safras no Brasil. Dados de sensoriamento remoto podem ser utilizados para monitoramento de culturas agrícolas de interesse estratégico. Este trabalho tem como objetivo estimar a área da cultura de soja para oito municípios do Estado do Paraná, em duas safras agrícolas 2003/2004 e 2004/2005, utilizando um método baseado na matriz de erros gerada a partir da classificação supervisionada de imagens Landsat 5/TM. As classificações foram realizadas por meio dos algoritmos Paralelepípedo e Máxima Verossimilhança, dando origem ao que foi denominado de máscara de soja. As estimativas de área obtidas a partir da máscara de soja, corrigidas pela matriz de erros, apresentaram alta correlação com as estimativas de área de soja oficial. Portanto, a estimativa de área plantada utilizando dados orbitais demonstrou ser viável, uma vez que é realizada de forma objetiva, direta, apresenta menor custo e pode ser feita com antecedência em relação aos métodos convencionais.

**PALAVRAS-CHAVES:** previsão de safras; classificação de imagens; Índice Kappa

## SOYBEAN CROP AREA ESTIMATION WITH LANDSAT 5/TM IMAGES BY ERROR MATRIX

**ABSTRACT:** Crop area is fundamental importance in crop yield forecasting in Brazil. Remote sensing data can be used for monitoring strategic agricultural areas. These work it has as objective the soybean crop area estimate for eight municipalities of the Paraná State, in two harvest 2003/2004 e 2004/2005, based on error matrix generated from supervised classification of Landsat 5/TM images. The classifications were realized by Parallelepiped and Maximum Likelihood algorithms, producing that called as soybean mask. The area estimation obtained from soybean mask corrected by error matrix, presented high correlation with official estimations and those obtained by direct expansion. Hence, a soybean crop area estimate using orbital data show to be viable that is realized objective and direct form, present less cost and can be made prior to the conventional methods.

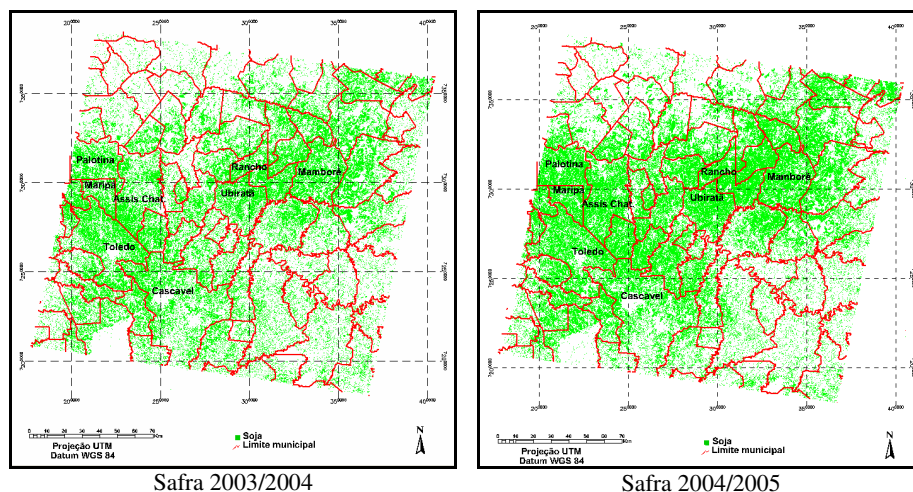
**KEYWORDS:** crop yield forecasting, images classification, Kappa Index.

**INTRODUÇÃO:** A soja vem se consolidando como principal cultura do agronegócio brasileiro, assumindo grande importância econômica nas exportações. O Estado do Paraná nas ultimas safras figura como o segundo maior produtor do Brasil, atrás apenas do Estado do Mato Grosso. A região Oeste do Paraná, local de estudo deste trabalho, responde por mais de 25% do total da produção do Estado. Nos últimos anos, a previsão de safras agrícolas no Brasil está sofrendo grandes transformações para se tornar menos subjetiva. Devido a isso, vem agregando ferramentas estatísticas e de acompanhamento de culturas. O sensoriamento remoto pode ser utilizado para proporcionar uma visão sinóptica de áreas cultivadas, gerando informações mais objetivas e precisas. As estimativas de área plantada em países mais desenvolvidos vêm sendo realizadas já há muito tempo utilizando-se sensoriamento remoto. Desde COHEN (1960), a técnica de matriz de erros, bem como as métricas de Exatidão Global e Índice Kappa, vêm sendo utilizadas para determinar a acurácia de classificações

digitais mediante utilização de imagens de satélites e, a partir disso, proceder a estimativa de área de classes de cobertura do solo. Segundo RICHARDS & JIA (1999), a matriz de erros permite ainda avaliar o desempenho da classificação correta de uma classe individual, particularmente quando um pequeno número de classes de uso do solo é de interesse, como por exemplo, na estimativa de área de uma cultura agrícola. Nesse contexto, o trabalho tem como objetivo realizar uma estimativa de área plantada de soja para oito municípios por meio da técnica de matriz de erros.

**MATERIAL E MÉTODOS:** A área de estudo é compreendida por oito municípios da região Oeste do Estado do Paraná, coberta pela cena do satélite Landsat 5/TM, órbita 223/ponto 77. As informações cartográficas dos limites municipais da região em estudo foram obtidas do IBGE, malha ano 2001. Para o estudo foram utilizadas imagens de duas safras anos 2003/2004 (seis imagens) e 2004/2005 (cinco imagens), selecionadas de forma a abranger todo o período do ciclo de desenvolvimento da cultura da soja na região, ou seja, desde do mês de Novembro (semeadura) a Fevereiro (colheita) de cada safra. As imagens foram processadas no software ENVI 4.0, com projeção cartográfica UTM, Zona 22 Sul, Datum WGS-84. A primeira etapa de processamento das imagens foi o georreferenciamento, a seguir foi executada a transformação radiométrica, ou seja, a conversão do número digital (ND) para valores de reflectância aparente (no topo da atmosfera), conforme CHANDER e MARKHAM (2003). Baseado em composições coloridas RGB-453 falsa-cor, foram coletadas as amostras de pixels que correspondem à cultura de soja, tomando como base a evolução de uma área agrícola ocupada com a cultura de soja durante as duas safras monitoradas, situada próxima ao município de Cascavel/PR. A partir disso, foram então executadas as classificações supervisionadas por meio dos algoritmos de classificação Paralelepípedo e Máxima Verossimilhança (MaxVer). Nas classificações foram utilizadas as bandas espectrais 3, 4 e 5 que proporcionam menor correlação entre as informações espectrais das três bandas, atribuindo maior chance de identificação dos objetos e, também, a banda 7 como forma de minimizar os efeitos atmosféricos pois é menos influenciada pelos efeitos de absorção e espalhamento dos aerossóis e vapor de água presentes na atmosfera, devido ao tamanho do comprimento de onda. Ao final, as imagens classificadas foram sobrepostas com o operador lógico *OR*, onde todos os pixels classificados como soja nas imagens que cobrem o ciclo de desenvolvimento da cultura na região em estudo, originaram a máscara de soja, nas duas safras monitoradas. Após gerar a máscara de soja foram realizados os cálculos da área plantada de soja para oito municípios da região em estudo. A geração das amostras para verificar a qualidade do mapa temático se baseou no chamado painel amostral que, resumidamente, caracteriza-se pela distribuição aleatória de pontos amostrais dentro dos limites dos municípios, com o objetivo de se realizar um levantamento das classes de uso do solo de cada ponto (LUIZ, 2003). Por fornecer uma boa relação custo-benefício, foram utilizados 100 pontos amostrais como referência terrestre em cada um dos oito municípios e nas duas safras monitoradas, após isso foi possível então realizar a avaliação da acurácia das classificações, por meio do Índice Kappa (COHEN, 1960). O método a ser utilizado neste estudo, proposto por JUPP (1989), normaliza a matriz de erros pelas linhas e por isso é denominado direto. Consiste na probabilidade de um pixel que foi previamente atribuído a uma certa classe, na realidade pertencer a todas as classes examinadas, pode ser estimada pela padronização das linhas e o total de pixels das classes de uso do solo determinadas pela imagem de satélite. Baseado nisso, este método, chamado de direto, é considerado mais natural para corrigir as probabilidades de designação das classes por meio das matrizes de proporções. As equações e a matriz de proporções utilizadas para aplicar o método de matriz de erros na correção das estimativas de área de soja podem ser vistas no trabalho de JUPP (1989).

**RESULTADOS E DISCUSSÃO:** O georreferenciamento produziu um erro quadrático médio (RMS) de aproximadamente dez metros. Essa precisão é satisfatória, pois considerando-se um pixel de 30 m do Landsat 5/TM, representa um deslocamento máximo de um terço do pixel. Após os processamentos realizados nas imagens e a partir das classificações supervisionadas e das sobreposições das várias datas das imagens em cada uma das safras monitoradas, foram construídas as máscaras de soja, conforme apresentado na Figura 1, destacando os oito municípios produtores de soja situados na região Oeste do Paraná. Posteriormente foram realizadas as avaliações da qualidade (acurácia) das informações geradas pelas classificações, a partir dos 100 pontos amostrados nos oito municípios estudados.



**Figura 1.** Máscaras de soja gerada a partir de imagens Landsat 5/TM.

A Tabela 1 apresenta os valores de área calculados por contagem de pixels nas máscaras de soja e os valores do Índice Kappa para os municípios estudados. A área foi obtida multiplicando-se o número total de pixels da classe Soja pela área do pixel Landsat 5/TM igual a 900 m<sup>2</sup> (30 x 30 m), dividindo-se por 10.000 para conversão em hectares (ha). O Índice Kappa envolve a matriz de erros completa, incorporando os elementos das linhas e colunas marginais referentes aos erros de classificação, além dos que se situam na diagonal principal. Os valores para todos os municípios estão acima de 0,624 para as duas safras. Quando comparados com os dados de avaliação da qualidade de classificações proposto por LANDIS e KOCH (1977), são considerados de qualidade muito boa ou excelente. Em complemento a isso, FOODY (2002) revisando os métodos de análise de acurácia da classificação sugere que os níveis recomendados para Índice Kappa devem estar em torno de 0,85, para se obter uma boa classificação que forneça dados precisos. De qualquer maneira, pode-se concluir que os classificadores apresentaram limitações ao efetuar o reconhecimento da cultura, apesar do padrão da soja ter sido bem delimitado nas amostras de treinamento.

**Tabela 1:** Área e Índice Kappa da máscara de soja nos oito municípios nas duas safras monitoradas.

| Município           | Área Máscara Soja (ha) |           | Índice Kappa |           |
|---------------------|------------------------|-----------|--------------|-----------|
|                     | safra                  |           | safra        |           |
|                     | 2003/2004              | 2004/2005 | 2003/2004    | 2004/2005 |
| Assis Chateaubriand | 48.724,11              | 66.968,55 | 0,708        | 0,809     |
| Cascavel            | 82.434,42              | 88.407,18 | 0,808        | 0,760     |
| Mamborê             | 46.961,46              | 56.586,06 | 0,799        | 0,909     |
| Maripá              | 16.324,11              | 21.029,4  | 0,650        | 0,777     |
| Palotina            | 34.002,09              | 45.741,33 | 0,635        | 0,859     |
| Rancho A. do Oeste  | 14.565,33              | 17.655,93 | 0,767        | 0,852     |
| Toledo              | 66.565,89              | 75.338,91 | 0,746        | 0,624     |
| Ubatuba             | 36.069,84              | 47.587,59 | 0,789        | 0,829     |

Devido ao limitante de acurácia inerente dos classificadores booleanos, a área dos oito municípios da máscara de soja foi corrigida por meio de matrizes de erros, conforme o método proposto por JUPP (1989). Diante disso, os dados de área da cultura da soja calculados pela matriz de erros para cada município estão apresentados na Tabela 2. Por fim, foram confrontadas as estimativas de área de soja calculadas a partir da máscara de soja baseada na matriz de erros, com a estimativa oficial realizada pela Secretaria da Agricultura e Abastecimento do Paraná, Departamento de Economia Rural (SEAB/DERAL, 2005), para os oito municípios nas duas safras, conforme mostrado na Tabela 2. O erro relativo (ER) entre as estimativas de área foi calculado pelo valor estimado menos o valor de referência, dividindo-se pelo valor de referência. Tomando como referência a estimativa oficial da SEAB/DERAL, a correlação com as áreas calculadas pela máscara de soja baseada em matriz de erros foi de aproximadamente 0,97 para as duas safras monitoradas, análises estatisticamente significativas. Isso indica que os dados são fortemente correlacionados e possuem a mesma tendência, movendo-se em perfeita proporção na mesma direção.

**Tabela 2:** Estimativas de áreas de soja e os erros relativos nas safras 2003/2004 e 2004/2005.

| Município           | Safr 2003/2004 |                |        | Safr 2004/2005 |                |        |
|---------------------|----------------|----------------|--------|----------------|----------------|--------|
|                     | Área (ha)      |                | ER (%) | Área (ha)      |                | ER (%) |
|                     | Oficial*       | Máscara Soja** |        | Oficial*       | Máscara Soja** |        |
| Assis Chateaubriand | 72.000,00      | 70.918,97      | -1,50  | 72.200,00      | 78.544,50      | 8,79   |
| Cascavel            | 82.000,00      | 78.388,29      | -4,40  | 87.700,00      | 83.171,75      | -5,16  |
| Mamborê             | 54.000,00      | 57.744,34      | 6,93   | 54.000,00      | 58.039,47      | 7,48   |
| Maripá              | 21.500,00      | 12.797,04      | -40,48 | 21.400,00      | 19.232,47      | -10,13 |
| Palotina            | 43.300,00      | 34.987,46      | -19,20 | 43.500,00      | 52.060,76      | 19,68  |
| Rancho A. do Oeste  | 15.400,00      | 20.560,44      | 33,51  | 18.000,00      | 20.768,23      | 15,38  |
| Toledo              | 68.600,00      | 72.194,60      | 5,24   | 69.300,00      | 77.732,80      | 12,17  |
| Ubiratã             | 51.200,00      | 48.409,06      | -5,45  | 51.400,00      | 49.015,67      | -4,64  |

Fonte: \*SEAB/DERAL (2005) - atualização 20/01/2005. \*\* Estimativa de área corrigida baseada em matriz de erros

De forma geral, observa-se que os valores do ER são relativamente altos para três dos oito municípios na safra 2003/2004 (Maripá - subestimou 40,48 %, Palotina - subestimou 19,20% e Rancho A. do Oeste - superestimou 33,51%), sendo que para os demais, os valores de ER ficaram abaixo de 7%. Para a safra 2004/2005 a mesma tendência também foi observada, ou seja, os três municípios continuam com os maiores valores de ER (Maripá - subestimou 10,13 %, Palotina - superestimou 19,20% e Rancho A. do Oeste - superestimou 15,38%), entretanto, o município de Toledo também apresentou alto ER em relação aos demais municípios para esta safra, superestimando 12,17%. Todos os outros municípios nas duas safras apresentaram valores de ER abaixo de 9%. As maiores diferenças na estimativa de área encontradas para três municípios, pode ter ocorrido pelo fato da estimativa de área de soja oficial ser realizada de forma subjetiva (considerando indicadores como venda de insumos, demanda por crédito e visitas a propriedades rurais) que pode embutir uma maior imprecisão, ao contrário da calculada por matriz de erros, feita de forma objetiva por meio de uma série de imagens Landsat 5/TM com resolução de 30 m. Apesar dessas diferenças, vale destacar, a grande similaridade entre todos os outros municípios. Cabe ressaltar, também, que ao final do ciclo vegetativo da soja na região de estudo foi possível obter a estimativa de área plantada total da cultura, enquanto que a estimativa oficial demanda um tempo bem maior até ser divulgada definitivamente.

**CONCLUSÃO:** Por meio da sobreposição das classificações geradas em cada uma das datas de aquisição das imagens, foi possível contemplar as áreas de soja nos diferentes estádios de desenvolvimento da cultura, compondo assim a máscara de soja. A estimativa de área utilizando imagens de satélites demonstrou ser viável, uma vez que é realizada de forma objetiva, direta, possui menor custo e pode ser realizada com antecedência em relação aos métodos convencionais. O método de estimativa de área baseada em matriz de erros comportou-se de forma satisfatória, uma vez que foi possível corrigir as áreas estimadas pela classificação, obtendo-se alta correlação com a estimativa oficial.

## BIBLIOGRAFIA:

- CHANDER, G.; MARKHAM, B. Revised Landsat 5 TM Radiometric Calibration Procedures and Postcalibration Dynamic Ranges. **IEEE Transactions on Geoscience and Remote Sensing**. v. 41, n. 11, p. 2764-2677, 2003.
- COHEN, J. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**. v. 20, n. 1, p. 37-46, 1960.
- FOODY, G. M. Status of Land Cover Classification Accuracy Assessment. **Remote Sensing of Environment**, v. 80, n. 1, p. 185-201, 2002.
- JUPP, D. L. B. The Stability of Global Estimates from Confusion Matrices. **International Journal of Remote Sensing**, v. 10, n. 9, p. 1563-1569, 1989.
- LANDIS, J. R.; KOCH, G. G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, v. 33, n. 1, p. 159-174, 1977.
- LUIZ, A. J. B. **Estatísticas agrícolas por amostragem auxiliadas pelo sensoriamento remoto**. INPE, São José dos Campos, SP, 2003. (Tese de Doutorado).
- RICHARDS, J. A.; JIA, X., **Remote sensing digital image analysis: an introduction**, 3. edition. New York: Springer Verlag, 363p., 1999.