

O uso de máscaras para seleção automática de áreas plantadas com soja no estado de São Paulo

Michelle Cristina Araujo Picoli ¹
Jansle Vieira Rocha ¹
Júlio César Dalla Mora Esquerdo ²
Rubens Augusto Camargo Lamparelli ³

¹ Universidade Estadual de Campinas – UNICAMP/FEAGRI
Caixa Postal 6011 - 13083-875 - Campinas - SP, Brasil
{michelle.picoli, jansle.rocha}@feagri.unicamp.br}

² Embrapa Informática Agropecuária
13083-886 – Campinas – SP, Brasil
julio@cnptia.embrapa.br

³ Centro de Pesquisas Meteorológicas Aplicadas a Agricultura - CEPAGRI
13083-970 - Campinas - SP, Brasil
rubens@cpa.unicamp.br

Abstract. This paper aimed to develop an automatic procedure for soybean crop identification in the State of São Paulo. Official statistics from IBGE (Instituto Brasileiro de Geografia e Estatística) were used as reference of soybean planted area. The method proposed by this study is based on the temporal variation of the NDVI (Normalized Difference Vegetation Index) in agricultural areas. Time series of Moderate Resolution Imaging Spectroradiometer (MODIS) imagery were acquired between January 2003 and May 2008. The HANTS (Harmonic Analysis of NDVI Time-Series) algorithm was used to filter the time series through low frequency sine functions to represent the seasonal effect of the vegetation development. The soybean mask was built according to the seasonal variation of the soybean pixels. Results have showed that, although the method has overestimated the soybean planted areas, it can help government agencies in generating agricultural statistics. At the municipal scale, the coefficient of determination (R^2) was always greater than 0,7.

Palavras-chave: remote sensing, area estimates, NDVI, sensoriamento remoto, estimativa de área, NDVI.

1. Introdução

Novas tecnologias vêm sendo desenvolvidas para auxiliar no monitoramento de culturas agrícolas utilizando dados de sensoriamento remoto. Atualmente as estimativas de safra no Brasil, em escala municipal, são realizadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE), as quais são baseadas em informações de agentes técnicos e econômicos relacionados ao setor agrícola (Sano et al., 1998), e, portanto são subjetivas e não possuem um erro associado. Logo, torna-se extremamente necessário criar novas metodologias para auxiliar os produtores e o governo brasileiro na estimativa das áreas plantadas e no monitoramento da cultura de forma precisa e objetiva.

A grande vantagem das imagens de satélite é a rapidez na coleta dos dados, o baixo custo e a objetividade dos dados. Trabalhos visando estimar área plantada utilizando imagens de satélite tem sido objeto de estudo de diversos autores (Alonso et al. (1991); Medeiros et al. (1996); Ippoliti-Ramilo et al. (2003); Motta et al. (2001); Fontana et al. (2001); Luiz (2003); Adami et al. (2005); Rizzi et al. (2006); Wardlow e Egbert (2008)), porém a aplicabilidade dessas metodologias ainda é limitada, devido ao volume de processamento, ao custo e à periodicidade de aquisição das imagens.

Para as culturas de verão, como é o caso da soja e do milho, é ainda mais complexa a aquisição de imagens, pois o ciclo dessas culturas vai de novembro a março, época em que a pluviosidade é maior e, portanto quase sempre há ocorrência de nuvens, e isso impossibilita a

obtenção de imagens com informação sobre as culturas.

Neste sentido, as imagens do sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS) a bordo dos satélites Terra e Aqua (NASA/EOS), favorecem a aquisição de imagens livres de cobertura de nuvens e viabilizam o monitoramento da cultura ao longo de seu ciclo vegetativo e produtivo devido à sua resolução temporal. Dentre os produtos gerados a partir dos dados coletados por este sensor, o índice de vegetação *Normalized Difference Vegetation Index* (NDVI) tem apresentado uma boa calibração e é apto para aplicações de classificação e detecção de mudanças (Miura et al., 2000). As imagens do NDVI são fornecidas a cada 16 dias, apresentam georreferenciamento, correção atmosférica e podem ser adquiridas em resolução espacial de 250 x 250 m.

Desta forma, o presente trabalho teve como objetivo estimar as áreas plantadas com soja no estado de São Paulo nas safras 2003/04, 2004/05, 2005/06 e 2006/07, por meio de técnicas automáticas de confecção de máscaras de área plantada utilizando dados de sensoriamento remoto de baixo custo. Essas estimativas baseadas em imagens devem auxiliar o governo e os produtores em suas estimativas bem como no monitoramento da expansão de área da cultura.

2. Metodologia de Trabalho

A área de estudo compreende o estado de São Paulo, localizado na região sudeste do Brasil entre as coordenadas geográficas de 44°W a 54°W e 19°S a 26°S.

Material

Para o desenvolvimento deste trabalho foram utilizadas as imagens do sensor MODIS, a bordo da plataforma Terra, com composição de 16 dias (produto MOD13, *Normalized Difference Vegetation Index* (NDVI), *tile* H13V11) adquiridas desde o dia 1 de janeiro de 2003 até o dia 25 de maio de 2007. Os dados do NDVI das imagens MODIS são gravados em números inteiros multiplicados por 10.000.

Como referência, foram utilizados dados do Instituto Brasileiro de Geografia e Estatística (IBGE) referente ao tamanho de área plantada com soja nos municípios do estado de São Paulo.

Tratamento das imagens

As imagens do sensor MODIS foram convertidas de 16 para 8 bits, e do formato HDF (*Hierarchical Data Format*) para o formato Geo-TIFF (*Georeferenced Tag Image File Format*) no software MODIS Reprojection Tool 4.0, desenvolvido pela USGS *Center for Earth Resources Observation and Science* (EROS).

Filtragem da série de imagens

Após uma análise preliminar na série temporal de NDVI, foi verificado que havia a necessidade de suavizar os dados e eliminar os ruídos. Para isso foi utilizado o algoritmo de análise harmônica HANTS (*Harmonic Analysis of NDVI Time-Series*), desenvolvido por Roerink et al. (2000). O conceito básico deste algoritmo é que o NDVI da vegetação apresenta um forte efeito sazonal que pode ser simulado por funções senoidais de baixa frequência, com diferentes fases e amplitudes. Nuvens e outros efeitos perturbadores, esporádicos ou não, provocam alterações nas séries temporais do NDVI e podem ser considerados ruídos de alta frequência. O algoritmo HANTS, aplicado por meio de rotinas IDL (*Interactive Data Language*) analisa a série temporal original do NDVI e identifica as áreas de baixas frequências. Então, a partir de transformações de Fourier, o algoritmo elimina valores discrepantes da série temporal, que representam altas frequências. O resultado é a filtragem da série, que se torna mais estável e sem a influência de ruídos ou variações do NDVI não causadas por efeitos naturais. Este algoritmo foi utilizado com sucesso na filtragem de séries temporais em recentes estudos (Jiang et al., 2008; Julien et al., 2006). A Figura 1 ilustra o

exemplo de uma série de dados NDVI do sensor MODIS, do dia juliano 353 de 2006 ao dia 145 de 2008, com e sem a aplicação do algoritmo HANTS em um pixel de soja.

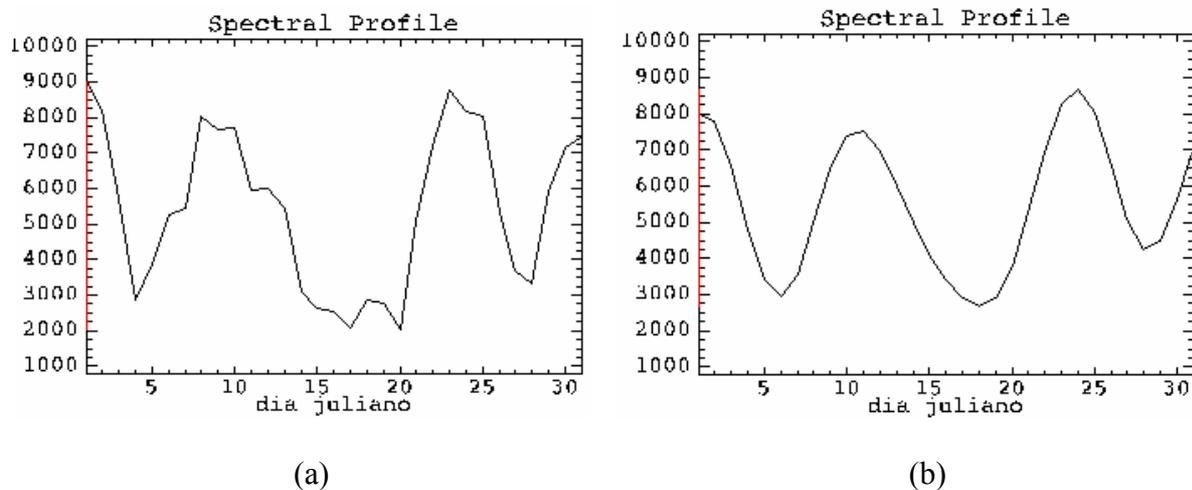


Figura 1. Valor do NDVI para uma série de imagens do sensor MODIS original (a) e a mesma série de imagens com aplicação do algoritmo HANTS, em um pixel de soja.

Extração dos dados e geração da máscara

A metodologia empregada na geração da máscara de soja foi baseada na variação sazonal e na grande amplitude de valores do NDVI que ocorreram nas áreas agrícolas ao longo do ciclo produtivo. A variação sazonal da biomassa nas áreas agrícolas influencia os índices de vegetação, cujo comportamento se assemelha a uma função senoidal, com altos valores nas épocas de pico vegetativo e baixos valores na época de colheita. Assim, a metodologia proposta considerou como pixels agrícolas aqueles cuja diferença entre os valores máximo e mínimo do NDVI ocorridos em determinadas épocas do ciclo esteve dentro de uma faixa pré-estabelecida. No caso da soja, a metodologia separou pixels que apresentaram altos valores de NDVI em janeiro (época do pico vegetativo) e baixos valores em setembro a outubro (preparo do solo e plantio), e cuja diferença esteve em um delta (diferença entre o maior valor de NDVI e o menor valor NDVI observado) pré-determinado, conforme é mostrado na Tabela 1. As imagens e os valores de delta foram selecionados após uma análise feita nos pixels puros de soja dos municípios cuja área plantada com esta cultura representava 60% ou mais da área total do município. Foram calculadas as médias dos valores de NDVI dos pixels com cana-de-açúcar e o desvio padrão, para selecionar os valores de delta.

Tabela 1. Data das imagens (ano e dia juliano) de mínimo e máximo, e, valores de delta mínimo e máximo selecionadas para o cálculo da área plantada com soja.

Safr	Data das imagens (ano e dia juliano)		Valor de delta	
	Máximo	Mínimo	mínimo	máximo
2003/04	2004001 e 2004017	2003257 a 2003289	3000	9000
2004/05	2005001 e 2005017	2004273 e 2004289	3000	9000
2005/06	2006001 e 2006017	2005257 a 2005289	3000	9000
2006/07	2007001 a 2007017	2006257 a 2006289	2500	9000

A metodologia foi aplicada por meio de uma rotina escrita em linguagem IDL. A rotina fez a leitura da série temporal de imagens de NDVI e, em seguida, gerou composições máximas e mínimas, contendo os maiores e menores valores de NDVI ocorridos em

determinados momentos ao longo da série temporal. Uma imagem diferença entre a composição máxima e mínima foi calculada, e foram separados apenas os pixels que apresentaram os valores dentro das faixas descritas na Tabela 1.

3. Resultados e Discussão

O resultado da classificação da máscara para áreas plantadas com soja é apresentado na Figura 2. Pode ser observado, nos mapas temáticos que a cultura se expandiu na safra 2004/05, porém entre as safras de 2005/06 e 2006/07 houve uma diminuição gradativa da área plantada com soja.



Figura 2. Áreas plantadas com soja selecionadas pela máscara, nas safras 2003/04 a 2006/07.

A Tabela 2 apresenta as estimativas de área plantada com soja selecionada pela máscara, para o estado de São Paulo, nas safras 2003/04, 2004/05, 2005/06 e 2006/07, e os dados oficiais do Instituto Brasileiro de Geografia e Estatística (IBGE).

Tabela 2. Área plantada com soja estimada pela máscara (ha), variância (ha), coeficiente de variação, área plantada com soja informada pelo IBGE (ha) e diferença (%) entre a área estimada pela máscara e a área informada pelo IBGE.

Safra	Área estimada pela máscara (a)	Variância	Coeficiente de variação ¹	Área IBGE (b)	Diferença 100((b-a)/a)
2003/04	878312	14025972	2,94	779880	-11,20
2004/05	1084400	16757916	2,60	781210	-27,95
2005/06	799156	9324298	2,63	656600	-17,83
2006/07	650800	4483539	2,24	525300	-19,28

Para as safras 2003/04 a 2006/07 a diferença entre o dado estimado (máscara) e o dado real (IBGE) mostrou-se razoável, devido ao fato da máscara ter classificado algumas áreas como sendo soja erroneamente. Essa falha pode ter ocorrido em áreas de plantio de milho, pois observou-se que a diferença entre os valores de máximo e mínimo de NDVI (delta) de áreas plantadas nessas áreas geralmente atingiam o mesmo valor observado para a soja, fazendo com que a máscara selecionasse tais áreas como sendo soja.

Quando o resultado foi analisado em escala municipal e comparado com os dados do IBGE, a área plantada com soja nos municípios atingiu um coeficiente de determinação (R^2) de 0,83 para a safra 2004, de 0,80 para a safra 2005, de 0,84 para a safra 2006 e de 0,70, ou seja, a área estimada pela máscara explicou 83%, 80%, 84% e 70% da variância das áreas plantadas com soja nos municípios do estado de São Paulo nas safras 2004, 2005, 2006 e 2007 respectivamente.

Os valores do coeficiente de variação (CV) mostraram que há grande dispersão nos dados, devido ao fato da análise ter sido feita para todos os municípios do estado de São Paulo, sendo que o plantio de soja se concentra em apenas algumas áreas. Portanto, o CV mostrou o quanto as áreas de soja são heterogêneas neste estado.

4. Conclusões

O método de identificação da cultura da soja utilizando a diferença entre períodos máximos e mínimos do NDVI da cultura se mostrou útil para auxiliar as estimativas de áreas plantadas com soja, podendo inclusive complementar e apoiar informações de órgão oficiais, como é o IBGE, e talvez aprimorá-las, já que este método é objetivo.

Devido ao baixo custo dessa metodologia é possível calcular a máscara de soja após o período de vigor vegetativo da cultura para que se possa monitorar também a colheita desta cultura.

Visto que a resolução espacial do sensor MODIS é baixa, é possível melhorar as estimativas utilizando imagens de média resolução.

Agradecimentos

À CONAB - Companhia Nacional de Abastecimento, através do Programa das Nações Unidas para o Desenvolvimento (PNUD), pelo financiamento do projeto, do qual se originou este trabalho, e ao engenheiro agrícola Jeferson Lobato Fernandes pelas sugestões que contribuíram para a execução do trabalho.

¹ Coeficiente de variação de Pearson (CV) é uma medida de dispersão referente à divisão entre o desvio padrão e a média. Não possui unidade de medida, porém muitas vezes é multiplicado por 100 e seu valor é então expresso em porcentagem.

Referências Bibliográficas

- 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**, n. 57, v. 1, p. 22 – 27, 2005.
- Alonso, F. G.; Soria, S. L.; Gozalo, J. M. C. Comparing two methodologies for crop area estimation in Spain using Landsat TM images and ground-gathered data. **Remote Sensing of Environment**, v. 35, n. 1, p. 29-35, 1991.
- 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**. Porto Alegre: Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, CEPARM/UFRGS. 2001, 116 p. (Série D: Relatório Técnico - N° 005/01).
- 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.
- Jiang, X.; Wang, D.; Tang, L.; Hu, J.; Xi, X. Analysing the vegetation cover variation of China from AVHRR-NDVI data. **International Journal of Remote Sensing**, v.29, n.17, p.5301-5311, 2008.
- Julien, Y.; Sobrino, J.A.; Verhoef, W. Changes in land surface temperatures and NDVI values over Europe between 1982 and 1999. **Remote Sensing of Environment**, v.103, n.1, p.43-55, 2006.
- Luiz, A. J. B. **Estatísticas agrícolas por amostragem auxiliadas pelo sensoriamento remoto**. 2003. 116 p. Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2003.
- Medeiros, A. M. P.; Rudorff, B. F. T.; Shimabukuro, Y. E. Imagens Landsat na estimativa de áreas de cana-de-açúcar, milho e soja. In: Simpósio Brasileiro de Sensoriamento Remoto, 8, Salvador, 1996. **Anais...** São José dos Campos: INPE, 1996. CD-ROM.
- Miura, T.; Didan, K.; Huete, A. R.; Rodriguez, P. E. A Performance Evaluation of the MODIS Vegetation Index Compositing Algorithm. **IEEE Transactions on Geoscience and Remote Sensing**, v. 38, n. 3, 2000.
- 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. 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, n. 1, p. 30-35, 2006.
- Roerink, G. J.; Menenti, M.; Verhoef, W. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. **International Journal of Remote Sensing**, v.21, n.9, p.1911-1917, 2000.
- 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.
- Wardlow, B. D.; Egbert, S. L. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. **Remote Sensing of Environment**, v. 112, n. 3, p. 1096-1116, 2008.