

Discordância global e análise de índices de acurácia para avaliação de mapeamentos de área de cana-de-açúcar no Paraná

Clóvis Cechim Júnior¹
Jerry Adriani Johann¹
João Francisco Gonçalves Antunes²

¹ Universidade Estadual do Oeste do Paraná - UNIOESTE
R. Universitária, 2069 - Caixa Postal 711
85819-110 - Cascavel - PR, Brasil
{juniorcechim, jerry.johann}@hotmail.com

² Embrapa Informática Agropecuária - CNPTIA
Av. André Tosello, 209 - Caixa Postal 6041
13083-886 - Campinas - SP, Brasil
joao.antunes@embrapa.br

Abstract. The use of images orbital remote sensing in mappings of agricultural crops has been very used in virtue of its availability free, the which allows the development of methodologies low cost and efficient for agricultural monitoring, being for area estimation of direct form in a short and medium term or to verify the expansion of cultivated areas. The objective of this paper was to evaluate the accuracy of maps from the analysis of global disagreement formed by the components of quantity and allocation, for this it was used the supervised classification model Fuzzy ARTMAP in temporal series of EVI images of the orbital sensor MODIS of the TERRA and AQUA satellites for the mapping of sugarcane in the 2014/2015 year-harvest in Paraná State, South of Brazil. The accuracy assessment of the classification for the 2014/2015 crop year reached global accuracy of 97.69%, with a Kappa index of 0.84. The overall disagreement reached 0.59 in the component allocation and 1.97 in the quantitative component. The accuracy indexes used indicate the efficiency of the mapping generated by the Fuzzy ARTMAP classification in state.

Key-words: remote sensing, image processing, space accuracy, sensoriamento remoto, processamento de imagens, acurácia espacial.

1. Introdução

Dentre os maiores produtores de cana-de-açúcar, segundo dados da FAOSTAT (2014), destacam-se o Brasil com a média histórica entre 1961 a 2013, com 271,4 milhões de toneladas, seguido pela Índia com cerca de 207,5 milhões de toneladas. O Brasil, além de ser maior produtor de cana-de-açúcar do mundo também, é o primeiro produtor mundial de açúcar, responsável por 20% da produção mundial e 40% das exportações mundiais. É também o 2º produtor mundial de etanol, responsável por 20% da produção mundial e 20% das exportações mundiais (UNICA, 2015).

De acordo com dados do IBGE (2016), o estado do Paraná produziu em média, no período de cinco anos compreendidos entre (2011 a 2015), 47,3 milhões de toneladas de cana-de-açúcar ao ano. O estado do Paraná representa 6,5% da área cultivada nacional e é o 4º maior produtor.

O mapeamento e monitoramento de áreas cultivadas são essenciais para a elaboração de políticas públicas para o desenvolvimento agrícola (Johann et al., 2012; Lunetta et al., 2006). E a partir dos produtos disponibilizados por alguns sensores orbitais, como os índices de vegetação, têm sido muito utilizados em estimativas de área cultivada e da produtividade das culturas agrícolas. Isso demonstra a importância do sensoriamento remoto, seja para a classificação automatizada de culturas agrícolas como no monitoramento da cultura de cana-de-açúcar que foi feito no Centro-Sul do Brasil pelo projeto Canasat (Rudorff et al., 2010).

Encaixam-se nessa conjectura os dados do sensor MODIS disponibilizados gratuitamente no banco de produtos MODIS na Base Estadual Brasileira (Esquerdo et al., 2010) usados para gerar mapeamentos e para estimar a área cultivada de diversas culturas como a cana-de-açúcar de forma rápida e objetiva sendo de extrema importância para todo o setor agrícola a nível nacional (Rudorff et al., 2010; Adami et al., 2012; Cechim Júnior et al., 2015; Duft et al., 2015; Antunes et al., 2015).

Porém para validar as metodologias geradas se faz uso de métricas estatísticas que tem a finalidade de avaliar o desempenho dos classificadores e a acurácia dos mapeamentos por meio de sua exatidão e precisão, as métricas mais usadas para comparação de mapas usada para esta avaliação são o índice kappa e a exatidão global juntamente com seus erros de omissão e inclusão.

Neste trabalho, será apresentada uma avaliação usando uma medida auxiliar conforme metodologia proposta por Pontius e Millones (2011), a partir do uso da discordância global que é formada pelas componentes quantidade e alocação, sendo utilizado o mapeamento de cana-de-açúcar do ano-safra de 2014/2015, feito por meio do modelo de classificação supervisionado *Fuzzy* ARTMAP usado por Antunes et al. (2015) no estado de São Paulo e Cechim Júnior et al. (2015) no estado do Paraná, comparado com o mapeamento feito pelo projeto Canasat usado como referência para a validação.

2. Material e Métodos

O resultado gerado na serie temporal de imagens de EVI do sensor MODIS (Huete et al., 1997; Huete et al., 1999; Huete et al., 2002) partir das classificações *Fuzzy* ARTMAP foram estatisticamente analisados para avaliar a acurácia e determinar o nível dos erros entre as estimativas de área cultivada. A acurácia do mapeamento gerado pela classificação *Fuzzy* ARTMAP para o ano-safra 2014/2015 foi avaliada com base em Pontius e Millones (2011). A matriz de erros foi transformada em matriz de proporções, que apresenta estimativas imparciais e não tendenciosas para toda a população, ou para toda a área em estudo.

A exatidão global mede a concordância das proporções classificadas corretamente na diagonal principal da matriz. O índice Kappa mede a concordância e faz a comparação de um mapa gerado por uma classificação com um mapa de referência que pode ter dois tipos de aleatoriedade. As distribuições aleatória de quantidade e de alocação foram usadas para avaliar a precisão do mapeamento. A discordância global é calculada pela matriz que separa em duas componentes, e estas podem ter diferentes interpretações.

A metodologia de Pontius e Millones (2011) deriva de uma única matriz de confusão que, como variáveis de entrada, devem ser inseridas os conjuntos amostrais das classes em estudos que são comparados com todo o conjunto populacional de pixels das classes presentes no mapeamento gerado.

As medidas estatísticas utilizadas foram a exatidão global, o índice Kappa e a discordância global composta pela componente de quantidade, que é a classificação de proporções incorretas de pixels nas classes e pela componente de alocação, que representa a distribuição espacial incorreta de pixels nas classes.

A acurácia das classificações geradas pela classificação *Fuzzy* ARTMAP para o ano-safra 2014/2015 foi avaliada a partir de amostras adquiridas por amostragem aleatória estratificada proporcional nas classes de cana-de-açúcar e não cana-de-açúcar, usando como referência os mapas temáticos do Canasat, dando origem à matriz de confusão (Tabela 1). Várias medidas de precisão da classificação podem ser derivadas a partir da matriz de confusão, porém, a exatidão global é uma das mais conhecidas (Foody, 2002).

A matriz de confusão é calculada pela comparação entre a localização e a classe de cada pixel da referência com o local correspondente da classe da imagem gerada pela classificação.

Cada coluna da matriz de confusão representa uma classe de referência, e os valores contidos na coluna correspondem ao número de pixels em cada classe da classificação que está ocorrendo quando se compara cada classe da referência.

Tabela 1 Matriz de confusão

		Referência		Soma	Total
		Cana j = 1	Não cana j = 2		
Classificação	Cana	i = 1	n ₁₁ (VP)	Σn _{1j}	N ₁
	Não Cana	i = 2	n ₂₂ (FN)	Σn _{2j}	N ₂

em que:

n_{ij} = amostras da classe de referência j ocorrendo na classe i; Σn_{ij} = soma das amostras da classe de referência j ocorrendo na classe i; N_i = total de amostras da classe i; VP = verdadeiro positivo; FP = falso negativo; VN = verdadeiro negativo.

A matriz de confusão pode mostrar a distribuição de classe em porcentagem para cada classe de referência. Os valores são obtidos ao se dividirem as contagens de pixels em cada coluna de referência pelo número total de pixels em determinada classe de referência.

Os erros de comissão representam os pixels que pertencem à outra classe, os quais são marcados como pertencentes à classe de interesse. Os erros de comissão são mostrados nas linhas da matriz de confusão.

Os erros de omissão representam os pixels que pertencem à classe de referência, mas a técnica de classificação não conseguiu classificá-los para a classe correta. Os erros de omissão são mostrados nas colunas da matriz de confusão. Por meio da matriz de confusão se obtém a matriz de proporções que evidencia a proporção das amostras de pixels, estimadas nas classes de referência (Tabela 2).

Tabela 2 Matriz de proporções

		Referência		Total
		Cana j = 1	Não cana j = 2	
Classificação	Cana	i = 1	p ₁₁	Σp _{1j}
	Não cana	i = 2	p ₂₁	Σp _{2j}
	Total		Σp _{i1}	Σn _{j2}

em que:

p_{ij} = proporções de amostras da classe de referência j representada na classe i;
 Σp_{ij} = soma da proporção de amostras da classe de referência j ocorrendo na classe i; onde p_{ij} é calculado pela Equação 1.

$$p_{ij} = \frac{(n_{ij})}{\left(\sum_{j=1}^J n_{ij}\right)} * \frac{(N_i)}{\left(\sum_{j=1}^J N_i\right)} \quad (1)$$

A análise dos resultados do mapeamento consistiu na avaliação da acurácia com os Índices Kappa e Exatidão Global (EG) (Congalton, 1991; Congalton e Green, 1999), (Equações 2 e 3), e na discordância global das classes cultivadas com e sem cana-de-açúcar.

$$IK = \frac{[n * \sum_{i=1}^r x_{ij} - \sum_{i=0}^r (x_i * x_j)]}{[m^2 - \sum_{i=1}^r (x_i * x_j)]} \quad (2)$$

$$EG = \frac{A}{m} * 100 \quad (3)$$

em que:

n = número de observações (pixels amostrais); A = acerto geral (pixels amostrais classificados corretamente); m = número de pixels amostrais; r = número de linhas da matriz de erro; x_{ij} = observações na linha i e coluna j; x_i = total marginal da linha i; x_j = total marginal da coluna j.

3. Resultados e Discussão

O resultado da classificação supervisionada por *Fuzzy ARTMAP* para o ano-safra 2014/2015 foi usado para avaliar os índices de acurácia (Figura 1) o que consistiu na comparação dos pixels de referência, adquiridos a partir do mapa proveniente Canasat. Foram usados 10.531 pixels para gerar a matriz de confusão (Tabela 3) e a matriz de proporções (Tabela 4). Desses, 907 pixels da classe de cana e 9.624 pixels da classe correspondente a não cana, a partir da matriz erro foram calculadas as demais medidas estatísticas.

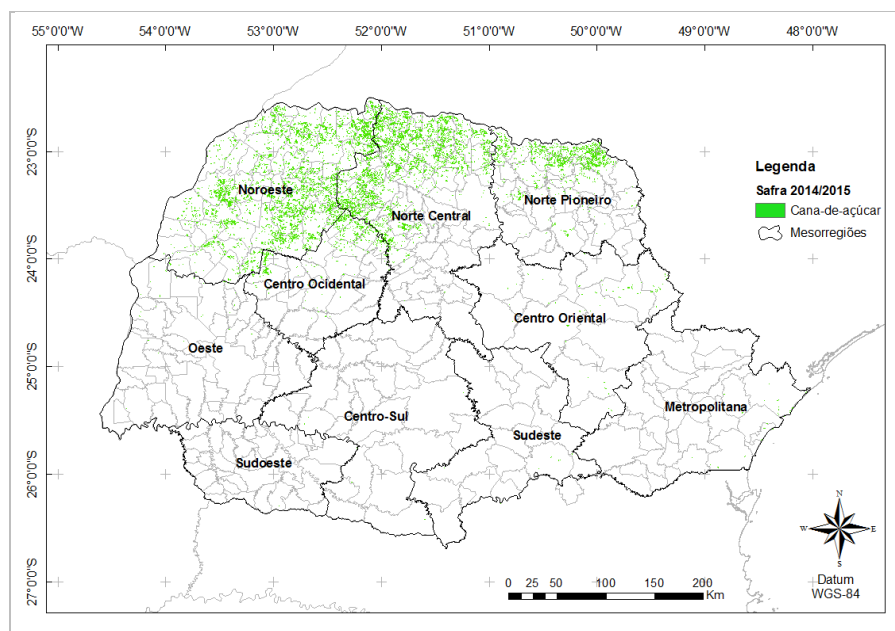


Figura 1. Modelo de classificação *Fuzzy ARTMAP* ano-safra 2014/2015.

A avaliação da acurácia da classificação para o ano-safra 2014/2015 atingiu a exatidão global de 97,69%, com um índice Kappa de 0,84. Conforme análise das medidas estatísticas obtidas, pode-se inferir que o mapeamento obtido pode ser considerado de qualidade temática muito boa, com alto valor de exatidão. Tais dados corroboram com o mapeamento temático

do projeto Canasat da safra 2010/2011 na região Centro-Sul com exatidão de 98% e com um erro de subestimativa de área plantada de cana-de-açúcar de 0,5%. A validação foi realizada por uma campanha de campo nos estados de São Paulo, Minas Gerais e Paraná, mostrando a excelente qualidade temática do mapeamento realizado pelo projeto. Sendo assim, também são esperados resultados com elevado nível de acurácia para as demais safras, devido ao método consistente de interpretação visual (Adami, et al., 2012).

Tabela 3 Matriz de confusão para determinação de índices de acurácia

Matriz de confusão					
Ano-safra 2014/2015		Referência		Total	
		Cana	Não cana	Amostrai	Populacional
	Cana	678	65	739	117.721
Classificação	Não cana	229	9.555	9.784	3.409.722
	Total	907	9.624	10.531	3.527.443

A partir da matriz de confusão, foi obtida a matriz de proporções, calculada por meio de planilhas eletrônicas do Excel, desenvolvidas por Pontius (2014). A matriz evidencia a proporção das amostras de pixels estimadas nas classes de referência (Tabela 4).

A Tabela 4 expressa a proporção estimada da área de estudo que é a classe (i) no mapa de comparação e a classe (j) no mapa de referência. Desta forma, a tabela apresenta as estimativas imparciais sobre as proporções para a área total em estudo. De acordo com a tabela, é possível o cálculo das métricas estatísticas, como os erros de omissão e comissão e a discordância de quantidade e discordância de alocação.

Tabela 4 Matriz de proporções

Matriz de confusão				
Ano-safra 2014/2015		Referência		Total
		Cana	Não cana	
	Cana	0,03	0,003	0,03
Classificação	Não cana	0,02	0,95	0,97
	Total	0,05	0,95	1,00

As medidas dos parâmetros quantidade e alocação de discordância resumem uma matriz cruzada. As demais medidas estatísticas, para avaliar a acurácia dos mapeamentos, foram calculadas de acordo com Pontius e Millones (2011).

Em função do índice Kappa calcular a acurácia fundamentando-se na aleatoriedade surge como opção a análise da Discordância Global, que é composta pela componente Quantidade, considerada como a classificação de proporções incorretas de pixels nas classes e pela componente Alocação que se refere à distribuição espacial incorreta de pixels nas classes. Tais componentes fornecem informações complementares e adicionais que auxiliam na explicação do erro nos mapeamentos. A Figura 2 evidencia o gráfico com os erros de omissão e comissão, para as classes avaliadas de cana-de-açúcar e não cana-de-açúcar, em relação à porcentagem da área em estudo.

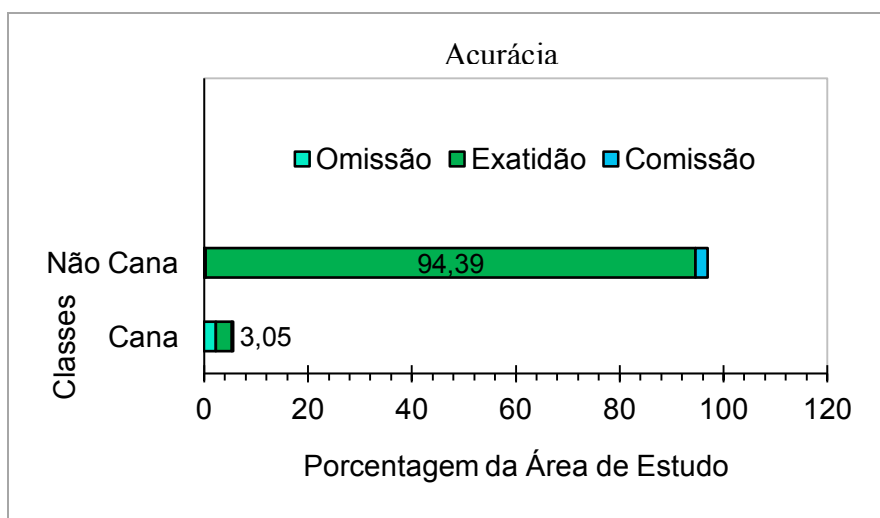


Figura 2. Erros de omissão e comissão para as classes de cana e não cana em relação à porcentagem da área de estudo.

Apesar das amostras de treinamento serem adquiridas de forma homogênea, esse nível mais elevado do erro de omissão da classe cana-de-açúcar (Figura 2) foi o mesmo observado por Antunes et al. (2015) e está associado à variabilidade espectral inerente da cultura ao longo do seu ciclo de desenvolvimento vegetativo, o que ocasiona maior confusão espectral no reconhecimento do padrão espectro-temporal da cultura pelo classificador.

As intensidades dos erros de omissão e comissão, para as classes de cana e não cana é dada em relação à porcentagem de cada uma das classes (Figura 3).

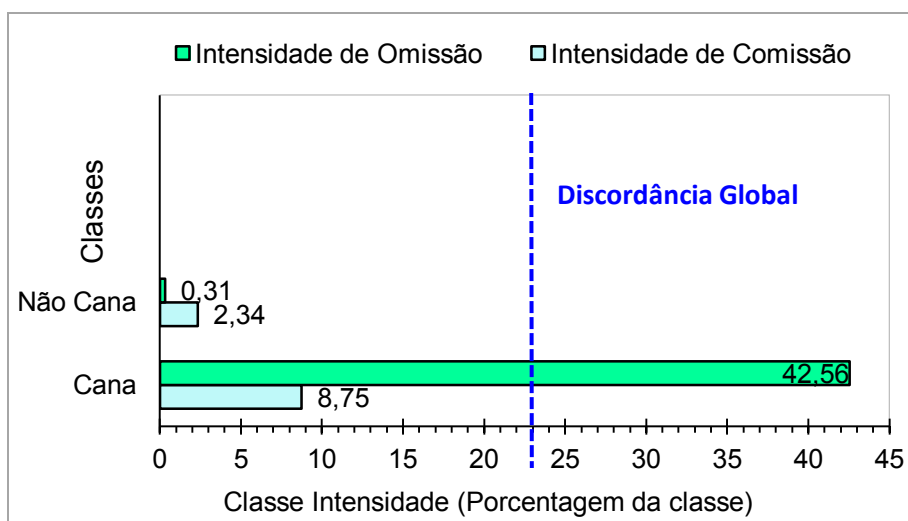


Figura 3. Intensidade dos erros de omissão e comissão para as classes de cana e não cana.

Duft et al. (2015) compararam a área estudada da bacia hidrográfica do Rio Paranaíba, os dados obtidos pelos mapeamentos da cultura de cana-de-açúcar gerados com séries de imagens temporais de EVI e NDVI do MODIS com os dados do Canasat e verificaram, para o ano-safra de 2009/2010, que houve uma subestimação de 6,2% para o mapeamento feito por EVI. Este resultado também foi encontrado no Paraná em regiões mais ao Sul do estado onde os cultivos de cana-de-açúcar não são muito significativos e em virtude da resolução do MODIS (250m) fez com que pixels de bordas de talhões não fossem selecionados pela máscara.

No gráfico da discordância global do ano-safra 2014/2015, formado pelas porcentagens da componente de alocação e quantidade (Figura 4), se observa que a contribuição da componente de quantidade na discordância total foi maior que a componente de alocação, o que implica maiores proporções incorretas de pixels designados nas classes.

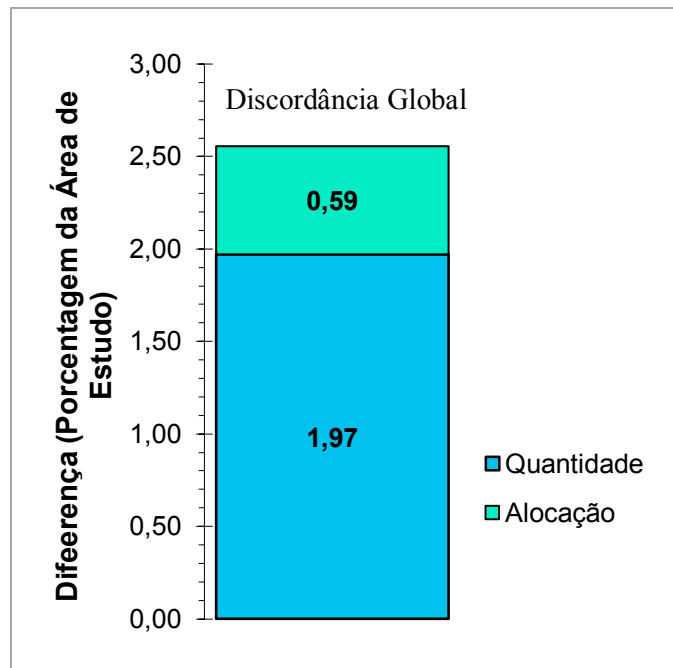


Figura 4. Discordância global formada pelas componentes quantidade e alocação.

4. Conclusões

O erro de omissão da classe cana-de-açúcar pode indicar uma tendência à subestimação de áreas de cultivo da cultura provavelmente em algumas regiões mais ao sul do estado onde a concentração de cana-de-açúcar não é muito significativa.

A discordância global formada pelas componentes quantidade e alocação demonstrou-se uma medida estatística auxiliar que pode ser usada juntamente com as medidas estatísticas convencionais como a exatidão global e o índice Kappa, na avaliação de acurácia de mapeamentos, o que permite explicar os erros gerados entre o mapa de referência e o mapeamento gerado independentemente do tipo de classificação.

5. Agradecimentos

À coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Capes e Fundação Araucária pelo apoio financeiro, ao Programa de Pós-graduação Stricto Sensu em Engenharia Agrícola (PGEAGRI) pela oportunidade bem como ao Laboratório de Estatística Aplicada (LEA) da UNIOESTE/Campus Cascavel, pela infraestrutura disponibilizada para realizar este trabalho.

6. Referências

Antunes, J. F. G; Lamparelli, R. A. C; Rodrigues, L. H. A. Mapeamento do cultivo da cana-de-açúcar por meio da classificação de séries temporais de dados MODIS. In: Simpósio Brasileiro de Sensoriamento Remoto (SSBR), 2015, João Pessoa. **Anais...** São José dos Campos: INPE, 2015.

Adami, M.; Mello, M. P.; Aguiar, D. A.; Rudorff, B. F. T.; Souza, A. F. A web platform development to perform thematic accuracy assessment of sugarcane mapping in South-Central Brazil. **Remote Sensing**. 2012, 4, 3201-3214.

Cechim Júnior, C.; Johann, J. A.; Antunes, J. F. G. Mapeamento de cana-de-açúcar por meio da classificação de séries temporais de EVI do sensor MODIS no estado do Paraná. In: VI – Simpósio em Estatística Espacial e Modelagem de Imagens, 2015, Toledo. **Anais...** Toledo: SEEMI, 2015.

Congalton, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v.37, p.35-46, 1991.

Congalton, R. G.; Green, K. **Assessing the accuracy of remotely sensed data: principles and practices**. Boca Raton: CRC Press, 1999. 160p.

Duft, D. G.; Picoli, M. C. M.; Scarpare, F. V.; Hernades, T. A. D.; Galdos, M. V. Comparação do desempenho de índices de vegetação do sensor MODIS para mapeamento sistemático da cana-de-açúcar. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 2015, João Pessoa. **Anais...** São José dos Campos: INPE, 2015.

Esquerdo, J. C. D. M.; Antunes, J. F. G.; Andrade, J. C. **Desenvolvimento do banco de produtos MODIS na Base Estadual Brasileira**. (Comunicado Técnico, 100) - Campinas: Embrapa Informática Agropecuária, 7 p., 2010.

FAOSTAT - Food and Agriculture Organization of the United Nations Statistics Division. **Production**, 2014. Disponível em: <<http://faostat3.fao.org/download/Q/QC/E>>. Acesso em: 14 mar. 2014.

Foody, G. M. Status of land cover classification accuracy assesment. **Remote Sensing of Environment**, v. 80, n. 1, p. 185-201, 2002.

Huete, A.; Didan, K.; Miura, T.; Rodriguez, E. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v. 83, n.1-2, p. 195-213, 2002.

Huete, A.; Justice, C.; Leeuwen, W. V. **MODIS vegetation index (MOD 13) algorithm theoretical basis document**. 1999. Disponível em: <http://modis.gsfc.nasa.gov/data/atbd_mod13.pdf>. Acesso em: 24 mar. 2016.

Huete, A.; Liu, H. Q.; Batchily, K.; Leween, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. **Remote Sensing of Environment**, n. 59, p. 440-451, 1997.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Banco de Dados Agregados: Sistema IBGE de Recuperação Automática – SIDRA**. 2016. Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em: 24 fev. 2016.

Johann, J. A.; Rocha, J. V.; Duft, D. G.; Lamparelli, R. A. C. Estimativa de áreas com culturas de verão no Paraná, por meio de imagens multitemporais EVI/Modis. **Pesquisa Agropecuaria Brasileira**, n. 9, v. 47, p. 1295-1306, 2012.

Lunetta, R. L.; Knight, F. K.; Ediriwickrema, J.; Lyon, J. G.; and Worthy, L. D. 2006. Landcover change detection using multi-temporal MODIS NDVI data. **Remote Sensing of Environment**, 105, 142-154.

Pontius, R. G.; Millones, M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. **International Journal of Remote Sensing**, v. 32, n. 15, p. 4407-4429, 2011.

Rudorff, B. F. T.; Aguiar, D. A.; Silva, W. F., Sugawara, L. M., Adami, M., Moreira, M. A. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data. **Remote Sensing**, v. 2, n. 4, p. 1057-1076, 2010.

UNICA - União da indústria de cana-de-açúcar. **Maior produtor mundial de cana-de-açúcar**. 2015. Disponível em: <<http://www.unica.com.br/faq/>>. Acesso em: 09 set. 2015.