

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/303519495>

Classificação orientada a objetos em imagens multitemporais Landsat aplicada na identificação de cana-de-açúcar e soja

Article · May 2016

CITATIONS

3

READS

228

6 authors, including:



Antonio Roberto Formaggio

National Institute for Space Research, Brazil

247 PUBLICATIONS 1,974 CITATIONS

[SEE PROFILE](#)



Isaque Daniel Rocha Eberhardt

Instituto de Pesquisa Econômica Aplicada - IPEA

32 PUBLICATIONS 32 CITATIONS

[SEE PROFILE](#)



Ieda Sanches

National Institute for Space Research, Brazil

57 PUBLICATIONS 296 CITATIONS

[SEE PROFILE](#)



Julio Cesar de Oliveira

Universidade Federal de Viçosa (UFV)

31 PUBLICATIONS 66 CITATIONS

[SEE PROFILE](#)

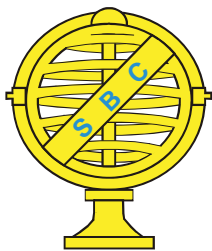
Some of the authors of this publication are also working on these related projects:



Integrated LOGistic Network Model (ILOGIM) [View project](#)



O uso da cobertura vegetal como bioindicador de impactos ambientais causados por vazamentos de hidrocarbonetos em dutos e sua detecção através de sensoriamento remoto. [View project](#)



Revista Brasileira de Cartografia (2016) N° 68/1: 131-143
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto
ISSN: 1808-0936

CLASSIFICAÇÃO ORIENTADA A OBJETOS EM IMAGENS MULTITEMPORAIS LANDSAT APLICADA NA IDENTIFICAÇÃO DE CANA-DE-AÇÚCAR E SOJA

*Object-Based Classification in Multi-Temporal Landsat Images Applied
to Identify Sugarcane and Soybean*

**Bruno Schultz¹, Antonio Roberto Formaggio¹,
Isaque Daniel Rocha Eberhardt², Ieda Del'Arco Sanches¹,
Júlio Cesar de Oliveira³ & Alfredo José Barreto Luiz⁴**

¹ Instituto Nacional de Pesquisas Espaciais – INPE
Divisão de Sensoriamento Remoto – DSR
Caixa Postal 515 – 12227-010 – São José dos Campos, SP – Brasil
{schultz, formagg, ieda,}@dsr.inpe.br

² Universidade de Brasília – UNB
Campus Universitário Darcy Ribeiro, Pós-graduação em Transportes
Anexo SG-12, 1° andar, CEP 70910-900, Brasília, DF
isaquedanielre@hotmail.com

³ Universidade Federal de Viçosa – UFV
Departamento de Engenharia Civil
CEP 36570-900, Viçosa, MG
oliveirajc@ufv.br

⁴ Empresa Brasileira de Pesquisa Agropecuária, Brasil
EMBRAPA Meio Ambiente
Rod SP 340, km 127,5 Cx Postal 69 13820-000 Jaguariúna, SP
alfredo.luiz@embrapa.br

Recebido em 19 de Novembro, 2015/Aceito em 06 de Fevereiro, 2015
Received on November 19, 2015/Accepted on February 06, 2015

RESUMO

A presente pesquisa teve por objetivo avaliar a potencialidade de dados multitemporais Landsat para classificação de cana-de-açúcar e de soja, conjuntamente, quando realizada via Análise de Imagens Orientada a Objetos (OBIA/Random Forest). Foi utilizado um segmentador multi-resolução (SM) para gerar os polígonos (objetos). Um conjunto de 500 segmentações foi criado pela variação dos parâmetros Fe (fator de escala), Fm (forma) e Cp (compacidade), e avaliado pelo Índice de Avaliação da Segmentação (IAVAS). Da segmentação que obteve menor IAVAS, foram extraídos os atributos espectrais das médias e desvios-padrão das bandas TM/Landsat-5 [setembro (S) e outubro (O) do ano 2000] e ETM+/Landsat-7 [fevereiro (F) e março (M) do ano 2001] dos objetos, e seus NDVIs. Estes atributos foram inseridos no algoritmo Random Forest (RF) e as exatidões das classificações foram testadas quanto ao uso dos seguintes conjuntos de datas: SOFM; SFM; OFM; SOF; FM; OF; SF; e F. O IAVAS definiu Fe (35), Fm (30) e Cp

(50) como melhores parâmetros de segmentação. As melhores exatidões de classificação Random Forest situaram-se em torno de 86%. Duas datas produziram melhor resultado que apenas uma, entretanto, o uso de mais de duas não produziu melhora significativa na exatidão final da classificação.

Palavras chaves: Segmentação, IAVAS, Mapeamento de Cana-de-Açúcar e Soja.

ABSTRACT

This research aimed to evaluate the potential of multi-temporal Landsat data for sugarcane and soybean classification, together, when used through Object-Based Image Analysis (OBIA/Random Forest). It was used a multi-resolution segmenting (SM) to generate the objects. A set of 500 segmentations was created varying the Fe parameters (scale factor), Fm (shape) and Cp (compactness), and evaluated by Index for the Evaluation of Segmentation (IAVAS). At the segmentation that obtained the lowest IAVAS, were extracted the spectral attributes of means and standard deviations of bands TM / Landsat-5 [September (S) and October (O) of the year 2000] and ETM + / Landsat-7 [February (F) and March (M) for the year 2001] of the objects, and their NDVIs. These attributes were inserted into the algorithm Random Forest (RF) and accuracies were tested using the following set of dates: (SOFM); (SFM and OFM); (FM, OF and SF); and (F). The IAVAS defines Fe (35), Fm (30) and Cp (50) as the best segmentation parameters for the study area. The best classification accuracies are situated around 86%. Two dates produced better results than just one, however the use of more than two failed to produce a significant improvement in the final classification accuracy.

Keywords: Segmentation, IAVAS, Sugarcane and Soybean Mapping.

1. INTRODUÇÃO

Em 2013, o Brasil produziu cerca de 768 milhões de toneladas de cana-de-açúcar e 81,7 milhões de toneladas de soja (FAOSTAT, 2015). Para esta produção ser possível, foram necessários 10,2 e 27,9 milhões de hectares de cana-de-açúcar e soja, respectivamente (IBGE, 2015). Ainda em 2013, 50 % do plantio nacional de cana-de-açúcar estava concentrado no estado de São Paulo (SP) (CANASAT, 2013). Esta cultura ocupava 20 % da área do estado, enquanto que a soja, cultivada no território nacional ocupava uma área equivalente à área de todo o estado de SP (IBGE, 2015).

Dada a importância da cana-de-açúcar e da soja no agronegócio para a economia do País, as informações de previsão de safras destas duas culturas devem ser fornecidas de forma confiável e em tempo hábil, visto que o governo nacional necessita destas informações para o controle das importações e exportações dos produtos agrícolas e o agricultor necessita tomar decisões acertadas sobre o plantio e a comercialização dos bens de produção (RIZZI & RUDORFF, 2005; ADAMI *et al.*, 2007).

No Brasil, o Instituto Nacional de Geografia e Estatística (IBGE) realiza anualmente o levantamento sistemático da produção agrícola por município (PAM). O método utilizado por este órgão é denominado subjetivo por se basear em consultas a especialistas e por

não apresentar o erro associado às estimativas geradas (RIZZI & RUDORFF, 2005; ADAMI *et al.*, 2007; LUIZ *et al.*, 2012).

As imagens obtidas por sensoriamento remoto (SR) são formas de captura indireta das informações espaciais da superfície terrestre (LUIZ *et al.*, 2012). Esse tipo de dado, associado ao uso de ferramentas de Sistemas de Informação Geográfica (SIG), favorece a obtenção de informações atualizadas sobre a dinâmica espaço-temporal de culturas agrícolas, como a soja (RIZZI & RUDORFF, 2005; LAMPARELLI *et al.*, 2008; SUGAWARA *et al.*, 2008; EBERHARDT *et al.*, 2015) e a cana-de-açúcar (RUDORFF & SUGAWARA, 2007; RUDORFF *et al.*, 2010; AGUIAR *et al.*, 2011), em grandes extensões territoriais.

Pesquisas relacionadas ao uso de dados de SR para estimativas de áreas de culturas agrícolas estão concentradas, principalmente, na classificação automática ou visual de dados de sensores ópticos Landsat-like (LL) (RIZZI & RUDORFF, 2005; SUGAWARA *et al.*, 2008; RUDORFF *et al.*, 2010; VIEIRA *et al.*, 2012), no uso de índices de vegetação derivados do sensor MODIS para confecção de mapas (EBERHARDT *et al.*, 2015) e na elaboração de painéis amostrais (ADAMI *et al.*, 2007; LUIZ *et al.*, 2012). A escolha dentre estas técnicas dependerá dos objetivos dos levantamentos e das condições da agricultura do local de estudo.

Atualmente, o principal entrave no uso de dados LL (resolução geométrica da ordem de 30m e repetitividade de aproximadamente 16 dias) para classificação automática em mapeamentos agrícolas está relacionado à não operacionalidade de um sistema rotineiro, que consiga avaliar a área agrícola macrorregional em curto período de tempo (RIZZI & RUDORFF, 2005; SUGAWARA *et al.*, 2008; EBERHARDT *et al.*, 2015) e, também, à relativamente baixa performance de exatidão produzida por algoritmos convencionais de classificação automática (WILKINSON, 2005). No entanto, este padrão de dados é válido, por possuir uma resolução espacial considerada apropriada para mapeamento de cana-de-açúcar e soja nos estados do Paraná e São Paulo (RUDOEFF & SUGAWARA, 2007; LAMPARELLI *et al.*, 2008; RUDORFF *et al.*, 2010).

Conforme Myint *et al.* (2009), Peña-Barragán *et al.* (2011), Vieira *et al.* (2012), Fernandes *et al.* (2012) e Garofalo *et al.* (2015), usando imagens LL, melhores resultados vêm sendo adquiridos na classificação automática de polígonos (objetos) do que o uso de classificação automática por pixels.

O uso de algoritmos de árvore de decisão (AD) J48 (VIEIRA *et al.*, 2012; RODRIGUES & KUX, 2015) e Random Forest (RF) (LONG *et al.*, 2013) para classificação de culturas em diferentes cenários agrícolas mundiais tem permitido a obtenção de exatidões globais (EG) superiores a 75 % (PEÑA-BARRAGÁN *et al.*, 2011; VIEIRA *et al.*, 2012; LONG *et al.*, 2013), em função do número de culturas classificadas.

O Índice de Avaliação da Segmentação (IAVAS), proposto por Oliveira *et al.* (2003), pode ser utilizado para selecionar os melhores parâmetros de uma segmentação necessária à classificação segundo o modelo Random Forest (RF) e OBIA, representando assim, uma medida objetiva da segmentação.

Neste contexto, no intuito de avaliar a potencialidade de dados multitemporais LL no mapeamento de cana-de-açúcar e de soja, conjuntamente, utilizando OBIA/RF e IAVAS, no presente trabalho foi feita a avaliação do desempenho de dados Landsat obtidos em diferentes datas dos ciclos fenológicos das duas citadas culturas, quando utilizados como fonte

de atributos espectrais no algoritmo RF, tendo como área de estudo três municípios do nordeste do Estado de São Paulo.

2. MATERIAL E MÉTODOS

Os tópicos descritos no material e métodos estão relacionados à área de estudo, ao calendário agrícola da cana-de-açúcar e da soja e aos métodos de pré-processamento e processamento dos dados de sensoriamento remoto.

2.1 Descrição da área de estudo e das culturas agrícolas tipicamente plantadas nessa região

A área de estudo compreende 121,4 mil ha, e corresponde aos municípios de Guará, Ipuã e São Joaquim da Barra, localizados na região nordeste do estado de São Paulo (Figura 1).

Esta região pode ser considerada representativa de parte das condições agrícolas ocorrentes nas regiões sul e sudeste do Brasil (Vieira *et al.*, 2012). Vários estudos que trataram da classificação de culturas, monitoramento, segmentação de imagens, amostragem e representação da resposta espectral das culturas ao longo do ano safra, foram desenvolvidos nesta região (Oliveira *et al.*, 2003; Sanches *et al.*, 2005; Vieira *et al.*, 2012).

As áreas agrícolas dos municípios citados são formadas principalmente por cana-de-açúcar (58,5 mil ha) e culturas temporárias, entre as quais, destacam-se o milho (18,8 mil ha) e a soja (31,8 mil ha) (IBGE, 2015).

Nessa região, a soja e o milho levam, em média, 120 a 150 dias para serem colhidos, respectivamente (MAPA, 2015). A cultura da soja é geralmente plantada no final de outubro e colhida entre meados de fevereiro e de março. O milho pode ser plantado no período da safra (outubro a fevereiro), ou como segunda safra, logo após a colheita da soja (denominado “milho safrinha”) (Figura 2). Em 2013, 74 % do milho plantado no estado de SP correspondeu ao plantio de milho safrinha (IEA, 2015).

A cana-de-açúcar é uma cultura semi-perene que pode ser colhida após um ciclo de 360 dias (cana-de-ano) ou de 540 dias (cana-de-ano-e-meio) após o plantio, e a colheita ocorre entre os meses abril a dezembro (Aguiar *et al.*, 2008; Aguiar *et al.*, 2011) (Figura 2).

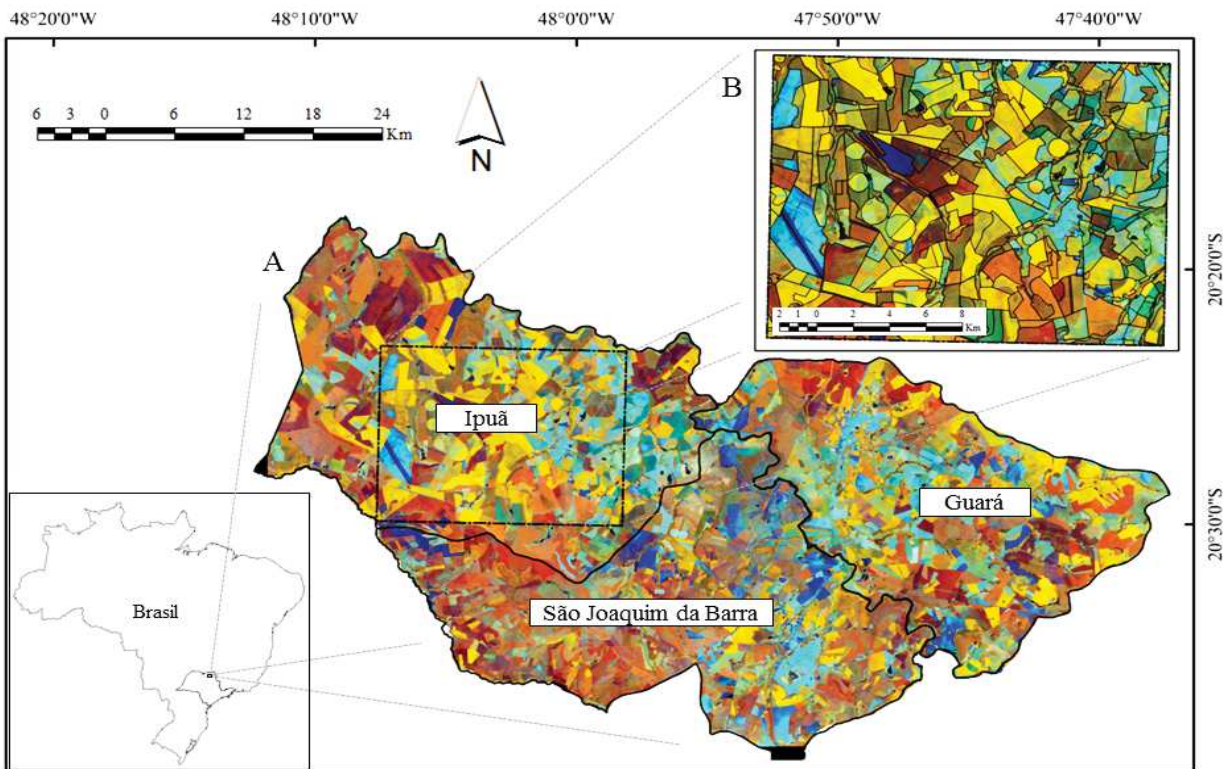


Fig 1 - Área de Estudo.

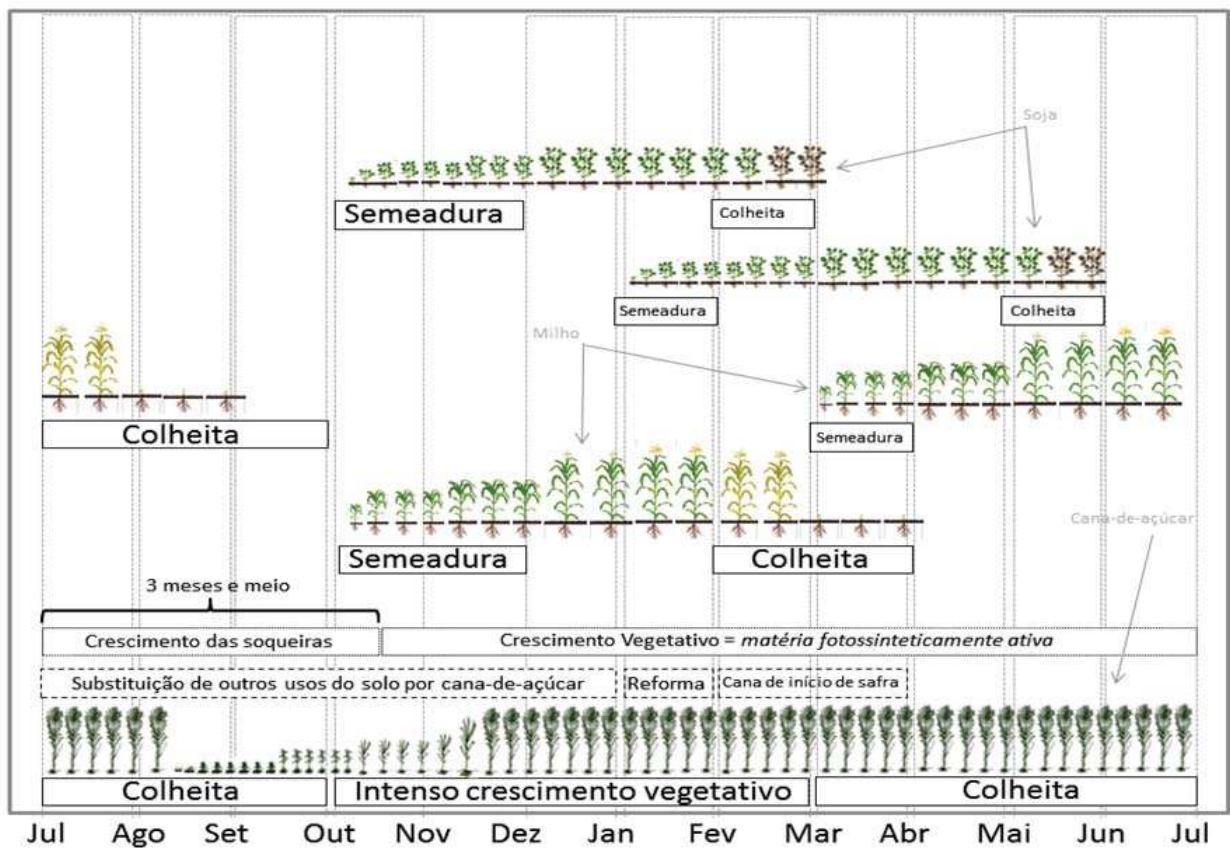


Fig. 2 - Calendário agrícola da cana-de-açúcar e soja para a região de estudo.

2.2 Pré-processamento e processamento dos dados de sensoriamento remoto

As cenas de sensoriamento remoto utilizadas no presente trabalho foram adquiridas junto ao Catálogo de Imagens do Instituto Nacional de Pesquisas Espaciais (INPE), disponível no endereço (<http://www.dgi.inpe.br/CDSR/>). As cenas que recobriram a área de estudo corresponderam à órbita 220 ponto 74 do “Worldwide Reference System” (WRS-2), oriundas dos dias 20/set./2000 (S) e 22/out./2000 (O) do sensor TM/Landsat-5; e 19/fev./2001 (F) e 23/mar./2001 (M) do sensor ETM+/Landsat-7.

Entre o conjunto de dados oferecidos, foram utilizadas apenas as bandas B3 (vermelho: 0,63–0,69 μm); B4 (infravermelho próximo: 0,76–0,90 μm) e B5 (infravermelho médio: 1,55–1,75 μm), provenientes de S, O, F e M, por terem maior relação com os aspectos fenológico-espectrais das culturas agrícolas (Sanches *et al.*, 2005).

O pré-processamento dos dados foi realizado conforme Vieira *et al.* (2012), tendo consistido no georrefenciamento, na correção radiométrica, e posterior correção atmosférica e normalização das cenas. Em seguida, no processamento de S, O, F e M, foi aplicado sobre B3, B4 e B5, o segmentador multi-resolução (SM) implementado no software Ecognition Developer $\text{\textcircled{R}}$ 64 (ED64) (Trimble Geospatial, 2009).

O algoritmo SM é do tipo crescimento de regiões (Espindola & Câmara, 2007) e agrupa os pixels de cada objeto em função da definição de seis parâmetros: Fator de Escala (Fe), Cor (Cr), Forma (Fm), Suavidade (Sv), Compacidade (Cp) e Peso para cada banda (Pb).

Conforme Baatz & Schäpe, (2000) e Fernandes *et al.* (2012), Fe determina a heterogeneidade máxima permitida para os objetos da imagem; Fm consiste na união dos parâmetros Cp (razão da área do objeto e o raio da circunferência abrangente) e Sv (expressa o limite entre os objetos, sendo que, quanto mais fractal, maior o índice de forma); Cr é inversamente proporcional ao parâmetro Fm; por fim, Pb é atribuído pelo usuário em função da ponderação (entre zero e um) que deseja utilizar para cada banda utilizada. Para o Fe, os valores variam de zero a um valor finito definido pelo usuário, enquanto que, para Fm, Cp, Sv e Cr, os valores variam de zero a um.

É importante destacar que, na plataforma ED64, apenas é necessária a definição dos parâmetros Fe, Fm, Cp e Pb.

Para a escolha da melhor segmentação a ser utilizada na fase de aplicação do algoritmo de classificação RF, foi gerado um conjunto de 500 segmentações possíveis para a área de estudo, conjunto este definido pela variação de Fe (variando de 5 a 100, em intervalos de 5) e Cp e Fm (variando de 0,1 a 0,9, em intervalos de 0,2). Para Pb, foi considerada uma ponderação constante e igual a um para as três bandas e datas Landsat utilizadas, ou sejam, B3, B4 e B5 das datas S, O, F e M, como recomendado por Vieira *et al.*, (2012).

No intuito de encontrar os polígonos gerados pelo SM que fossem os mais semelhantes possíveis aos polígonos realmente encontrados in situ, foi utilizado o Índice de Avaliação da Segmentação (IAVAS), proposto por Oliveira *et al.* (2003). Este índice avalia a coerência entre um conjunto de segmentações geradas por um algoritmo de segmentação (que, no caso presente, foi o SM) e uma segmentação de referência (que, no presente estudo, foi gerada manualmente via tela do computador).

Os valores de IAVAS mais próximos de zero representam as segmentações mais semelhantes à segmentação de referência, sendo que, para encontrar a melhor segmentação, foi implementado um algoritmo em linguagem R (R Development Core Team, 2014), obtendo-se então os melhores valores para os parâmetros Fe, Fm e Cp.

Para utilizar o RF, há uma fase de fornecimento de amostras de treinamento, uma vez que este algoritmo é do tipo supervisionado. Para alocar as amostras de treinamento, foi utilizado o mapa de referência adaptado a partir dos mapas gerados por Vieira *et al.* (2012). É oportuno destacar que este mapa de referência foi confeccionado via o programa ED64, tendo por base a melhor segmentação determinada pelo menor valor de IAVAS, conforme descrito anteriormente. Neste mapa da área de estudo, foram formados 1971 polígonos (cada um correspondendo a uma parcela agrícola).

Com base em Vieira *et al.* (2012), foi definido que um número adequado de amostras de treinamento deveria ser da ordem de 320 amostras

(16% do total). Desta forma, foram selecionadas, aleatoriamente, 103 amostras de cana-de-açúcar, 87 de soja e 130 de outros (todas as outras classes que não eram cana-de-açúcar ou soja).

O próximo passo foi a extração de atributos pelo ED64, usando as ferramentas Feature View e Export Results. Visando evitar problemas de sobreajuste por excesso de parâmetros (Breiman, 2001), no presente trabalho, para as quatro datas de dados Landsat utilizadas, foram considerados apenas os seguintes atributos espectrais: médias das bandas B3, B4 e B5; desvios-padrão das mesmas três bandas e NDVIs.

De posse dos citados atributos, para determinar como a multitemporalidade das imagens poderia fornecer os melhores resultados de classificações, foram testados oito diferentes conjuntos de datas (Tabela 1), aplicando o algoritmo RF implementado em linguagem R (R Development Core Team, 2014).

O RF gerou um conjunto de 1000 árvores de decisão (AD) para cada um dos oito conjuntos de datas (Tabela 1) e, desta forma, neste momento, chegou-se a 8000 AD.

Tabela 1: Conjunto de imagens obtido pela variação do número de datas Landsat utilizadas para treinamento do modelo de classificação Random Forest (RF)

Nº imagens	Nome do modelo
1 - Uma	F (fev.)
2 - Duas	SF (set. + fev.)
3 - Duas	OF (out. + fev.)
4 - Duas	FM (fev. + mar.)
5 - Três	SOF (set. + out. + fev.)
6 - Três	OFM (out. + fev. + mar.)
7 - Três	SFM (set. + fev. + mar.)
8 - Quatro	SOFM (set. + out. + fev. + mar.)

O passo seguinte realizado pelo RF foi calcular a confiabilidade percentual com que cada segmento pôde ser atribuído a uma determinada classe (cana-de-açúcar, soja ou outros). Por exemplo, se o RF prediz que um determinado polígono apresenta 88% de probabilidade a posteriori para ser classificado como cana-de-açúcar, o algoritmo atribuirá, para este polígono, a classe cana-de-açúcar e não outra classe qualquer. Desta forma, foram obtidas oito classificações, uma para cada conjunto de

datas (Tabela 1).

Finalmente, para a avaliação das exatidões das oito classificações RF, foram obtidas as matrizes de confusão, os coeficientes de exatidão global (EG) e o Kappa de Cohen (K), segundo Congalton & Green (2009), considerando o mapa de referência obtido conforme exposto anteriormente.

As diferenças entre as exatidões das oito classificações foram avaliadas estatisticamente por meio do teste de Z, comparando o desempenho do modelo SOFM (todas as datas) com os demais conjuntos de datas, ao nível de 5% de significância.

3. RESULTADOS E DISCUSSÃO

Os resultados do presente trabalho serão apresentados em dois tópicos principais: o primeiro refere-se à avaliação do IAVAS como ferramenta para selecionar a melhor combinação entre parâmetros do segmentador SM; o segundo refere-se ao desempenho do OBIA/RF, mediante o uso de diferentes datas de imagens Landsat para classificar cana-de-açúcar e soja conjuntamente.

3.1 Avaliação Quantitativa da Segmentação pelos IAVAS

A Figura 3 ilustra o comportamento dos 500 IAVAS testados, mostrando os parâmetros Fe (eixo x), Fm (eixo y) e Cp (range variando de 0,1 a 0,9 para cada Fm). O menor valor de IAVAS foi 4,52 (melhor resultado), ou seja, com os parâmetros Fe = 35, Fm = 0,3 e Cp = 0,5.

Conforme Oliveira *et al.* (2003), Vieira *et al.* (2012) e Fernandes *et al.* (2012), as mais indicadas formas de avaliar o desempenho de uma segmentação são feitas: (a) através da comparação entre a segmentação de referência e a segmentação obtida pelo uso dos parâmetros do melhor IAVAS; e (b) comparando a melhor segmentação IAVAS com a imagem estudada, via sobreposição.

Desta forma, a Figura 4 ilustra a coerência entre a segmentação de referência (gerada manualmente via tela do computador) e a obtida via melhor IAVAS (4,52). Observa-se que existe coerência entre ambas. Adicionalmente, ao comparar os segmentos do melhor IAVAS com a imagem utilizada, também fica corroborada esta coerência. Portanto, pode-se depreender que o IAVAS possui a capacidade de selecionar valores adequados dos parâmetros de uma segmentação.

Trabalhando também em áreas agrícolas,

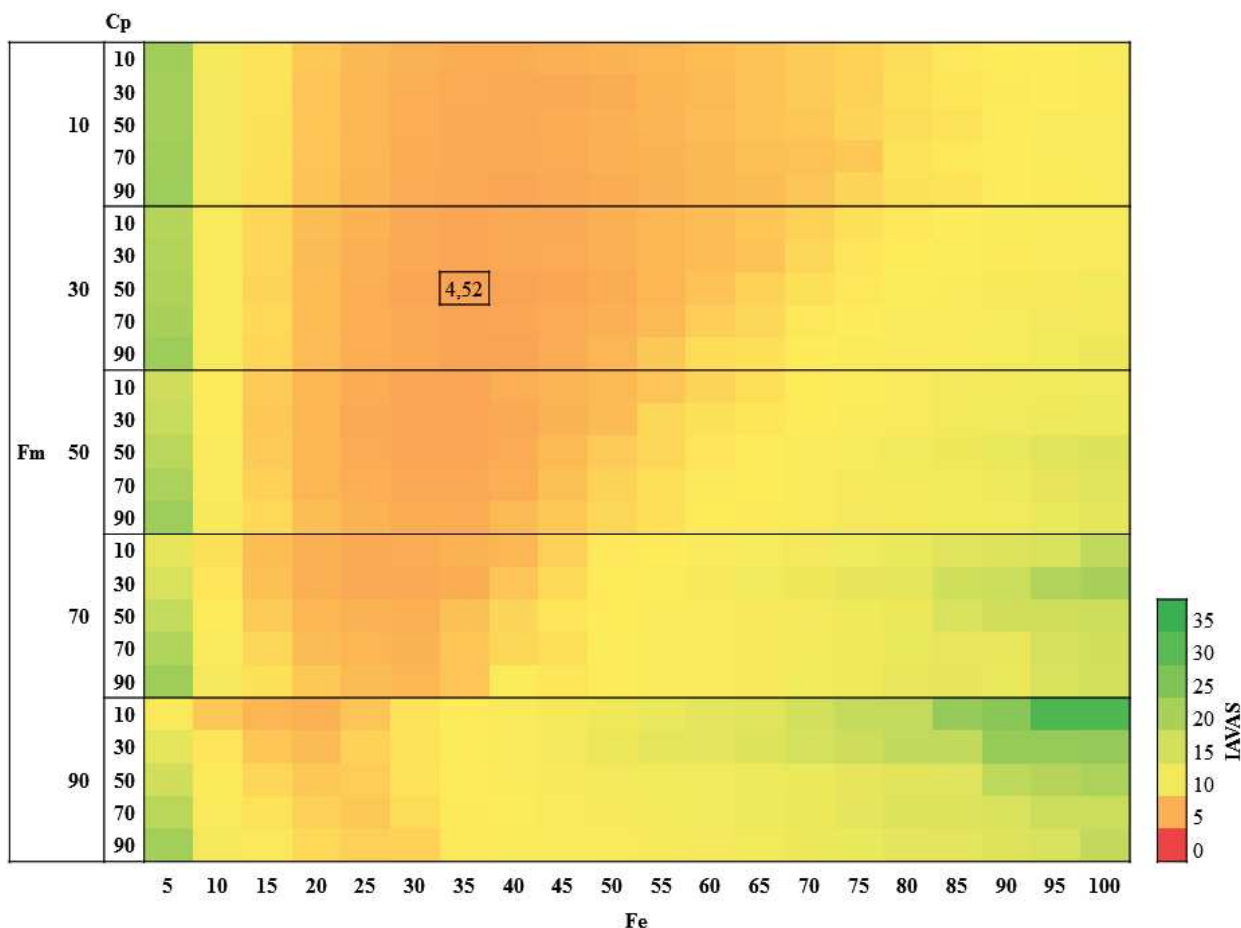


Fig 3 - Ilustração do comportamento dos 500 IAVAS testados para encontrar a melhor segmentação da área de estudo. No eixo x, o parâmetro Fe (variando de 5 a 100) e no eixo y, os parâmetros Fm (variando de 0,1 a 0,9) e Cp (variando de 0,1 a 0,9). Os valores de IAVAS variaram de 4,52 a 32,09 (escala de cores).

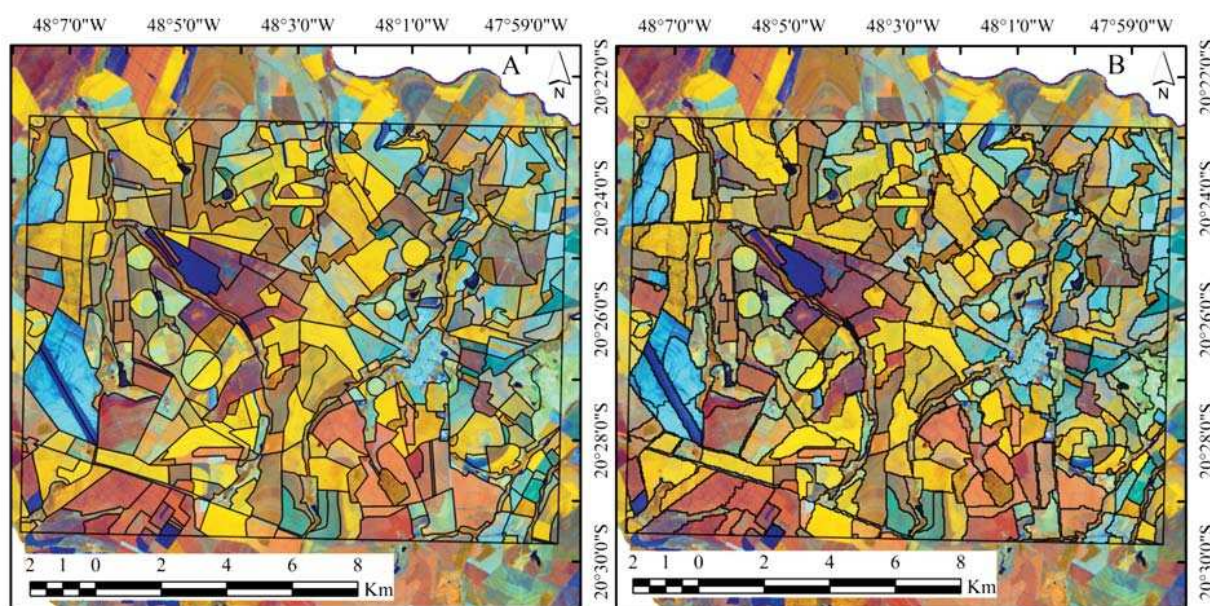


Fig 4 - Ilustração do comportamento dos 500 IAVAS testados para encontrar a melhor segmentação da área de estudo. No eixo x, o parâmetro Fe (variando de 5 a 100) e no eixo y, os parâmetros Fm (variando de 0,1 a 0,9) e Cp (variando de 0,1 a 0,9). Os valores de IAVAS variaram de 4,52 a 32,09 (escala de cores).

outros autores obtiveram melhores segmentações quando utilizaram valores de C_p e F_m de 0,1 e 0,7 (Peña-Barragán *et al.*, 2011) e 0,22 e 0,32 (Vieira *et al.*, 2012), sendo que, no presente trabalho, estes parâmetros foram 0,3 e 0,5, respectivamente. É possível constatar que os valores utilizados para o parâmetros F_m , em todos os trabalhos supra referidos, são baixos.

De fato, de acordo com Baatz e Schäpe (2000) e Fernandes *et al.* (2012), quanto menor o valor de F_m e maior o valor de C_p , menos fractal (menos aleatório) é o limite entre os objetos. Uma vez que as parcelas agrícolas são delimitadas por fatores antrópicos decorrentes de tipos de culturas plantadas, presença de estradas, hidrografia, tipos de relevos, etc, os valores de F_m deverão ser menores para segmentação de áreas agrícolas. Na figura 3 isto é perceptível, pois como a área do presente estudo é basicamente agrícola, os menores valores de IAVAS foram encontrados na parte superior do gráfico, local em que são posicionados os menores valores de F_m .

Por outro lado, quanto ao parâmetro F_e , Peña-Barragán *et al.* (2011) utilizaram o valor 100 e Vieira *et al.* (2012) o valor 20. No presente estudo, o valor determinado pelo melhor IAVAS foi de 35. Peña-Barragán *et al.* (2011) trabalharam com pixels de 15m e Vieira *et al.* (2012) com pixels de 30m, assim como no presente estudo, em que também foram utilizados pixels de 30m. Desta forma, podemos considerar que o F_e da ordem de 35 está coerente com Vieira *et al.* (2012), em função do tamanho de pixel.

3.2 Desempenho da classificação Random Forest

A classificação RF é utilizada no processo de OBIA (Object Based Image Analysis) e vem sendo avaliada positivamente na área do sensoriamento remoto agrícola e é baseada na geração de grande número de árvores de decisão, sendo do tipo supervisionada. Após uma segmentação, é necessária a fase denominada treinamento, com base em atributos extraídos dos dados de sensoriamento remoto, o que permitirá ao classificador atribuir a qual classe ele pertence cada polígono (objeto). Isto pode ser realizado de várias formas, e as mais citadas na literatura são: ou por mineração de dados (PEÑA-BARRAGÁN *et al.* 2011, VIEIRA *et al.*

2012; RODRIGUES & KUX, 2015) ou por um conjunto (floresta) de árvores de decisão (LONG *et al.*, 2013).

No caso do presente estudo, após a geração de um conjunto de 1000 AD, cujos atributos eram provenientes dos oito conjuntos de datas de dados Landsat (Tabela 1), a classificação RF calculou a porcentagem de acerto de predição da classe atribuída a cada um dos 1971 polígonos da área de estudo.

Assim, a RF aplicada para os oito diferentes conjuntos de datas (SOFM, SFM, OFM, SOF, FM, OF, SF e F) gerou, então, oito classificações, permitindo, posteriormente; avaliar tanto o desempenho da RF, quanto se os usos de diferentes números de datas apresentavam performances de mapeamento significativamente diferentes. Para esta avaliação foi utilizado um mapa de referência gerado a partir de dados obtidos *in situ* (Vieira *et al.*, 2012) e, desta forma, oito matrizes de confusão foram obtidas.

A Figura 5 ilustra os valores de EG e de Kappa para cada conjunto de datas avaliado. Verifica-se que a EG variou de 78,28 a 85,95%, sendo o valor mais baixo obtido quando analisada apenas uma data (F). Isso era esperado, visto que culturas agrícolas são alvos dinâmicos, passam por diferentes fases fenológicas, e, portanto, o aspecto multitemporal tem de ser levado em consideração nos algoritmos para classificação

Long *et al.* (2013) comentam que um maior número de datas recobrimo os calendários agrícolas das culturas em classificação, em geral favorece a formação de um modelo de conhecimento mais vantajoso. Os maiores valores de EG foram constatados para FM (85,95%) e para SOFM (85,88%). Nesse caso, percebe-se que o uso de um número maior de datas de imagens Landsat (e.g. S, O, F, M) não necessariamente apresentará o melhor desempenho de classificação simplesmente por recobrir uma faixa de tempo maior dos dois calendários agrícolas em estudo (cana-de-açúcar e soja).

O importante é incluir as datas, dentro do ano safra das culturas da cana-de-açúcar e soja, de maior importância para a identificação/separação das culturas analisadas. Breiman (2001) comenta que um exagerado número de atributos, nem sempre favorecerá um maior desempenho do classificador RF. Por isso, é importante conhecer

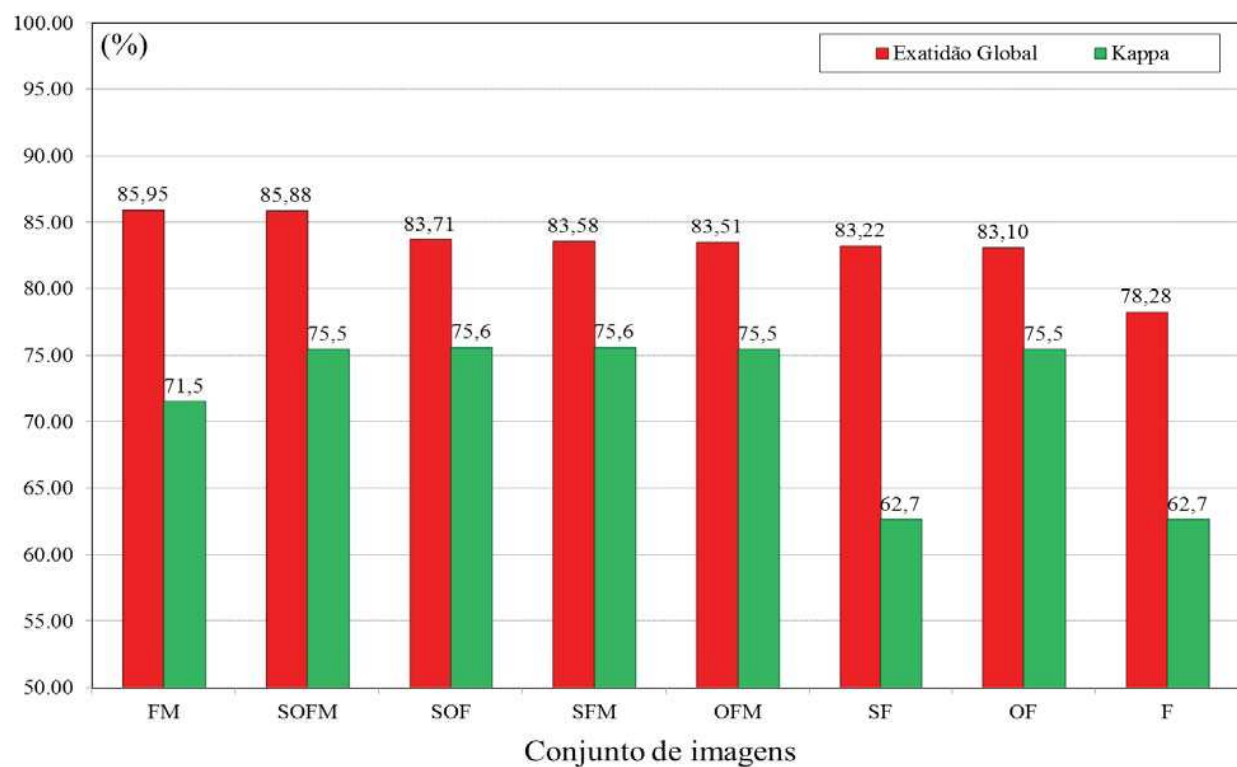


Fig. 5 - Avaliação do mapeamento dos oito modelos RF utilizados para a classificação automática de cana-de-açúcar, soja e outros na variação dos conjuntos de datas das imagens Landsat. Onde: S (Setembro), O (Outubro), Fevereiro (F) e Março (M). Da esquerda para a direita, o resultado da exatidão global (EG) (barra verde) está em ordem decrescente.

adequadamente os calendários agrícolas das culturas presentes, para que consigamos usar as datas-chaves (dentro do calendário agrícola da cana-de-açúcar e soja) que possam propiciar o melhor desempenho do modelo de conhecimento RF. Com isto, também se diminuirá o problema causado por “overfitting”. Com base nos resultados obtidos no presente estudo, verificou-se que as datas “chave” para a classificação de soja e cana-de-açúcar correspondem aos meses de fevereiro e março.

As datas F e M, usadas conjuntamente, destacaram-se por terem fornecido ao classificador atributos que melhor representaram as duas culturas em relação aos demais alvos presentes na área de estudo. De fato, em estudos com dados multitemporais LL, Rizzi & Rudorff (2005), Sugawara *et al.* (2008) e Rudorff *et al.* (2010) já haviam constatado que imagens obtidas em fevereiro e março destacam-se para separação entre culturas semi-perenes (no caso, cana-de-açúcar), temporárias (soja e milho) e as demais classes (outros, ou seja, não-cana-de-açúcar e não-soja).

Se compararmos os valores de EG obtidos pelos modelos que incluíram dados de fevereiro e março (FM, SOFM, SFM e OFM, EG 83,51 a 85,95 %) no presente estudo aos EGs dos trabalhos de Vieira *et al.* (2012) (94%, para duas classes, quatro datas de um ano safra, usando J48 e mineração de dados), Peña-Bárragan *et al.* (2011) (79%, para onze classes, duas datas e utilizando também J48 e mineração de dados) e Long *et al.* (2013) (85%, para três classes, duas datas e RF), perceberemos que tais indicadores de exatidão são relativamente similares.

Verifica-se então que os algoritmos classificadores utilizados no processo de OBIA apresentam desempenhos que podem ser considerados substanciais para formação dos objetos da área de estudo, destacando-se que alguns parâmetros devem ser levados em conta como: número de classes de interesse, número de datas dos dados de sensoriamento remoto dentro do ano safra e métodos de formação dos modelos de conhecimento.

O parâmetro estatístico Kappa de Cohen, em estudos de classificação de imagens, é

geralmente calculado como uma medida do acerto das classificações versus concordâncias aleatórias (Congalton & Green, 2009). Assim, a Tabela 2 mostra as significâncias dos parâmetros Kappa resultantes dos diferentes conjuntos de datas testados (Tabela 1) em relação ao conjunto de todas as quatro datas (SOFM), visando verificar se o uso de um maior número (ou menor) de datas acarreta diferenças significativas (5% de confiança pelo teste de Z) de qualidade de classificação.

Tabela 2: Significâncias entre os Kappas estatísticos resultantes do conjunto de imagens SOFM para os demais conjuntos utilizados para classificação final de cana-de-açúcar, soja e outros, em imagens multitemporais Landsat

Comparação dos modelos RF	Teste de (Z)	p-Valor
SOFM vs. F	9,13	0,0001 *
SOFM vs. SF	3,00	0,0014 *
SOFM vs. FM	0,01	0,5000 ns
SOFM vs. OF	0,01	0,5000 ns
SOFM vs. SOF	0,01	0,5000 ns
SOFM vs. OFM	0,12	0,4520 ns
SOFM vs. SFM	0,01	0,5000 ns

Nota: ^a A análise foi obtida pelo resultado do Kappa de cada um dos modelos, ^(ns) não significativo ao nível de 95 % de confiança (*) significativo ao nível de 95 %.

Verifica-se na Tabela 2 que houve dois casos que apresentaram diferença significativa de desempenho em relação ao uso de quatro datas no ano safra, F e SF, os quais tiveram os mais baixos valores de Kappa (ambos com 62,66%). Para os demais conjuntos de datas (SOFM x FM, SOFM x OF, SOFM x SOF, SOFM x OFM e SOFM x SFM) os respectivos Kappas são significativamente iguais ao uso de quatro datas, ao nível de 5%, pelo teste Z.

De fato, com relação à questão do número de datas multitemporais no ano safra para classificação de culturas agrícolas, uma dificuldade já realçada por Rizzi & Rudorff (2005), Sugawara *et al.* (2008) e Eberhardt *et al.* (2015) é a de se conseguir imagens LL livres de nuvens e, portanto, quanto menos se depender do número de datas de imagens, maior será a probabilidade de sucesso, quando se pretende construir um sistema operacional de mapeamentos agrícolas com tais tipos de dados.

Se compararmos os resultados de Kappa deste trabalho de 78,3 com os resultados encontrados nos trabalhos de Peña-Barragán *et al.* (2011) e Long *et al.* (2013) ambos com Kappa de 0,75 e desenvolvidos em áreas agrícolas americanas, percebemos que são similares. No entanto, o calendário agrícola do Brasil é diversificado. Isto dificulta a inserção de imagens em datas estratégicas para classificação das culturas agrícolas no Brasil com o uso de sensoriamento remoto.

Assim, mesmo que soubéssemos as datas ideais em relação ao ciclo fenológico para a classificação de uma cultura, dificilmente podemos contar que sempre iremos conseguir imagens dessas fases da cultura no Brasil, tanto por causa da diversidade do calendário agrícola, que faz com que tenhamos no mesmo momento no campo a cultura em diversas fases do ciclo fenológico como por causa da ocorrência de nuvens na época principal dos plantios. Assim, para culturas de ciclo anual e curto, devemos inicialmente trabalhar com todas as imagens disponíveis, mesmo que haja redundância de informação nelas, pois dificilmente teremos o conjunto ideal de datas.

Para obter melhores resultados de mapeamento de cana-de-açúcar e soja, o uso de imagens multi-sensores é um método eficaz e que obteve satisfatório resultados de classificação com uso de sensores ativos como os radares, por exemplo (MORAN *et al.*, 2002; LARRAÑAGA *et al.*, 2011).

Outra forma de melhorar os resultados obtidos com a classificação automática destas culturas seria utilizar imagens da banda pancromática do sensor OLI fusionada com as bandas dos demais canais espectrais deste sensor (MULIANGA *et al.*, 2015), para melhorar a resolução espacial dos pixels e assim diminuir a quantidade de pixels impuros presentes nos objetivos formados pela segmentação.

4. CONCLUSÃO

O IAVAS demonstrou capacidade de selecionar, de forma automática, os valores mais adequados dos parâmetros (Fe, Fm e CP) para uma segmentação em dados multitemporais Landsat, gerada via algoritmo segmentador multi-resolução, como o implementado no ED64.

Duas datas de imagens Landsat posicio-

nadas no final da safra de verão (F e M, no presente caso) foram suficientes para obter uma exatidão global, para cana-de-açúcar e soja sendo classificadas conjuntamente via o algoritmo Random Forest e OBIA, da ordem de aproximadamente 86%.

Atributos espectrais das bandas 3 (0,63–0,69 μm), 4 (0,76–0,90 μm) e 5 (1,55–1,75 μm) do TM e ETM+/Landsat, como médias, desvios-padrão e NDVIs foram suficientes para classificação conjunta de três classes: cana-de-açúcar, soja e outros (não-cana e não-soja) para as condições agrícolas do nordeste do Estado de São Paulo, usando RF.

5. AGRADECIMENTOS

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo provimento da bolsa de doutorado que possibilitou a elaboração desta pesquisa, bem como ao Programa Ciência sem Fronteiras (CNPq/CAPES via Projeto 402.597/2012-5).

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMI, M.; MOREIRA, M.A.; RUDORFF, B.F.T.; FREITAS, C.C.; FARIA, R.T.; DEPPE, F. Painel amostral para estimativa de áreas agrícolas. **Pesquisa Agropecuária Brasileira**, v. 2, n.1, p.81-88, 2007.

AGUIAR, D.A.; RUDORFF, B.F.T.; RIZZI, R.; SHIMABUKURO, Y. Monitoramento da colheita da cana-de-açúcar por meio de imagens MODIS, **RBC - Revista Brasileira de Cartografia**, v. 60, n. 4, p. 375-383, 2008. CITAÇÃO ERRADA VEJA NO SITE!

AGUIAR, D.A.; RUDORFF, B.F.T.; SILVA, W.F.; ADAMI, M.; MELLO, M.P. Remote Sensing Images in Support of Environmental Protocol: Monitoring the Sugarcane Harvest in São Paulo State, Brazil, **Remote Sensing**, v.3, n.12, p. 2682-2703, 2011.

BAATZ, M.; SCHÄPE, A. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. **Journal of Photogrammetry and Remote Sensing**, v.58, n.10, p. 12-23, 2000.

BREIMAN, L. Random forests. **Machine Learning**, v.45, n.3, p. 5-32, 2001.

CANASAT: mapeamento da cana via imagens de satélite de observação da terra. Disponível em: <<http://www.dsr.inpe.br/laf/canasat>>. Acesso em: 07 out. 2014.

CONGALTON, R.G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. Boca Raton: 2nd, CRC Press, 2009. 183 p.

EBERHARDT, I.D.R.; LUIZ, A.J.B.; FORMAGGIO, A.R.; SANCHES, I.D. Detecção de áreas agrícolas em tempo quase real com imagens Modis. **Pesquisa Agropecuária Brasileira**, v.50, n.7, p. 605-615, 2015.

ESPÍNDOLA, G. M.; CÂMARA, G. Ajuste de parâmetros no algoritmo de segmentação de imagens do software eCognition®. In: XIII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, Florianópolis, 2007. Anais. São José dos Campos: INPE, Instituto Nacional de Pesquisas Espaciais, 2007 p. 5739-5744.

FERNANDES, R.R.; NUNES, G.N.; SILVA, T.S.F. Classificação orientada a objetos na caracterização da cobertura da terra no Araguaia. **Pesquisa Agropecuária Brasileira**, v.47, n. 9, p. 1251-1260, 2012.

FAOSTAT. Browser Data. Disponível em: <<http://faostat3.fao.org/browse/Q/QC/E>>. Acesso em 10 mai. 2015.

GAROFALO, D. F. T.; MESSIAS, C. G.; LIESENBERG, V.; BOLFE, E. L.; FERREIRA, M. C. Análise comparativa de classificadores digitais em imagens do Landsat-8 aplicados ao mapeamento temático. **Pesquisa Agropecuária Brasileira**, v. 50, n. 7, p. 593-604, 2015.

IBGE. **Sistema IBGE de Recuperação Automática**. Disponível em: <<http://www.sidra.ibge.gov.br/>> Acesso em: 08 de mar. 2015.

IEA. **Área e Produção de milho de primeira safra, de segunda safra e irrigado, em 2014, no estado de São Paulo**. Disponível em: <<http://ciagri.iea.sp.gov.br/nia1/>> Acesso em: 08 de mar. 2015.

LAMPARELLI, R.A.C., CARVALHO, W.M.O.; MERCANTE, E. Mapeamento e semeaduras de soja (*Glycine max* (L.) Merr.) mediante dados MODIS/TERRA e TM/Landsat 5: um

- comparativo. **Engenharia Agrícola**, v. 28, n. 2, p. 334–344, 2008.
- LARRAÑAGA, A.; ÁLVAREZ-MOZOS, J.; ALBIZUA, L. Crop classification in rain-fed and irrigated agricultural areas using Landsat TM and ALOS/PALSAR data. **Canadian Journal of Remote Sensing**, v.37, n.15, p. 157–170, 2011.
- LONG, J.A.; LAWRENCE, R.L.; GREENWOOD, M. C.; MILLER, P. R. Object-Oriented Crop Classification Using Multitemporal ETM + SLC-off Imagery and Random Forest. **GIScience and Remote Sensing**, v.50, n.4, p.418-436, 2013.
- LUIZ, A.J.B.; FORMAGGIO, A.R.; EPIPHANIO, J.C.N.; ARENAS-TOLEDO, J.M.; GOLTZ, E.; BRANDÃO, D. Estimativa amostral objetiva de área plantada regional, apoiada em imagens de sensoriamento remoto. **Pesquisa agropecuária Brasileira**, v.47, n.9, p.1279-1287, 2012.
- MAPA. Ministério da Agricultura Pesca e Abastecimento. Disponível em: < <http://www.agricultura.gov.br/politica-agricola>> Acesso em: 08 de mar. 2015.
- MYINT, S. W.; GOBER, P.; BRAZEL, A.; GROSSMAN-CLARKE, S.; WENG, Q. Per-Pixel vs. Object-Based Classification of Urban Land Cover Extraction Using High Spatial Resolution Imagery. **Remote Sensing of Environment**, v.115, n.5, p.1145–61, 2011.
- MORAN, M.S.; HYMER, D.C.; QI, J.; KERR, Y. Comparison of ERS-2 SAR and Landsat TM imagery for monitoring agricultural crop and soil conditions. **Remote Sensing of Environment**, 2002, v. 79, n.4, 243–252
- MULINGA, b.; BÉGUÉ, A.; CLOUVEL, P.; TODOROFF, Mapping Cropping Practices of a Sugarcane-Based Cropping System in Kenya Using Remote Sensing, **Remote Sensing**, v.7, n.11, p. 14428-14444, 2016.
- OLIVEIRA, J. C.; FORMAGGIO, A. R.; EPIPHANIO, J. C. N.; LUIZ, A. J. B. Index for the Evaluation of Segmentation (IAVAS): An Application to Agriculture. **Mapping Sciences and Remote Sensing**, v.40, n.3, p.155-169, 2003.
- PEÑA-BARRAGÁN, J. M.; MOFFATT, K.; MGUGI, R. E. P.; SIX, J. Object-Based Crop Identification Using Multiple Vegetation Indices , Textural Features and Crop Phenology. **Remote Sensing of Environment**, v.115, n.6, p.13013-1316, 2011.
- R DEVELOPMENT CORE TEAM, R: A Language and Environment for Statistical Computing, **R Foundation for Statistical Computing**, Vienna, Austria, 2014
- RIZZI, R.; RUDORFF, B. F. T. Estimativa da área de soja no Rio Grande do Sul por meio de imagens Landsat. **Revista Brasileira de Cartografia**, v.3; n.57, p.226-234, 2005.
- RODRIGUES, T. C. S.; KUX, H. J. H. Uso do *software* livre *Interimage* para a classificação orientada a objetos em áreas de expansão urbana na Ilha do Maranhão. **Revista Brasileira de Cartografia**, v. 67, n. 3, p. 493-506, 2015. CITAÇÃO ERRADA VEJA NO SITE!
- 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. CITAÇÃO ERRADA VEJA NO SITE!
- RUDORFF, B. F. T.; SUGAWARA, L. M. Mapeamento da cana-de-açúcar na Região Centro-Sul via imagens de satélites. **Informe Agropecuário**, v.28, n.241, p.79-86, 2007. CITAÇÃO ERRADA VEJA NO SITE!
- SANCHES, I. D.; EPIPHANIO, J. C. N.; FORMAGGIO, A. R. Culturas Agrícolas Em Imagens Multitemporais Do Satélite Landsat. **Agricultura de São Paulo**, v.52, n.1, p.83-96, 2005.
- SUGAWARA, L. M.; RUDORFF, B. F. T. Viabilidade de Uso de Imagens Do Landsat Em Mapeamento de Área Cultivada Com Soja No Estado Do Paraná. **Pesquisa Agropecuária Brasileira**, v.43, n.12, p.1777-1783, 2008.
- TRIMBLE GEOSPATIAL. **Definiens eCognition developer. Version 8**. Munich: Trimble GeoSpatial, 2009.
- VIEIRA, M. A.; FORMAGGIO, A. R.; RENNÓ,

C. D. ATZBERGER, C.; AGUIAR, D. A.; MELLO, M. P. Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. **Remote Sensing of Environment**, v.123, p. 553-562, 2012.

WILKINSON, G. G. Results and implications of a study of fifteen years of satellite image classification experiments. **IEEE Transactions on Geoscience and Remote Sensing**, v.43, n.3, p.433-440, 2005.