

Mapeamento de espécies arbóreas em floresta tropical utilizando imagens hiperespectrais

Matheus Pinheiro Ferreira ¹
Maciel Zortea ²
Daniel Capella Zanotta ³
Yosio Edemir Shimabukuro ¹
Carlos Roberto de Souza Filho ⁴

¹ Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
{mpf, yosio}@dsr.inpe.br

² Norwegian Computing Center - NR
Gaustadalléen, 23 – NO-0373, Oslo, Noruega
maciel.zortea@nr.no

³ Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - IFRS
Rua Eng. Alfredo Huch, 475 - 96201-460 - Rio Grande - RS, Brasil
daniel.zanotta@riogrande.ifrs.edu.br

³ Universidade Estadual de Campinas - UNICAMP
Rua João Pandiá Calógeras, 51 - 13083-870 - Campinas - SP, Brasil
beto@ige.unicamp.br

Abstract. Tree species mapping in tropical forests using remote sensing imagery is challenging with the benefit to provide valuable insights for ecologists and land managers. Hyperspectral data proved to be feasible for this task, but it is still unclear how different classification methods perform. In this work, we evaluated Linear Discriminant Analysis (LDA), Radial Basis Function Support Vector Machines (RBF-SVM) and Random Forest (RF) for tree species discrimination and mapping in a tropical forest using airborne hyperspectral data. The effects of dimensionality reduction on classification performance were also assessed by selecting sets of 10, 20 and 30 bands. At the pixel level, LDA performed better than other methods (Average Accuracy (AA) =84.7%) using all (260) spectral bands for classification. However, RBF-SVM produced the best map of species using 30 selected bands (AA =90.4%) in an object based approach (OBIA). OBIA increased the AA of species mapping for all tested methods and reduced spatial noise.

Palavras-chave: hyperspectral remote sensing, image processing, ecology, sensoriamento remoto hiperespectral, processamento de imagens, ecologia.

1. Introdução

Levantamentos florísticos de florestas tropicais são fundamentais para sua conservação e manejo sustentável. No entanto, sua regularidade é comprometida por exigir dispendiosos inventários fitossociológicos em áreas normalmente de difícil acesso. Imagens orbitais possuem a vantagem de cobrir grandes áreas, mas tem sua utilização restrita para esse fim, devido, entre outros fatores, às resoluções espacial e espectral (Nagendra e Rocchini, 2008). Imagens obtidas por sensores hiperespectrais a bordo de aeronaves, possuem resolução espacial com suficiência para visualização de copas, individualizando, assim, a resposta espectral das árvores. Além disso, o elevado número de bandas desses sensores permite melhor caracterizar as espécies, através, por exemplo, da seleção de atributos que maximizem a variabilidade interespecífica. Estudos recentes demonstraram a potencialidade de dados hiperespectrais para identificação remota de espécies arbóreas em ambientes tropicais, tanto

ao nível foliar (Ferreira et al., 2013) quanto de dossel (Féret e Asner, 2013). Particularmente, ao nível de dossel, as técnicas para o mapeamento de espécies se encontram em estágios iniciais de desenvolvimento. Não há consenso na literatura sobre as potencialidades e limitações entre os diversos métodos de processamento de imagens. Ademais, pouco se conhece sobre o desempenho de sensores hiperespectrais em aplicações dessa natureza.

Nesse contexto, o objetivo desse trabalho é avaliar métodos de classificação para identificação de espécies arbóreas em uma floresta tropical, utilizando imagens hiperespectrais aerotransportadas. Primeiramente, modelos de classificação foram treinados e testados com amostras de *pixels* de diversas espécies, visando avaliar seu desempenho quanto à exatidão. Em seguida, os melhores modelos foram escolhidos para classificar *pixels* de uma imagem e produzir mapas das espécies arbóreas. A acurácia desses mapas foi avaliada com uma referência produzida por interpretação visual e dados de campo.

2. Material e Métodos

2.1 Área de estudo

A área de estudo compreende a Reserva de Santa Genebra (RSG) (22°48' - 22°50' S, 47°06' - 47°07' W), localizada no município de Campinas/SP (Figura 1). Esta área possui 251,8 hectares, onde a formação florestal predominante é a floresta estacional semi-decidual (Veloso et al., 1991). A RSG é um remanescente bem preservado de Floresta Atlântica e abriga alta diversidade de espécies arbóreas (Guaratini et al., 2008).

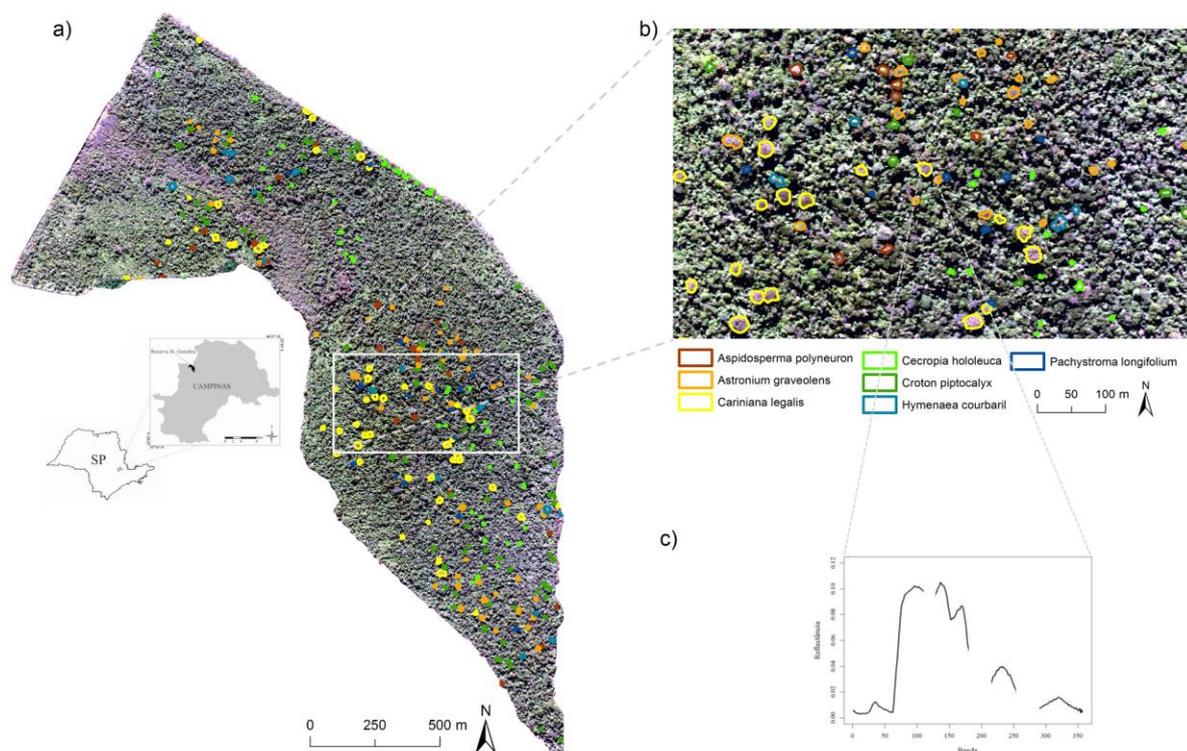


Figura 1. (a) Mapa de localização da Reserva de Santa Genebra. Os polígonos representam árvores georreferenciadas e identificadas ao nível de espécie. A imagem no *background* corresponde a uma composição colorida cor real (R-638 nm G-548 nm B-460 nm); (b) Recorte utilizado para realização deste trabalho com 7 espécies alvo de identificação sobre composição colorida cor real (R- 638 nm G-548 nm B-460 nm); (c) Resposta espectral média dos *pixels* que compõe uma copa de *Cariniana legalis* (Jequitibá-rosa).

2.2 Imagens hiperespectrais

Dados hiperespectrais foram adquiridos pelo instrumento dual ProSpecTIR-VS, composto pelos sensores AisaEAGLE e AisaHAWK (*Spectral Imaging, Inc.*) que cobrem a região do visível/infravermelho próximo (VNIR, 400-970 nm) e infravermelho de ondas curtas (SWIR, 970-2450 nm) do espectro eletromagnético, respectivamente. As imagens do VNIR e SWIR foram adquiridas simultaneamente, produzindo 357 bandas espectrais com 1 metro de resolução espacial. Um sistema de navegação interna (INS) forneceu os dados de Posicionamento Global (GPS) para geração do Mapa de Geometria Interna (IGM), que contém as informações necessárias para o georreferenciamento de cada *pixel* da imagem. A aquisição dos dados foi realizada em condições de céu aberto no dia 07 de junho de 2010, iniciando-se às 15:27h, com duração de 28 minutos. Dez linhas de voo foram necessárias para cobrir toda extensão da RSG.

2.3 Dados de campo

Inicialmente as copas de 400 árvores emergentes, claramente visualizadas em uma composição colorida de cor real com escala fixa de 1:1.500, foram manualmente delineadas. Em campo, com o auxílio de um dispositivo GPS foi possível visitar 310 árvores, sendo as mesmas identificadas ao nível de espécie (Figura 1-a).

2.4 Pré-processamento das imagens hiperespectrais e dos dados de campo

As imagens de cada linha de voo foram georreferenciadas com as respectivas *Geometric Look-up-Table* (GLT) geradas a partir dos arquivos IGM e mosaicadas. A correção atmosférica foi realizada sobre o mosaico com o *software Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes* (FLAASH), baseado no modelo de transferência radiativa MODTRAN-4. Bandas compreendidas nas regiões do espectro afetada por vapor d'água atmosférico, em torno de 1.400 e 1.900 nm, foram excluídas do conjunto hiperespectral. Bandas ao redor de 970 nm também foram excluídas, pois estão localizadas na faixa de transição entre os sensores e apresentam baixa relação sinal/ruído. No total, 260 bandas, abrangendo comprimentos de onda entre 450-2450 nm, foram retidas para uso posterior.

Após as correções geométrica e atmosférica, a resposta espectral dos *pixels* que compõe as árvores identificadas em campo foi extraída para formar uma base de dados composta de 310 árvores de sete espécies, conforme apresentado na Tabela 2.

Tabela 2. Lista das espécies, número de copas, *pixels* e mínimo e máximo Diâmetro à Altura do Peito (DAP).

Espécie	Nº de copas	<i>Pixels</i>	DAP (cm)
<i>Aspidosperma polyneuron</i>	25	2630	48-96
<i>Astronium graveolens</i>	59	5914	26-57
<i>Cariniana legalis</i>	50	13312	52-111
<i>Cecropia hololeuca</i>	60	2234	13-25
<i>Croton piptocalyx</i>	83	6375	25-71
<i>Hymenaea courbari</i>	18	3663	61-95
<i>Pachystroma longifolium</i>	15	914	49-63

2.5 Classificação

2.5.1 Seleção de amostras de treinamento e teste

Para realização deste trabalho, foi escolhida uma região localizada na porção central da RSG, com dimensões de 458x648 *pixels*, ou 29,7 hectares (Figura 1-b). De acordo com os dados de campo, no interior desta área, há 76 árvores das espécies listadas na Tabela 2. Os *pixels* que compõe estas árvores foram excluídos da base de dados original, garantindo, dessa

forma, a não utilização dos mesmos para treinamento dos classificadores. As 76 árvores serão utilizadas na avaliação da acurácia dos mapas de espécies.

A base de dados foi particionada nos conjuntos de treinamento e teste, em uma proporção de 60% e 40%, respectivamente. Além disso, buscou-se respeitar a árvore que originou a amostra, sendo que os *pixels* de teste sempre virão de árvores diferentes das utilizadas para o treinamento. Estudos anteriores demonstram que a desconsideração da árvore na formação das amostras de treinamento e teste pode superestimar a exatidão de classificação (Baldeck e Asner, 2014).

2.5.2 Métodos de classificação

A classificação de imagens hiperespectrais normalmente requer um procedimento de seleção de bandas, devido à elevada dimensionalidade deste tipo de dado. Os principais objetivos da seleção de atributos é evitar o efeito de Hughes (Hughes, 1968), ou seja, redução da acurácia com o aumento da dimensionalidade, e diminuir o custo computacional. Neste trabalho, foi utilizado o método *Sequential Forward Selection* (SFS) tendo como critério de seleção a separabilidade espectral estimada pela distância de Jeffreys-Matusita (JM) (Bruzzone et al., 1995). SFS é uma estratégia de busca que seleciona variáveis (bandas) de forma a maximizar a separabilidade espectral interespecífica. Neste trabalho, as classificações foram realizadas com 10, 20, 30 e 260 bandas, para verificar o efeito do número de atributos na exatidão de classificação.

Buscou-se avaliar o desempenho de métodos que têm se destacado na aplicação em questão como: Análise Discriminante Linear (LDA) (Féret e Asner, 2013), *Support Vector Machines* (SVM) (Dalponte et al., 2014) e *Random Forest* (RF) (Ghosh et al., 2013). LDA é uma técnica da estatística multivariada que busca obter uma combinação linear de características observadas, ou seja, bandas, que apresente o melhor poder de discriminação entre classes. Essa combinação linear é obtida assumindo-se que as classes possuem distribuição normal e compartilham da mesma matriz de covariância. SVM é uma técnica de aprendizado computacional baseada em *kernels*, que realizam a classificação construindo hiperplanos no espaço multidimensional para separar amostras de diferentes classes (Vapnik, 1995). Dentre os vários tipos de *kernels* existentes, este estudo aplicou somente o *kernel Radial Basis Function* (RBF), que possibilita o ajuste do parâmetro de penalidade (C) e gamma (γ), estimados em pontos de uma grade de parâmetros previamente definida. O algoritmo *Random Forest* (RF) (Breiman, 2001) é a combinação de diversas árvores de decisão (*Forest*), resultantes de um processo de partição aleatória (*Random*) do conjunto de dados. Cada árvore gera um voto que indica a decisão sobre a escolha da classe de uma amostra, assim, a classe que obteve o maior número de votos será atribuída a mesma. A classificação com RF depende do número de árvores e do número de variáveis, estipulados em 1000 e 10 neste trabalho.

Os modelos de classificação com melhor desempenho foram selecionados para classificar a imagem e gerar mapas de espécies. Adotou-se aqui, além da forma tradicional *pixel-a-pixel*, uma abordagem baseada em objeto. Primeiramente, a imagem em composição cor verdadeira (Figura 1-b) foi segmentada por um algoritmo de crescimento em regiões (Zanotta, 2014). Em seguida, os *pixels* de cada segmento foram classificados por cada método e uma regra de votação por maioria foi aplicada para definir o segmento à classe mais frequente. Considerando que o objetivo é mapear apenas algumas espécies arbóreas (Tabela 2), no processo de classificação, foram considerados apenas os *pixels* com altos valores de probabilidade de pertencer às classes de treinamento. Dessa forma, *pixels* classificados com menor confiabilidade não serão rotulados.

2.5.3 Avaliação de acurácia

O desempenho dos classificadores foi atestado a partir das amostras de teste. Foram calculados os valores de exatidão média e global ao se utilizar 10, 20, 30 e 260 bandas para classificação das espécies. A acurácia dos mapas de espécies gerados com os diferentes métodos foi estimada tendo-se como referência as 76 árvores identificadas em campo. Nesse caso, buscou-se avaliar a exatidão média e global, bem como a porcentagem de *pixels* corretamente classificados.

3. Resultados e Discussão

A Tabela 3 apresenta os valores de exatidão média e global, calculadas a partir das amostras de teste (seção 2.5.1), com 10, 20, 30 e 260 bandas utilizadas pelos métodos de classificação.

Tabela 3. Exatidão Média (EM) e Exatidão Global (EG) de classificação de sete espécies arbóreas tropicais dos métodos *Random Forest* (RF), *Radial Basis Function Support Vector Machines* (RBF-SVM) e Análise Discriminante Linear (LDA). Os melhores modelos de classificação estão destacados em cinza.

	RF		RBF-SVM		LDA	
	EM (%)	EG (%)	EM (%)	EG (%)	EM (%)	EG (%)
10 Bandas	59,3	72,5	71,3	79,4	72,1	75,1
20 Bandas	67,8	77,2	80,6	85,5	82,2	85,2
30 Bandas	67,9	77,4	80,9	85,8	83,7	87,1
260 Bandas	64,8	76,9	80,7	84,6	84,7	89,4

O procedimento de seleção de bandas diminuiu a dimensionalidade dos dados e elevou os valores de exatidão para os métodos RF e RBF-SVM. No entanto, esse procedimento impactou de forma negativa a classificação pelo método LDA, que obteve melhores resultados com base em todas as 260 bandas. Os modelos com maiores valores de exatidão foram escolhidos para produção dos mapas de espécies. A Tabela 4 apresenta os valores de exatidão média, global e a porcentagem de *pixels* corretamente classificados.

Tabela 4. Exatidão Média (EM), Exatidão Global (EG) e Porcentagem de *Pixels* Corretamente Classificados (PCC) ao se utilizar os métodos *Random Forest* (RF), *Radial Basis Function Support Vector Machines* (SVM) e Análise Discriminante Linear (LDA) para o mapeamento de espécies arbóreas tropicais, por classificação *pixel-a-pixel* e orientada a objeto. O método com melhor exatidão está destacado em cinza.

Método	Classificação <i>pixel-a-pixel</i>			Classificação orientada a objeto		
	EM (%)	EG (%)	PCC (%)	EM (%)	EG (%)	PCC (%)
RF	70,7	81,4	70,5	76,5	87,8	63,8
RBF-SVM	90	94,4	66,9	90,4	95,2	58,2
LDA	86,6	92,1	86,7	87,0	92,0	86,4

O procedimento de classificação orientada a objeto elevou a exatidão média para todos os métodos testados. Além disso, esse procedimento reduziu consideravelmente o ruído espacial, homogeneizando as imagens classificadas (Figura 2). O método com melhores resultados foi o RBF-SVM. Embora esse método tenha classificado uma porcentagem menor de *pixels* da referência (Tabela 4), o mapa produzido foi o que mais se assemelhou visualmente da mesma (Figura 2). É importante destacar que o mapa referência não representa o censo das sete espécies estudadas. Na realidade, buscou-se avaliar a capacidade dos métodos em classificar corretamente as 76 árvores que foram identificadas em campo. Na prática, é muito difícil realizar o censo de todas as espécies que ocorrem na área. Rotular todos os *pixels* da imagem

a certo número de espécies pode implicar numa generalização equivocada. Por esse motivo, optou-se por classificar apenas os *pixels* com alta probabilidade de pertencer a uma determinada classe de treinamento. A quantidade de *pixels* classificados por cada método variou consideravelmente. Embora o método LDA tenha atingido os melhores valores de exatidão ao nível de *pixel* (Tabela 3), o mapa classificado apresentou grande confusão com a classe *Pachystroma longifolium* (Figura 2). De forma semelhante, o método *Random Forest* classificou uma grande quantidade de *pixels*/segmentos como *Croton piptocalyx* (Figura 2).

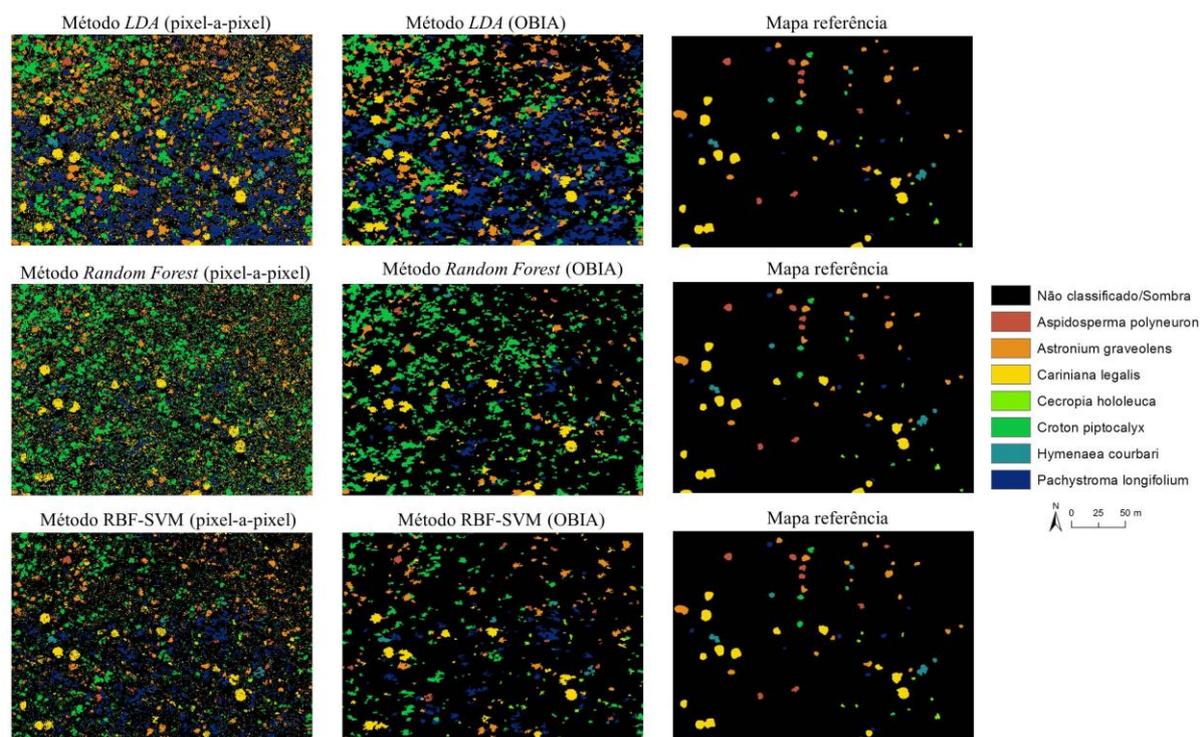


Figura 2. Mapas de espécies arbóreas pelos métodos de classificação Análise Discriminante Linear (LDA), *Random Forest* (RF) e *Radial Basis Function Support Vector Machine* (RBF-SVM) obtidos por classificação *pixel-a-pixel* e orientada a objeto (OBIA).

4. Conclusões

Neste trabalho foram avaliados métodos de classificação para o mapeamento de espécies arbóreas em floresta tropical utilizando dados hiperespectrais aerotransportados. A Análise Discriminante Linear (LDA) obteve os melhores resultados de exatidão média ao nível de *pixel*. No entanto, o método *Radial Basis Function Support Vector Machines* (RBF-SVM) obteve melhores resultados para geração de um mapa de espécies. A classificação orientada a objeto mostrou-se superior à tradicional *pixel-a-pixel* e também reduziu consideravelmente o ruído espacial. Trabalhos futuros irão concentrar-se no aprimoramento da exatidão de classificação, por meio da inclusão de atributos que ajudem a aumentar a separabilidade espectral interespecífica, como, por exemplo, índices de vegetação. Além disso, pretende-se aprimorar a classificação orientada a objeto através de melhorias no processo de segmentação.

Agradecimentos

Agradecemos a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela concessão de auxílio financeiro para o desenvolvimento desta pesquisa, através do projeto nº 2013/11.589-5, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

pelo auxílio financeiro para realização do trabalho de campo e a empresa FotoTerra por gentilmente ceder as imagens hiperespectrais.

5. Referências bibliográficas

Baldeck, C. A.; Asner, G. P. Improving Remote Species Identification through Efficient Training Data Collection. **Remote Sensing**, v. 6, n. 4, p. 2682-2698, 2014.

Breiman, Leo. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.

Bruzzzone, L.; Roli, F.; Serpico, S. B. An extension of the Jeffreys-Matusita distance to multiclass cases for feature selection. **IEEE Transactions on Geoscience and Remote Sensing**, v. 33, n. 6, p. 1318-1321, 1995.

Dalponte, M.; Ørka, H. O.; Ene, L. T.; Gobakken, T.; Næsset, E. Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. **Remote sensing of environment**, v. 140, p. 306-317, 2014.

Féret, J.B.; Asner, G. P. Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy. **IEEE Transactions on Geoscience and Remote Sensing**, v. 51, n. 1, p. 73-84, 2013.

Ferreira, M. P.; Grondona, A. E. B.; Rolim, S. B.; Shimabukuro, Y. E. Analyzing the spectral variability of tropical tree species using hyperspectral feature selection and leaf optical modeling. **Journal of Applied Remote Sensing**, v. 7, 073502-1-073502-13, 2013.

Ghosh, A.; Fassnacht, F. E.; Joshi, P. K.; Koch, B. A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. **International Journal of Applied Earth Observation and Geoinformation**, v. 26, p. 49-63, 2013.

Guaratini, M. T. G.; Gomes, E. C. P.; Tamashiro, J. Y.; Rodrigues, R. R. Composição florística da Reserva Municipal de Santa Genebra, Campinas, SP. **Revista Brasileira de Botânica**, v. 31, n. 2, p. 323-337, 2008.

Hughes, G. F. On the mean accuracy of statistical pattern recognizers. **IEEE Transactions on Information Theory**, vol. IT-14, n. 1, p. 55-63, 1968.

Nagendra, H.; Rocchini, D. High resolution satellite imagery for tropical biodiversity studies: the devil is in the detail. **Biodiversity and Conservation**, v. 17, n. 14, p. 3431-3442, 2008.

Vapnik, V. N. **The Nature of Statistical Learning Theory**. New York: Springer-Verlag, 1995.

Veloso, H.P.; Rangel Filho, A.L.R.; Lima, J.C.A. **Classificação da vegetação brasileira, adaptada a um sistema universal**. IBGE, Rio de Janeiro, 1991.

Zanotta, D. C. **Uma metodologia adaptativa baseada em objetos para detecção semi-automática de desmatamento na Amazônia por imagens de sensoriamento remoto**. 2014. 139 p. Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE).