

ARTIGO ORIGINAL

Comparação entre processos de mapeamento da cobertura do solo urbano com ênfase na cobertura arbórea

Comparing mapping processes of urban land cover with emphasis in tree cover

Luciana Cavalcante Pereira¹ , Hilton Thadeu Zarate do Couto¹ 

¹Escola Superior de Agricultura "Luiz de Queiroz" – ESALQ, Universidade de São Paulo – USP, Piracicaba, SP, Brasil

Como citar: Pereira, L. C., & Couto, H. T. Z. (2020). Comparação entre processos de mapeamento da cobertura do solo urbano com ênfase na cobertura arbórea. *Scientia Forestalis*, 48(127), e3264. <https://doi.org/10.18671/scifor.v48n127.13>

Resumo

O mapeamento da cobertura do solo urbano pode ser executado por diferentes processos. Compararam-se seis processos em seis bairros de Rio Claro (São Paulo/ Brasil) quanto à concordância na quantificação dos diversos tipos de cobertura e quanto à acurácia na quantificação da cobertura arbórea. Os processos avaliados foram: *Dot grid*; Classificação Automática Supervisionada (CAS); CAS com uso do *NDVI*; Classificação Automática Não Supervisionada (CANS); CANS com uso do *NDVI*; e Vetorização aplicada à cobertura arbórea. Estes concordaram na quantificação dos tipos de cobertura com Kappa superior a 0,75. Avaliou-se a acurácia da quantificação da cobertura arbórea pelo teste de Tukey, não havendo diferenças significativas entre as médias dos processos em relação à Vetorização (dados de referência), exceto para CAS que apresentou média superestimada. A combinação do *NDVI* à imagem multiespectral aproximou as médias das classificações automáticas à da Vetorização, corrigindo inclusive a superestimativa mencionada. Avaliou-se o tempo de execução dos processos.

Palavras-chave: Mapeamento da cobertura do solo urbano; Acurácia da quantificação da cobertura arbórea; Concordância entre processos de mapeamento.

Abstract

Mapping urban land cover can be performed by different processes. Six processes were compared in six neighborhoods of Rio Claro (Sao Paulo/ Brazil) regarding the agreement on the quantification of the various types of coverage and the accuracy of tree cover quantification. The evaluated processes were: *Dot grid*; Supervised Automatic Classification (CAS); CAS with *NDVI*; Automatic Not Supervised Classification (CANS); CANS with *NDVI*; Vectorization on tree cover. These methods were compatible with the quantification of the types of coverage with Kappa above 0.75. The accuracy of the tree cover quantification was evaluated by Tukey test, and there were no significant differences between means of the processes in relation to the vectorization (reference data), except for CAS, which presented an overestimated mean. The combination of the *NDVI* with the multispectral image approximated means of automatic classifications to that of the vectorization, correcting even the overestimation mentioned. The execution time of the processes was evaluated.

Keywords: Urban land cover mapping; Accuracy of tree cover quantification; Agreement between mapping processes.

Fonte de financiamento: Fundação de Amparo à Pesquisa do Estado de São Paulo através do processo nº 2012/11828-7.

Conflito de interesse: Nada a declarar.

Autor correspondente: lcavalpereira@gmail.com

Recebido: 1 fevereiro 2019.

Aceito: 11 setembro 2019.

Editor: Paulo Henrique Müller Silva.



Este é um artigo publicado em acesso aberto (*Open Access*) sob a licença *Creative Commons Attribution*, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

INTRODUÇÃO

O mapeamento da cobertura do solo urbano fornece informações fundamentais para a compreensão da Ecologia Urbana. Esta cobertura pode ser separada em: áreas vegetadas - compostas por áreas de vegetação herbácea e pela cobertura arbórea (copa de árvores, arbustos e palmeiras); áreas impermeabilizadas - recobertas por asfalto, cimento e outros materiais dos pavimentos, assim como o solo exposto compactado e os telhados das construções; e reservatórios hídricos - como rios, represas, lagos e piscinas.

A proporção dos tipos de cobertura resulta em fenômenos ambientais que afetam significativamente a vida. Excessiva impermeabilização, por exemplo, pode levar ao surgimento de ilhas de calor e à ocorrência de enchentes e deslizamentos. Por outro lado, a cobertura arbórea realiza funções como interceptação da água das chuvas e diminuição do *run off*, redução da poluição da água, filtragem e umidificação do ar, amenização da temperatura, portanto pode ser devidamente alocada para proporcionar conforto, saúde e segurança. Também está associada à presença de diversas espécies, servindo-lhes de habitat e corredor para passagem.

Para o manejo da cobertura do solo urbano por meio de diagnósticos corretos e intervenções eficazes, torna-se necessária a utilização de técnicas que forneçam o mapeamento acurado, preferencialmente de baixo custo para possibilitar repetições que acompanhem o dinamismo da cidade (Miller, 1996).

Antecede o mapeamento da cobertura, a aquisição de uma imagem da área, que pode ser obtida no nível suborbital (fotografias aéreas e videografia) ou no nível orbital (imagens de satélite). A imagem é produto da captura por parte de sensores das ondas eletromagnéticas refletidas pelos objetos na superfície terrestre e é caracterizada pela faixa do espectro eletromagnético armazenada em cada pixel e pela resolução espacial, que expressa o nível de detalhamento do solo. No contexto urbano, a resolução espacial geralmente deve ser alta devido ao tamanho diminuto dos alvos (Jensen, 2009).

Na década de 60, difundiram-se os métodos vetoriais de mapeamento, nos quais os tipos de cobertura do solo eram separados por fotointerpretação, utilizando-se pontos, linhas, e polígonos para demarcar e delimitar feições. A princípio, eram executados em papel e aplicados principalmente a áreas rurais e naturais. Para áreas urbanas, devido à alta heterogeneidade da cobertura que aumenta drasticamente a quantidade de feições e as classes de cobertura, foram adaptados métodos nos quais a forma e o tamanho dos polígonos eram pré-determinados e depois o tipo de cobertura identificado, como o *Dot grid Method*.

O *Dot grid Method* (Método da Grade de Pontos) envolve a fotointerpretação da cobertura encontrada sob pontos distribuídos sistematicamente numa imagem da área. A proporção de uma determinada cobertura corresponde ao número de pontos que cai sobre ela dividido pelo total de pontos na área. Nowak et al. (1996) analisaram as técnicas utilizadas para o levantamento de cobertura arbórea em 68 cidades norte-americanas e observaram que em 42 delas havia sido utilizado o *Dot grid*.

Na década de 80, a informática permitiu que os métodos vetoriais fossem executados com mais facilidade por meio dos sistemas de informação geográfica. Também possibilitou o surgimento dos métodos matriciais ou por *raster*, nos quais a interpretação digital ou automática de imagens multiespectrais é utilizada para identificação do padrão espectral dos pixels. Os primeiros *softwares* de geoprocessamento faziam apenas a interpretação digital, posteriormente possibilitaram a execução de métodos híbridos com entrada de informações de interpretação visual que aprimoraram muito o mapeamento (Moreira, 2011).

A Classificação Automática Não Supervisionada (CANS) é um processo em que o *software* separa os pixels da imagem em grupos por faixa de assinatura espectral e posteriormente o intérprete determina a qual classe de cobertura cada grupo pertence por meio da similaridade de distribuição com a imagem. A Classificação Automática Supervisionada (CAS) segue o caminho inverso da CANS: o intérprete informa ao *software* quais os padrões espectrais dos tipos de cobertura por meio de pequenas amostras da imagem e o *software* classifica a imagem pixel por pixel (Moreira, 2011).

Nas classificações automáticas, o confundimento entre classes de cobertura ocorre quando suas assinaturas espectrais são próximas, dessa forma quanto maior a resolução espectral da imagem mais precisa é a separação. Por isso são utilizadas imagens multispectrais, que contêm bandas na faixa visível do espectro eletromagnético (0,4 a 0,7 μm) e na faixa do infravermelho próximo (0,7 a 1,1 μm) e médio (1,1 a 3,0 μm). A separação acurada da vegetação de outros tipos de cobertura, como asfalto e rios, só é possível devido à utilização da banda do infravermelho, pois as folhas absorvem a radiação fotossinteticamente ativa, que abrange a faixa do vermelho e do azul e refletem passivamente a faixa do verde, porém possuem mecanismos internos que fazem o infravermelho ser intensamente refletido, diferenciando-as dos demais materiais (Carleer et al., 2005).

Com o intuito de ressaltar o comportamento espectral da vegetação foram criados índices baseados na álgebra de bandas, que podem ser combinados às imagens espectrais. O *Normalized Difference Vegetation Index (NDVI)*, um dos mais difundidos, utiliza a banda do vermelho (*R*), na qual a vegetação aparece relativamente escura devido à absorção para fotossíntese, e a banda do infravermelho próximo (*NIR*), onde aparece clara devido à reflexão ativa. Quanto mais folhas a planta tem e mais vigorosas, mais esses comprimentos de onda são afetados, viabilizando diferenciar inclusive vegetação herbácea de cobertura arbórea (Weier & Herring, 2000).

Neste artigo, processos de mapeamento do solo são comparados quanto à concordância na quantificação dos tipos de cobertura, à acurácia na quantificação da cobertura arbórea e ao tempo de execução. Os processos são: *Dot grid*; Classificação Automática Supervisionada; Classificação Automática Supervisionada com uso do *NDVI*; Classificação Automática Não Supervisionada; Classificação Automática Supervisionada com uso do *NDVI*; Vetorização, aplicada apenas à cobertura arbórea.

MATERIAL E MÉTODOS

Área de estudo

Compreende quatro bairros contíguos (3,17 km^2) do perímetro urbano de Rio Claro, São Paulo, delimitados de acordo com o zoneamento realizado pela Prefeitura Municipal em 2007 (Figura 1). O Centro de Rio Claro foi o maior bairro analisado (1,84 km^2) e foi dividido em três porções, também chamadas de bairro - Centro Norte (0,63 km^2), Centro Médio (0,71 km^2) e Centro Sul (0,50 km^2) - para que os processos de mapeamento fossem aplicados em áreas de tamanhos próximos. Alto do Santana (0,5 km^2) e Vila Operária (0,44 km^2) são bairros ao norte do Centro e Cidade Nova (0,39 km^2) a leste. A escolha dos bairros se baseou primeiramente na quantidade de cobertura arbórea, o Centro é visivelmente pouco arborizado em relação aos demais bairros. As particularidades são o lago do Parque Municipal localizado na Vila Operária e a maior quantidade de vegetação herbácea no bairro Cidade Nova.

Os pontos extremos da área estão situados entre as latitudes 22°23'16" S e 22°25'26" S e as longitudes 47°33'03" O e 47°33'35" O; a elevação média do terreno é de 610 m em relação ao nível do mar. Na região, o clima é do tipo Cwa de acordo com a classificação de Köppen, caracterizado por temperatura média anual mínima de 15,1 °C e máxima de 28,0 °C, e por precipitação média de 1.366,8 mm por ano. As chuvas de verão acumulam acima de 600 mm entre dezembro e fevereiro, e durante a estiagem de inverno entre junho e agosto, chove menos de 100 mm (Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura, 2013). Predominam Argissolos Vermelhos e Vermelhos-Amarelos e são encontrados Chernossolos e Neossolos Litólicos. A vegetação original é composta por Cerrado e Floresta Estacional Semidecidual.



Figura 1 - Área de estudo: À Esquerda. Localização do Município de Rio Claro no Estado de São Paulo (Brasil); Ao Centro. Destaque em cinza da área do Perímetro Urbano no Município de Rio Claro; À Direita. Bairros analisados: Alto do Santana (contorno em azul), Vila Operária (em amarelo), Centro Norte, Centro Médio e Centro Sul (em verde), Cidade Nova (em vermelho)

Imagem de satélite da área de estudo

Foi adquirida uma imagem multiespectral do satélite WorldView II (resolução espacial de 0,5 m) composta pelas bandas do vermelho (*R*), verde (*G*), azul (*B*) e infravermelho próximo (*NIR*), fusionada, georreferenciada para *WGS 1984 datum* e sistema de coordenadas *Universal Transversal Mercator (UTM)* Zona 23 Sul, e ortorretificada. Foi capturada em 27 de julho de 2011, durante a estiagem de inverno, quando há menor oclusão pelas nuvens e a vegetação herbácea está menos vigorosa (exceto em áreas irrigadas).

Foi combinada à imagem multiespectral uma banda *NDVI* obtida por meio da Calculadora *Raster* do *software* gratuito Quantum GIS (1.8.0) (Equação 1):

$$NDVI = \frac{(NIR - VIS)}{(NIR + VIS)} \quad (1)$$

Onde *VIS* é a banda do vermelho (*R*) e *NIR* a banda do infravermelho próximo. Os valores de *NDVI* variam de -1 a 1, quanto mais próximo de 1, mais a *VIS* é refletida e não há vegetação, quanto mais próximo de +1 a *NIR* é refletida e há vegetação (National Aeronautics and Space Administration, 2000).

Mapeamento da cobertura do solo urbano

Para mapeamento da cobertura do solo urbano foram estabelecidas 12 classes de cobertura: cobertura arbórea, vegetação herbácea, lago, piscina, asfalto, cimento, solo exposto, telha cerâmica, telha cinza, telha metálica, sombra, outros (Silva Filho et al., 2005). Foram executados 6 processos de mapeamento do solo: (1) Vetorização, aplicada apenas à cobertura arbórea; (2) *Dot grid*; (3) Classificação Automática Supervisionada utilizando a imagem multiespectral (CAS); (4) Classificação Automática Supervisionada utilizando a imagem multiespectral combinada ao *NDVI* (CAS_*NDVI*); (5) Classificação Automática Não Supervisionada utilizando a imagem multiespectral (CANS); (6) Classificação Automática Não Supervisionada utilizando a imagem multiespectral combinada ao *NDVI* (CANS_*NDVI*). Para cada processo, foram calculadas as porcentagens dos tipos de cobertura e registrado o tempo de execução por bairro. Abaixo segue a descrição básica da execução dos processos:

Vetorização: executada no *software* Quantum GIS (1.8.0). Para cada bairro foi estabelecido um projeto composto pela camada *raster* da imagem com exibição das bandas *R*, *G*, *B* (similar a uma fotografia aérea colorida), sobre a qual foi colocada uma camada vetorial. Nesta foram delimitados os polígonos da cobertura arbórea com as ferramentas de edição do *software*.

Dot grid: executado com auxílio do Quantum GIS (1.8.0), no qual foi construído um projeto para cada bairro formado por 3 camadas: i. a camada *raster* da imagem Worldview II multiespectral com exibição das bandas *R*, *G*, *B*; ii. uma grade vetorial composta por polígonos de 10 x 10 m; iii. uma camada de pontos regulares dispostos no centro dos polígonos da grade vetorial. Cada polígono da grade vetorial foi classificado de acordo com a cobertura do solo demarcada pelo ponto ao centro.

Classificação Automática Supervisionada: executada no *software* gratuito Multispec (3.3), utilizando-se a imagem multiespectral e a imagem multiespectral combinada ao *NDVI*. Optou-se pela exibição das bandas *NIR*, *G*, *B* da imagem, na qual a vegetação é realçada pela coloração avermelhada (falsa cor), para visualização da área e seleção das amostras das classes de cobertura. No Multispec, podem ser fornecidos dois conjuntos de amostras chamados de treinador e de auditor. O auditor pode ser entendido como um segundo treinador, por meio do qual o *software* reclassifica os pixels da imagem e compara os resultados com aqueles do treinador. Foram fornecidas 5 amostras de cada classe de cobertura para o treinador e 5 para o auditor. O tamanho das amostras foi de 3 x 3 pixels, porém quando a classe se apresentou muito fragmentada, foram selecionadas amostras menores. Os mesmos conjuntos de amostras foram utilizados para a classificação da imagem sem e com *NDVI*, por meio do algoritmo *Maximum Likelihood*. As estatísticas do *software* a partir do treinador e auditor foram utilizadas como parâmetro de qualidade, de forma que a classificação foi considerada satisfatória quando a Acurácia Geral e Estatística Kappa foram superiores a 80% e as acurácias das classes superiores a 50%. A comparação visual com a imagem também foi utilizada como indicativo da qualidade.

Classificação Automática Não Supervisionada: executada no Multispec (3.3). Por meio do algoritmo *Isodata*, foram separados 24 *clusters* ou grupos de pixels por faixas de assinatura espectral. A assinatura espectral foi composta *a priori* pelas camadas *R*, *G*, *B* e *NIR* e posteriormente pelas camadas *R*, *G*, *B* e *NIR* combinadas ao *NDVI* (*CANS* e *CANS_NDVI*, respectivamente). Os *clusters* foram classificados por identificação visual de acordo com o tipo de cobertura de maior ocorrência e quando apresentavam duas classes ocorrendo paralelamente, o número de pixels do *cluster* foi dividido igualmente entre elas.

Uma pessoa iniciante em geoprocessamento executou os processos para se avaliar a facilidade de assimilação e aplicação. O tempo de execução não foi registrado para a *CAS_NDVI* por utilizar os mesmos treinadores e auditores da *CAS*, e para *CANS_NDVI*, pois seria influenciado pela experiência adquirida pelo usuário ao executar a *CANS*. Obtiveram-se o tempo médio, mínimo e máximo, intervalo de confiança com 99% de probabilidade e coeficiente de variação. A área dos bairros transformada para 0,5 km² e os tempos proporcionais utilizados para comparar *Dot grid* e da Vetorização.

Nas classificações automáticas, foram observados os principais confundimentos entre classes. Devido à dificuldade de se separar as classes piscina de telha metálica, excluiu-se piscina, e de se separar solo exposto de telha cerâmica, separou-se solo exposto apenas em parte (Pereira-Rollo et al., 2012).

Análise estatística

Comparação dos processos envolvendo todas as classes de cobertura

Na comparação entre os processos de mapeamento envolvendo todas as classes de cobertura, a Vetorização não foi incluída por abranger apenas a cobertura arbórea. Aplicou-se a Estatística Kappa (*K*) estabelecida por Cohen (1960) para avaliar a concordância das porcentagens das classes obtidas entre pares de processos. Utilizou-se parte da matriz de erro (Tabela 1), na qual os totais marginais das linhas correspondem às porcentagens obtidas por um processo, os totais marginais das colunas às porcentagens de outro processo, e a diagonal às porcentagens que concordam entre os dois processos. Cálculo do Kappa (*K*):

$$K = \frac{P_o - P_a}{1 - P_a} \quad (2)$$

$$P_o = \frac{\sum_{i=1}^M n_{ii}}{N} \quad (3)$$

$$P_a = \frac{\sum_{i=1}^M l_i * c_i}{N^2} \quad (4)$$

Onde: P_o é a proporção de acordo observado (proporção da área que é classificada nas mesmas classes de cobertura pelos dois processos); P_a a proporção de acordo devida ao acaso; M , o número de classes presentes na matriz de erro; n_{ii} , a porcentagem da área categorizada na mesma classe (corresponde aos valores na diagonal da matriz); N , porcentagem da área contemplada pela matriz (no caso 100%); l_i e c_i são as porcentagens de uma mesma classe obtidas pelos dois processos.

Tabela 1 - Exemplo de matriz de erro acompanhada dos cálculos de acurácia e Estatística Kappa (Adaptado de Jensen, 2005)

Dados de Referência				
Classificação	Classe 1	Classe 2	Classe 3	Total da Linha
Classe 1	8	1	3	12
Classe 2	2	9	0	11
Classe 3	0	0	7	7
Total da Coluna	10	10	10	30

$Acurácia\ Geral = (8 + 9 + 7) / 30 = 80\%$

<p>Acurácia do Produtor (<i>Erro de Omissão</i>)</p> <p>Classe 1 = $8/10 = 80\%$ (20%)</p> <p>Classe 2 = $9/10 = 90\%$ (10%)</p> <p>Classe 3 = $7/10 = 70\%$ (30%)</p>	<p>Acurácia do Usuário (<i>Erro de Comissão</i>)</p> <p>Classe 1 = $8/12 = 67\%$ (33%)</p> <p>Classe 2 = $9/11 = 82\%$ (18%)</p> <p>Classe 3 = $7/7 = 100\%$ (0%)</p>
---	--

Cálculo da Estatística Kappa (K)

$$P_o = \frac{(8 + 9 + 7)}{30} = 0,8 \quad P_a = \frac{[(10 * 12) + (10 * 11) + (10 * 7)]}{30^2} = 0,33$$

$$K = \frac{P_o - P_a}{1 - P_a} = \frac{0,8 - 0,33}{1 - 0,33} = 0,7 = 70\%$$

Comparação dos processos envolvendo apenas cobertura arbórea

Para comparação entre os processos quanto à quantificação da cobertura arbórea, cada bairro foi compreendido como uma repetição. A Análise de Variância (Fisher, 1918) foi utilizada para testar a hipótese nula de que não houve diferença entre os processos de mapeamento na obtenção das porcentagens de cobertura arbórea ($H_0: \bar{X}_{\text{vetor}} = \bar{X}_{\text{Dot grid}} = \bar{X}_{\text{cas}} = \bar{X}_{\text{cas_NDVI}} = \bar{X}_{\text{cans}} = \bar{X}_{\text{cans_NDVI}}$). Para que este teste estatístico seja usado, é necessário que: (1) as amostras sejam obtidas aleatoriamente; (2) os valores da variável tenham distribuição normal; (3) as variâncias dos processos de mapeamento sejam homogêneas. O primeiro pressuposto não foi observado, uma vez que os bairros foram pré-selecionados conforme a ocorrência de cobertura arbórea. Como consequência, o bairro foi incluído como uma das fontes de variação no cálculo do F (Tabela 2). O teste Box-Cox (Box & Cox, 1964) foi utilizado para indicar se o segundo e o terceiro pressupostos foram cumpridos ou se seria necessária uma transformação dos dados. No caso, os dados precisaram passar pela transformação

logarítmica ($X' = \ln X$) para apresentarem distribuição normal e homocedasticidade das variâncias.

Tabela 2 - Equações do **Quadro da** Análise de Variância para variável Cobertura Arbórea

Fontes da Variação	GL	SQ	QM	F calculado
Processo	$(w - 1)$	$\sum_{i=1}^k l_i * (\bar{X}_i - \bar{X})^2$	$SQ_{\text{processo}} / GL_{\text{processo}}$	$(QM_{\text{processo}} + QM_{\text{bairro}}) / QM_{\text{res}}$
Bairro	$(n - 1)$	$\sum_{i=1}^n k_i * (\bar{X}_i - \bar{X})^2$	$SQ_{\text{bairro}} / GL_{\text{bairro}}$	
Resíduo	$GL_{\text{total}} - (GL_{\text{processo}} + GL_{\text{bairro}})$	$SQ_{\text{total}} - (SQ_{\text{processo}} + SQ_{\text{bairro}})$	$SQ_{\text{res}} / GL_{\text{res}}$	
Total	$N - 1$	$\sum_{i=1}^k \sum_{j=1}^{l_i} (X_{ij} - \bar{X})^2$		

GL: graus de liberdade; SQ: soma dos quadrados das diferenças; QM: quadrado médio; w: número de processos (6); n: número de repetições, ou seja, de bairros (6); N: número de observações ($n * w = 36$); l: processo observado; k: bairro observado; i: contador

Pela Análise de Variância, observou-se que houve diferença entre os processos para obtenção da cobertura arbórea (rejeitou-se H_0), pois F calculado foi maior que o valor crítico de F igual a 5,03 para $\alpha = 5\%$, $GL_{\text{processo}} + GL_{\text{bairro}} = 10$, $GL_{\text{resíduo}} = 25$. Na sequência, o teste de Tukey (Zar, 2010) foi utilizado para determinar entre quais pares de processos de mapeamento não houve diferença significativa ($H_0: \bar{X}_B = \bar{X}_A$). Para tanto, as médias dos processos foram colocadas em ordem decrescente e foi calculada a diferença entre pares ($\bar{X}_B - \bar{X}_A$). Em seguida, calculou-se o erro padrão (EP):

$$EP = \sqrt{\frac{s^2}{n}} \quad (5)$$

Onde: s^2 é a variância do resíduo ou QM_{res} ; n é o número de repetições ou bairros em cada processo (6).

E calculou-se a diferença mínima significativa (Δ) a partir da amplitude estudentizada ou q tabelado:

$$\Delta = q * \sqrt{EP} \quad (6)$$

Onde: q é igual a 4,358 (para $\alpha = 5\%$, $GL_{\text{resíduo}} = 25$, número de processos = 6).

Rejeitou-se H_0 para os pares de processos cuja diferença entre médias foi superior à diferença mínima significativa. A Vetorização foi utilizada como processo de referência para avaliação da acurácia dos demais processos por ser puramente baseada na fotointerpretação e na delimitação dos limites das classes.

RESULTADOS E DISCUSSÃO

Concordância entre processos de mapeamento (Estatística Kappa)

A Estatística Kappa foi utilizada para avaliar a concordância entre processos de mapeamento abordando a equivalência entre eles na obtenção de porcentagem das classes de cobertura. Na matriz de erro foram incluídos os totais de porcentagem de classe de dois processos e as porcentagens de coincidência entre eles a partir de toda área dos bairros, em detrimento da verificação em campo de unidades amostrais e da localização dos erros de

omissão e comissão (Congalton, 1991; Jensen, 2005). Assim informações a respeito de 5 processos de mapeamento executados em 6 bairros, portanto 60 pares de processos, puderam ser sumarizadas em uma única tabela (Tabela 3).

Tabela 3 - Estatística Kappa (K) obtida entre processos de mapeamento do solo aplicados em 6 bairros de Rio Claro (SP), sendo o sistema de classificação: cobertura arbórea, vegetação herbácea, lago, piscina, asfalto, cimento, solo exposto, telha cerâmica, telha cinza, telha metálica, sombra e outros

	Centro Norte	Centro Médio	Centro Sul	Alto do Santana	Cidade Nova	Vila Olímpia
<i>Dot grid</i> X CAS	0,8712	0,8362	0,8759	0,8675	0,9057	0,8816
<i>Dot grid</i> X CAS_NDVI	0,8665	0,8380	0,8666	0,8392	0,8877	0,9028
<i>Dot grid</i> X CANS	0,8231	0,8481	0,7648	0,8108	0,7613	0,8637
<i>Dot grid</i> X CANS_NDVI	0,8162	0,8064	0,7853	0,8293	0,8226	0,8130
CAS X CAS_NDVI	0,8974	0,9967	0,9606	0,9664	0,9776	0,9636
CAS X CANS	0,8611	0,8443	0,8080	0,7508	0,7821	0,7798
CAS X CANS_NDVI	0,8346	0,7944	0,8594	0,7976	0,8570	0,7957
CAS_NDVI X CANS	0,8222	0,8462	0,8291	0,7528	0,8024	0,8007
CAS_NDVI X CANS_NDVI	0,7880	0,7938	0,8564	0,7989	0,8793	0,7700
CANS X CANS_NDVI	0,9463	0,8482	0,8515	0,8465	0,8700	0,8232

CAS: Classificação Automática Supervisionada; CAS_NDVI: Classificação Automática Supervisionada com uso do NDVI; CANS: Classificação Automática Não Supervisionada; CANS_NDVI: Classificação Automática Não Supervisionada com uso do NDVI

Observou-se $0,8007 \leq K \leq 0,9967$ entre 46 pares (76,7%), nos 14 pares restantes (23,3%) $0,7508 \leq K \leq 0,7989$. Os pares formados por *Dot grid*, CAS ou CAS_NDVI apresentaram $K > 0,8$ em todos os bairros, assim como os pares de CANS com CANS_NDVI. No entanto, quando CANS e CANS_NDVI foram pareados com *Dot grid*, CAS e CAS_NDVI, obteve-se $K > 0,8$ entre 22 pares (61,1%) e $0,7508 \leq K \leq 0,7989$ entre 14 pares (38,9%) (Tabela 3).

Divisões arbitrárias dos valores de Kappa têm sido estabelecidas para padronizar as avaliações de concordância. Fleiss (2003), pesquisador do equacionamento da Estatística Kappa, estabeleceu a seguinte divisão: $K > 0,75$ - concordância excelente; $0,40 < K \leq 0,75$ - concordância satisfatória à boa; $K \leq 0,40$ - concordância pobre. Landis & Koch (1977) sugeriram uma divisão mais rigorosa que se tornou popular: $0,8 < K < 1$ - concordância quase perfeita; $0,6 < K \leq 0,80$ - concordância substancial; $0 < K \leq 0,60$ - de pobre a moderada. Segundo Fleiss (2003), a concordância foi excelente entre todos os pares de processo. Para Landis & Koch (1977), a concordância foi quase perfeita entre 46 pares de processo (76,7%) e substancial entre os 14 pares restantes (23,3%). Como enfatizado, os valores de Kappa abaixo de 0,8 se concentraram nos pares de processo em que um dos constituintes era CANS ou CANS_NDVI e o outro *Dot grid*, CAS ou CAS_NDVI, especificamente em 38,9% dos pares assim configurados. De maneira generalizada, os processos podem ser considerados equivalentes na obtenção das porcentagens de classes de cobertura, apesar de as classificações automáticas não supervisionadas poderem apresentar menor concordância com os demais métodos.

Observa-se que a discordância esteve presente entre todos os pares de processo ($K \neq 1$), pois está associada a características dos processos de mapeamento dentre outros fatores (Figura 2). Neste sentido, o *Dot grid* tem como fonte de erro a extrapolação da classe sob o ponto para toda área da quadrícula, e é claro que quanto menor a quadrícula e a fragmentação da classe de cobertura, menor deve ser o erro. A presença periódica de uma determinada classe (por exemplo, o asfalto) também pode levar a erro, porém uma forma de evitá-lo é a aleatorização da posição do ponto na quadrícula (Nowak et al., 1996) (Figura 2b). Nas classificações automáticas, o refinamento da resolução espacial aumenta a variabilidade da reflectância dentro das unidades de cobertura do solo, que por sua vez reduz a separabilidade dentro do espaço espectral disponível. A diminuição da separabilidade tende

a reduzir a acurácia das classificações pixel a pixel, inclusive do algoritmo *Maximum Likelihood* (Carleer et al., 2005). No entanto, Hester et al. (2008) demonstraram que elas podem gerar mapas altamente precisos, apesar das sobreposições espectrais entre classes.

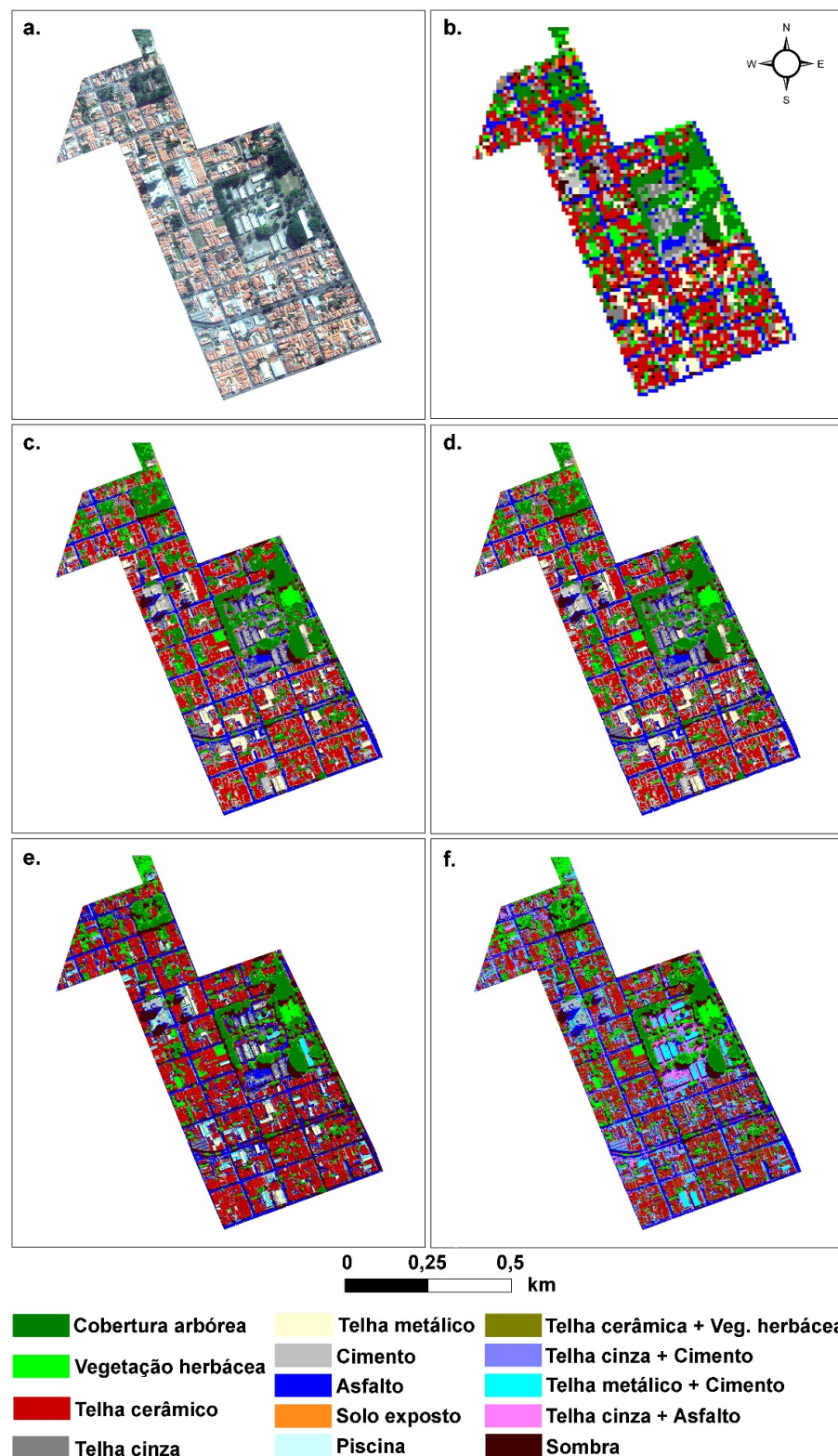


Figura 2 – Mapeamento do solo urbano no bairro Alto do Santana (Rio Claro – SP, Brazil): a. Imagem Worldview II (0,5 m de resolução); b. *Dot grid*; c. Classificação Automática Supervisionada (CAS); d. Classificação Automática Supervisionada com uso do NDVI (CAS_NDVI); e. Classificação Automática Não Supervisionada (CANS); f. Classificação Automática Não Supervisionada com uso do NDVI (CANS_NDVI)

Confundimentos devido às sobreposições espectrais são reportados em literatura, como entre solo exposto e superfícies impermeáveis e entre tipos de vegetação (Thomas et al., 2003; Hester et al., 2008). É claro que se as classes que se sobrepõem ocorrem em maior proporção, especificamente dentro da faixa espectral da sobreposição, o confundimento será mais intenso, sobressaindo o efeito sal pimenta (mistura de pixels de classes diferentes). Tratando-se da classificação automática supervisionada, em que as faixas espectrais de cada classe são informadas pelo usuário, pode haver situações em que seja possível separar uma das classes apenas em parte deixando os pixels da sobreposição para a outra, como optou-se em fazer com o solo exposto em relação à telha cerâmica, ou simplesmente pode-se excluir uma delas se ocorrer em pequena proporção e seus pixels serão incorporados pela outra classe, caso da classe piscina em relação à telha metálica (Figura 2c, 2d).

Por outro lado, nas classificações automáticas não supervisionadas, os *clusters* são estabelecidos a partir de faixas espectrais pré-definidas de amplitudes iguais que acabam por reunir classes diferentes com assinaturas espectrais próximas. Consequentemente, os confundimentos podem ser mais pronunciados, como o observado entre telha metálica e cimento, por as classes ocorrerem concomitantemente em alguns *clusters*, assim optou-se por dividir os pixels entre elas (Figura 2e, 2f). O procedimento convencional é aumentar o número de *cluster*, porém apenas dividir os pixels do *cluster* entre as classes foi muito mais prático. Como a cobertura do solo urbano é muito heterogênea, mais *clusters* podem aumentar a dificuldade do usuário em determinar as classes ao invés de resolver a confusão entre elas. Certamente os confundimentos terão importância de acordo com os objetivos do levantamento, sendo assim, supondo-se que se deseja estimar a quantidade e distribuição da cobertura arbórea, confundimentos que não envolvam esta classe serão pouco relevantes.

Diversas alternativas para lidar com a lacuna de resolução espectral que leva às sobreposições de classes nas imagens de alta resolução espacial e aos confundimentos nas classificações estão em desenvolvimento. Associações de técnicas podem ser utilizadas, como por exemplo, vetorização de um tipo de cobertura seguida de classificação automática do restante da imagem, uso paralelo de classificação automática e lidar, dentre outras (Hester et al., 2008; Tooke et al., 2009). Análises orientadas a objeto, que incorporam informações sobre textura, forma, contexto, além do padrão espectral, têm se mostrado consistentes nos levantamentos de cobertura do solo, porém são técnicas ainda em estudo e disponíveis a alto-custo (Trimble, 2013). Pode-se optar pelo uso de filtros que suavizam a heterogeneidade de reflectância dos objetos na imagem (Hall, 2007). E há também os índices baseados na álgebra de bandas que podem ampliar e realçar informações espectrais, e consequentemente, melhorar as classificações automáticas (Zha et al., 2003; Sawaya et al., 2003).

Acurácia da quantificação da cobertura arbórea

Os processos de mapeamento a cerca apenas da cobertura arbórea (%) foram comparados por Análise de Variância. F calculado foi igual a 35,55, superior ao valor crítico de F ($F_{5\%,10,25} = 5,03$; valor-p < 0,001), portanto rejeitou-se a hipótese nula de que não havia diferença significativa entre as médias de cobertura arbórea obtidas pelos processos de mapeamento do solo ($H_0: \bar{X}_{Dot\ grid} = \bar{X}_{cas} = \bar{X}_{cas_NDVI} = \bar{X}_{cans} = \bar{X}_{cans_NDVI} = \bar{X}_{vetor}$) (Tabela 4).

Tabela 4 - Quadro da Análise de Variância para variável Cobertura Arbórea (%)

Fontes da Variação	GL	SQ	QM	F calculado	Valor-p
Processo	5	0,893	0,178	37,55	<0,001
Bairro	5	5,407	1,082		
Resíduo	25	0,419	0,017		
Total	35	6,720			

GL: graus de liberdade; SQ: soma dos quadrados das diferenças; QM: quadrado médio

A partir do teste de Tukey, avaliou-se a diferença entre as médias de cobertura arbórea dos processos de mapeamento, sendo a diferença mínima significativa (Δ) igual a 0,2305. Assim formaram-se 4 grupos de médias e apenas a CAS apresentou diferença significativa em relação à Vetorização (Tabela 5).

Tabela 5 - Grupos de Média de Tukey para a variável Cobertura Arbórea (%)

Grupos de Média de Tukey			Média	Processo
	A		2.57605	CAS
B	A		2.48773	<i>Dot grid</i>
B	A	C	2.45749	CAS_NDVI
B	D	C	2.30420	Vetorização*
	D	C	2.24076	CANS_NDVI
	D		2.11690	CANS

CAS: Classificação Automática Supervisionada; CAS_NDVI: Classificação Automática Supervisionada com uso do NDVI; CANS: Classificação Automática Não Supervisionada; CANS_NDVI: Classificação

A partir da pressuposição tradicionalmente aceita de que a fotointerpretação é correta (Congalton, 1991), a Vetorização da cobertura arbórea pode servir como dado de referência para aferir a consistência da quantificação da cobertura arbórea pelos demais processos. Ainda se baseando naquela pressuposição, fica explicado porque o *Dot grid* quantificou acuradamente a cobertura arbórea, não se diferenciando significativamente da Vetorização, apesar da extrapolação da classe sob o ponto para toda área da quadrícula.

Nas classificações automáticas, com a incorporação da banda do infravermelho próximo nas imagens de sensoriamento remoto, tornou-se mais precisa a separação da cobertura vegetal dos demais tipos de cobertura do solo. No entanto, a sobreposição espectral entre classes de cobertura vegetal permaneceu sendo um desafio. Neste experimento, ocorreram confundimentos principalmente entre vegetação herbácea, árvores isoladas e árvores da periferia dos aglomerados, que podem ser explicados pelo menor volume de folhas das árvores em ambas as condições e ao maior vigor da vegetação herbácea em algumas áreas, resultando em uma reflectância do infravermelho próximo semelhante entre elas. Por isso a média da CAS acabou se afastando da média da Vetorização e se apresentou superestimada, no entanto foi semelhante à média do *Dot grid* e da CAS_NDVI.

A adição da banda do NDVI à imagem multiespectral se mostra vantajosa por possibilitar ampliar e enfatizar informações espectrais de forma simples e sem custos adicionais. No caso, a utilização do NDVI corrigiu a superestimativa da CAS e melhorou o desempenho da CANS, a ponto de apresentar a média mais próxima à da Vetorização que o próprio *Dot grid*.

Tempo de execução dos processos de mapeamento

Tradicionalmente a fotointerpretação é aceita como correta (Congalton, 1991), no entanto, quanto mais o método depende dela, mais tempo é necessário para sua execução. À vista disso, a interpretação digital utilizada nos processos híbridos proporciona a redução desse tempo, porém imagens multiespectrais de alto custo são necessárias para a separação acurada das classes.

Observou-se que o *Dot grid* se mostrou moroso (\bar{x} = 13,18 h/bairro), enquanto as classificações automáticas foram executadas em curto espaço de tempo (\bar{x} = 1,75 h/bairro para CAS; \bar{x} = 1,04 h/bairro para CANS). O alto valor do CV da CAS em relação à CANS (61,18% e 18,60%, respectivamente) demonstra que a primeira requer maior tempo para assimilação pelo usuário, principalmente na capacidade de seleção de amostras das classes. Com certo treino, a CAS pode ser executada em períodos tão curtos quanto à CANS (Mín.= 0,75 h para CAS; Mín.= 0,73 h para CANS) (Tabela 6).

Tabela 6 - Estatísticas do tempo de execução dos processos de mapeamento da cobertura do solo urbano (os tempos da Vetorização e do *Dot grid* foram estabelecidos a partir da aproximação da área dos bairros para 0,5 km²)

Processo de Mapeamento	\bar{x} (h)	Mín. (h)	Máx. (h)	IC (99%)	CV (%)
Vetorização*	3,50	1,91	4,66	3,50 \pm 2,21	30,75
<i>Dot grid</i> *	1,89	0,88	2,81	1,98 \pm 1,89	46,55
<i>Dot grid</i>	13,18	11,23	14,56	13,18 \pm 3,06	11,28
CAS	1,75	0,75	3,50	1,75 \pm 1,77	61,38
CANS	1,04	0,73	1,10	1,04 \pm 0,32	18,60

Onde: \bar{x} : média; Mín.: valor mínimo observado; Máx.: valor máximo observado; IC: Intervalo de Confiança; CV: Coeficiente de Variação; CAS: Classificação Automática Supervisionada; CANS: Classificação Automática Não Supervisionada; * Aplicada (o) apenas à cobertura arbórea

Comparando-se Vetorização e *Dot grid* aplicados apenas à cobertura arbórea, observa-se que o tempo de execução médio do primeiro processo foi de 3,5 h/bairro enquanto do segundo foi 1,89 h/bairro, pois a Vetorização requer que o usuário desenhe cada polígono. No *Dot grid* basta selecionar as quadrículas que representam a classe. O CV para ambos foi alto (30,75% e 46,55%, respectivamente), devido principalmente às diferenças na quantidade de cobertura arbórea entre bairros (Tabela 6).

CONCLUSÃO

Os processos de mapeamento aplicados à cobertura do solo urbano se mostraram equivalentes na quantificação das diferentes classes de cobertura de acordo com a Estatística Kappa (K). Os pares formados por *Dot grid* e classificações automáticas supervisionadas apresentaram maior concordância ($K > 0,8$) que os pares formados por um desses processos com as classificações automáticas não supervisionadas ($K > 0,75$). Vale ressaltar que as classificações automáticas não supervisionadas foram executadas de forma diferenciada da convencional, quando se dividiu o número de pixels entre classes que ocorreram concomitantemente no *cluster*.

Quando os processos foram avaliados apenas quanto à quantificação da cobertura arbórea, tendo a Vetorização como dado de referência, apenas a média da Classificação Automática Supervisionada se mostrou superestimada pelo teste de Tukey. As médias da classificação automática supervisionada e não supervisionada se aproximaram da média da Vetorização quando o *NDVI* foi combinado à imagem multiespectral, inclusive proporcionando a correção da superestimativa. Portanto o *NDVI* foi útil para ampliar e realçar as informações espectrais da imagem, melhorando a acurácia das classificações automáticas.

Na decisão pelo processo de mapeamento a ser aplicado, deve-se levar em consideração o tempo, os recursos financeiros disponíveis e o custo para execução. Nas áreas urbanas estudadas de aproximadamente 0,5 km², os métodos de fotointerpretação (*Dot grid* e Vetorização) demandaram maior tempo de execução que os métodos híbridos. Porém podem ser executados com imagens de menor custo, por exemplo, fotografias aéreas digitalizadas, mesmo em preto e branco. Já os processos híbridos dependem de imagens multiespectrais com a banda do infravermelho próximo, que são obtidas com tecnologias ainda onerosas, para fornecer dados precisos.

Os *softwares* gratuitos apresentaram bom desempenho durante sua utilização, porém sua eficiência poderá ser diferente para aplicação em imagens de áreas maiores que 0,70 km².

AGRADECIMENTO

Agradecemos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento.

REFERÊNCIA BIBLIOGRÁFICA

- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformation. *Journal of the Royal Statistical Society. Series A (General)*, 26(2), 211-252.
- Carleer, A. P., Debeir, O., & Wolff, E. (2005). Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering and Remote Sensing*, 71(11), 1285-1294. <http://dx.doi.org/10.14358/PERS.71.11.1285>.
- Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura – CEPAGRI. (2013). *Clima dos municípios paulistas: Rio Claro*. Campinas. Recuperado em 28 de setembro de 2013, de http://www.cpa.unicamp.br/outras-informacoes/clima_muni_494.html
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. <http://dx.doi.org/10.1177/001316446002000104>.
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35-46. [http://dx.doi.org/10.1016/0034-4257\(91\)90048-B](http://dx.doi.org/10.1016/0034-4257(91)90048-B).
- Fisher, R. A. (1918). The correlation between on the supposition of Mendelian inheritance. *Philosophical Transactions of the Royal Society of London*, 52, 399-433.
- Fleiss, J. (2003). *The measurement of interrater agreement: statistical methods for rates and proportions* (3rd ed., Chap. 18, pp. 598-626). New York: John Wiley. <http://dx.doi.org/10.1002/0471445428.ch18>.
- Hall, M. (2007). Smooth operator: smoothing seismic interpretations and attributes. *The Leading Edge*, 26(1), 16-20. <http://dx.doi.org/10.1190/1.2431821>.
- Hester, D. B., Cakir, H. I., Nelson, S. A. C., & Khorram, S. (2008). Per-pixel classification of high spatial resolution satellite imagery for urban land-cover mapping. *Photogrammetric Engineering and Remote Sensing*, 74(4), 463-471. <http://dx.doi.org/10.14358/PERS.74.4.463>.
- Jensen, R. J. (2005). Land-use classification map accuracy assessment. In R. J. Jensen (Ed.), *Introductory digital image processing: a remote sensing perspective* (2nd ed., Chap. 8, pp. 247-256). Upper Saddle River: Prentice Hall.
- Jensen, R. J. (2009). Sensoriamento remoto da paisagem urbana. In R. J. Jensen, *Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres* (K. C. Clarke, trad., Cap. 13, pp. 445-510). São José dos Campos: Parênteses.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. PMID:843571. <http://dx.doi.org/10.2307/2529310>.
- Miller, R. W. (1996). *Urban forestry: planning and managing urban greenspaces* (2nd ed., 502 p.). New Jersey: Prentice-Hall.
- Moreira, M. A. (2011). *Fundamentos do sensoriamento remoto e metodologias de aplicação* (4. ed., 422 p.). Viçosa: UFV.
- National Aeronautics and Space Administration – NASA. (2000). *Measuring vegetation (NDVI & EVI)*. Recuperado em 6 de março de 2014, de <http://earthobservatory.nasa.gov/Features/MeasuringVegetation/>
- Nowak, D. J., Rowntree, R. A., McPherson, E. G., Sisinni, S. M., Kerkmann, E. R., & Stevens, J. C. (1996). Measuring and analyzing urban tree cover. *Landscape and Urban Planning*, 36(1), 49-57. [http://dx.doi.org/10.1016/S0169-2046\(96\)00324-6](http://dx.doi.org/10.1016/S0169-2046(96)00324-6).
- Pereira-Rollo, L. C., Vettorazzi, C. A., Polizel, J. L., Silva-Filho, D. F., & Couto, H. T. Z. (2012). Comparação entre dois processos de mapeamento da cobertura do solo urbano. In *Anais do 2º Congresso Brasileiro de Ecologia de Paisagens*. Piracicaba: USP. Recuperado em 28 setembro de 2013, de http://cmq.esalq.usp.br/wiki/lib/exe/fetch.php?media=publico:projetos:processos_mapeamento.pdf
- Sawaya, K. E., Olmanson, L. G., Heinert, N. J., Brezonik, P. L., & Bauer, M. E. (2003). Extending satellite remote sensing to local scales: land and water resource monitoring using high-resolution imagery. *Remote Sensing of Environment*, 88(1-2), 144-156. <http://dx.doi.org/10.1016/j.rse.2003.04.006>.
- Silva Filho, D. F., Pivetta, K. F. L., Couto, H. T. Z., & Polizel, J. L. (2005). Indicadores de floresta urbana a partir de imagens aéreas multiespectrais de alta resolução. *Scientia Forestalis*, (67), 88-100.
- Thomas, N., Hendrix, C., & Congalton, R. G. (2003). A comparison of urban mapping methods using high-resolution digital imagery. *Photogrammetric Engineering and Remote Sensing*, 69(9), 963-972. <http://dx.doi.org/10.14358/PERS.69.9.963>.

- Tooke, T. R., Coops, N. C., Goodwin, N. R., & Voogt, J. A. (2009). Extracting urban vegetation characteristic using spectral mixture analysis and decision tree classifications. *Remote Sensing of Environment*, 113(2), 398-407. <http://dx.doi.org/10.1016/j.rse.2008.10.005>.
- Trimble. (2013). *eCognition software*. Recuperado em 12 de novembro de 2013, de <http://www.ecognition.com/>
- Weier, J., & Herring, D. (2000). *Measuring vegetation*. NASA Earth Observatory. Recuperado em 28 de setembro de 2013, de <http://earthobservatory.nasa.gov/Features/MeasuringVegetation/>
- Zar, J. H. (2010). Multiple comparisons. In J. H. Zar (Ed.), *Biostatistical analysis* (5th ed., Chap. 10, pp. 226-244). New Jersey: Prentice-Hall.
- Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), 583-594. <http://dx.doi.org/10.1080/01431160304987>.

Contribuição dos Autores: LCP: conceptualization, funding acquisition, data curation, writing; PGL: conceptualization, funding acquisition, supervision, formal analysis.