

Comparação da exatidão de métodos de classificação supervisionada e não supervisionada a partir do índice kappa na microbacia do Ribeirão Duas Águas em Botucatu/SP

Danilo Fávero Tangerino¹
Rodrigo Tommasello Lourenço¹

Universidade Estadual Paulista – UNESP/FCA¹
Campus de Botucatu
Rua José Barbosa de Barros, nº 1780 Fazenda Lageado 18610-307
danilofaverot@yahoo.com.br
rtlourenço@yahoo.com.br

Abstract

This study aims to compare the accuracy between the methods of supervised classification MAXVER and unsupervised CLUSTER with the kappa index in the watershed of Ribeirão Duas Águas in Botucatu/SP. From that goal, we can understand what the best method to accomplish the classification of types of use in the area of study, as well as define the best procedure for the improvement on the environmental planning. It's been used the *software* Idrisi 15- *The Andes Edition* for this goal, and several papers about the theme, and the observation after the study is that, for this situation, the best method to do it is the supervised classification MAXVER, because of its best accuracy in the methodology of work and the intervention of the user to determinate the training sites of each class of land use.

1. Introdução

A apropriação da natureza pelo homem sempre foi uma condição para que o processo de fixação de comunidades fosse possível, assim como para que a sociedade se organizasse, de maneira que podemos considerar a natureza a pedra fundamental para os processos que vivemos atualmente no capitalismo. Como fonte de infindáveis recursos naturais, a natureza é alvo constante de criação e recriação proveniente da criatividade humana, sendo tida como a matéria-prima para quaisquer questões ou necessidades humanas.

Esse processo se intensificou com a Revolução Industrial no século XVIII, quando as máquinas movidas principalmente a vapor tomaram conta de vários processos antes realizados por seres humanos ou animais. Então surgem os problemas ambientais, que emergiram diante de um quadro insustentável de êxodo urbano, expansão urbana, e necessidade de cada vez mais matérias-primas, tudo isso sem a existência de um planejamento ambiental e urbano consistente e que visasse a conservação dos recursos naturais.

Essa exploração desordenada criou na sociedade uma necessidade de desenvolver tecnologias e métodos que viessem a auxiliar na “compreensão da dinâmica natural e ação sobre a natureza, de modo a mitigar e/ou recuperar os danos já causados pela ação humana” (DEMARCHI et al, 2011, p. 235). E dentre as tecnologias utilizadas atualmente para o manejo correto do uso da terra, o planejamento ambiental, delimitação de áreas de proteção ambiental, conservação tanto do solo quanto dos recursos hídricos, dentre outros, estão o Sensoriamento Remoto e o Geoprocessamento.

Segundo Jensen (2009), o termo *sensoriamento remoto* pode ser definido como a medição ou aquisição de informação de alguma propriedade de um objeto ou fenômeno, por um dispositivo de registro que não esteja em contato físico ou íntimo com o objeto ou fenômeno em estudo, podendo ser dividido em orbital, quando as informações são coletadas

por sensores localizados em órbitas ao redor do planeta, realizando a coleta de dados e informações sobre a superfície terrestre em intervalos de tempo e espaço previamente estabelecidos, cujos produtos principais são as imagens de satélite; ou sub-orbital, quando esse processo é realizado por equipamentos aerotransportados não localizados em órbitas, cujos produtos principais são as fotografias aéreas (PIROLI, 2010a)

Os produtos referentes às varreduras feitas, seja por sistemas orbitais ou sub-orbitais, são produtos brutos, com os quais, sem o tratamento, de nada servem para os objetivos descritos por Demarchi et al (2011). Portanto, há a necessidade de realizar um processamento das imagens digitais, a fim de prepará-las para utilização. Segundo Novo (1992), as técnicas para esse processo podem ser classificadas em três conjuntos: pre-processamento, realce e classificação. Dentre os conjuntos, o mais prático, referente à análise propriamente dita, é o conjunto das técnicas de classificação, que

visam o reconhecimento automático de objetos da cena a partir da análise quantitativa dos níveis de cinza. Os objetos são classificados em diferentes categorias em função de algum critério de decisão objetivo. (NOVO, 1992, p. 218)

De acordo com Rosa (2009), a classificação de imagens diz respeito à interpretação das mesmas com auxílio de um computador, tratando-se essencialmente, da atribuição de um significado específico a um pixel ou a um conjunto de pixels, em função de suas propriedades numéricas. Esse processo de classificação se caracteriza pela atribuição aos conjuntos de pixels que possuem características espectrais comuns entre si, uma assinatura espectral que nos diz que aquele conjunto de pixels representa um tipo específico de uso da terra, como água, pastagem, reflorestamento, área urbana etc.

As técnicas de classificação de imagens podem ser divididas em não-supervisionada e supervisionada. A primeira não requer qualquer informação sobre classes de interesse, somente examinando sem auxílio do analista, os agrupamentos espectrais do alvo e dividindo-os em classes. Esse tipo de classificação trabalha com uma lógica conhecida como análise de agrupamento (*clustering*), sendo útil quando não se possui informações sobre a área imageada, sendo as classes definidas no algoritmo de classificação (ROSA, 2009). Uma das técnicas de classificação não supervisionada é a opção CLUSTER, que usa uma técnica de seleção dos pixels pelos picos dos histogramas, e quando os picos foram identificados, todos os possíveis valores são associados ao pico mais próximo, e as classes tendem a cair no ponto médio entre os picos (EASTMAN, 2006).

A classificação supervisionada consiste na identificação prévia das classes de informação (tipos de cobertura do solo), chamadas áreas de treinamento, que nada mais são do que representações dos do comportamento médio das classes que serão mapeadas automaticamente (NOVO, 1992). Esse tipo de classificação é utilizada quando se possui algum conhecimento sobre as classes que devem ser representadas na imagem pelo computador. Uma das principais técnicas utilizadas para a realização de classificações supervisionadas é a MAXVER (máxima verossimilhança), que consiste na classificação da imagem ponto a ponto, a partir de classes fornecidas pelo usuário. (ROSA, 2009). Segundo Novo (1992), é importante que as amostras de treinamento (ou *training sites*) sejam bastante homogêneas e que representem as classes de interesse, ou o resultado pode ser inadequado ao esperado para a classificação. O MAXVER é um método “pixel a pixel”, que considera a ponderação das distâncias entre as médias dos níveis digitais das classes cujas amostras de treinamento foram definidas, utilizando parâmetros estatísticos e ajuste segundo uma distribuição gaussiana. (INPE, 2009)

Para que possamos identificar se as classificações realizadas, seja por técnicas não supervisionadas ou supervisionadas, devemos embatê-las com a chamada *verdade terrestre*,

que se caracteriza pela real utilização do solo na área de estudo. Esse processo é feito a partir de índices de acurácia, ou exatidão, que realizam uma comparação entre as matrizes de pixels da classificação temática e os da verdade terrestre. Segundo Rosenfield e Fitzpatrick-Lins (1986), dentre os diversos métodos existentes para essa avaliação, o índice Kappa é o mais recomendado por utilizar todas as células da matriz ao invés de somente os elementos diagonais, o que garante uma maior acurácia em relação aos outros métodos devido à medição da probabilidade de um pixel estar corretamente classificado, em relação à probabilidade de estar incorretamente classificado (DEMARCHI et al, 2010).

O índice Kappa possui intervalos determinados que caracterizam a acurácia das classificações realizadas em relação à verdade terrestre, determinados por Landis e Koch (1977) e adaptados por Piroli (2010b), apresentadas a seguir.

Valor de Kappa	Qualidade da classificação
<0,00	Péssima
0,00 – 0,20	Ruim
0,20 – 0,40	Razoável
0,40 – 0,60	Boa
0,60 – 0,80	Muito boa
0,80 – 1,00	Excelente

Todos estes conceitos baseiam nossa abordagem em relação ao tema proposto, que tem como objetivo a comparação entre os métodos de classificação supervisionada Máxima Verossimilhança (MAXVER) e a classificação não supervisionada CLUSTER, embatendo-os com a verdade terrestre para a estimativa do índice Kappa, e analisar qual dos dois métodos é mais eficaz na representação dos usos presentes na área de estudo.

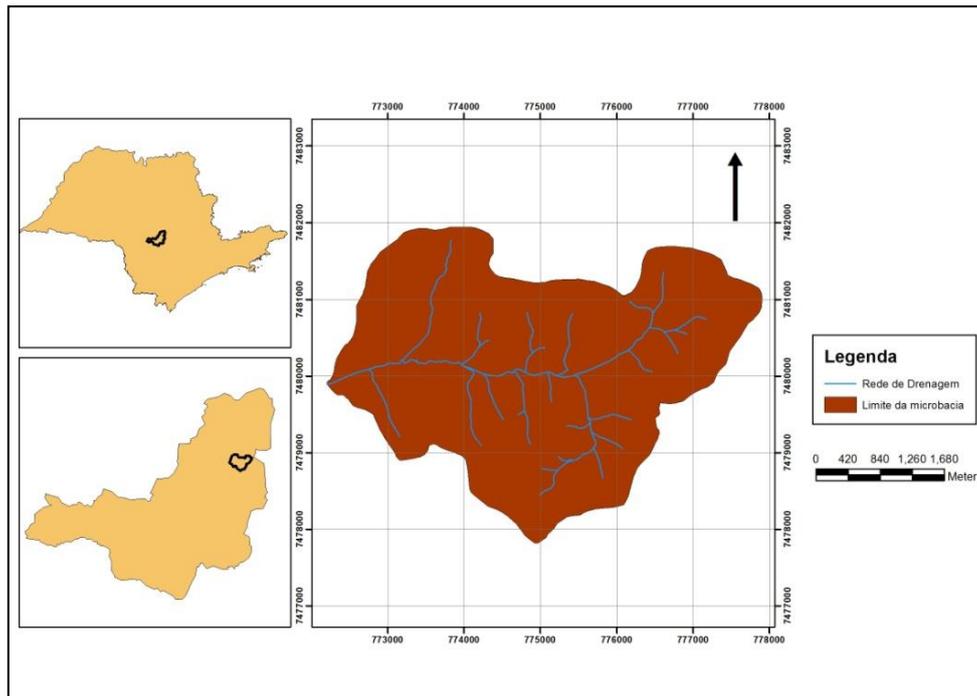
2. Metodologia de Trabalho

A microbacia do Ribeirão Duas Águas está inserida dentro do município de Botucatu (SP), sendo um afluente do Córrego Água da Lúcia, fazendo parte da bacia do Rio Capivara (Figura 1). A microbacia possui uma área de 1.363,73 hectares, e, segundo Jorge (2000), uma vegetação natural de Cerradão, sendo formada por Argissolos Vermelhos-Amarelos distróficos (CARREGA, 2006) e clima, Cfa segundo a classificação climática de Köppen-Geiger, com características de clima temperado chuvoso.

Para a confecção deste trabalho, foram utilizadas cartas topográficas digitalizadas de escala 1:50.000 que abrangem a área de estudo, de seguintes nomenclaturas: Botucatu (folha SF-22-R-IV-3) e Barra Bonita (folha SF-22-Z-B-VI-1), úteis para a delimitação da microbacia do Ribeirão Duas Águas, assim como para a digitalização da rede de drenagem. (IBGE,1973)

As cartas topográficas foram georreferenciadas no datum horizontal SIRGAS-2000, nas bandas 1, 2 e 3 pelo *software* Idrisi 15 – *The Andes Edition*, sendo montado um mosaico das duas cartas, para a confecção do limite da microbacia.

Figura 1. Localização da microbacia do Ribeirão Duas Águas em Botucatu-SP



Foram utilizadas imagens do programa SPOT (Sattelite Pour l’Observation de la Terre), série SPOT 5, datadas de 10/04/2009 a 27/06/2010. O programa foi planejado e projetado como um sistema operacional e comercial de observação da Terra, tendo suas atividades iniciadas em 1986 com o SPOT 1, operando hoje em dia com as séries SPOT 2, SPOT 4 e SPOT 5 (ROSA, 2009). A série SPOT 5, origem das imagens aqui utilizadas, possui uma órbita circular, heliosíncrona e polar, com período de visita de 26 dias e horário de passagem às 10:30 da manhã, operando com 3 sensores: HRS, VEGETATION-2 E HRG, este último a fonte de nossas imagens.

A partir das imagens, foi feito um mapa da *verdade terrestre* dos tipos de uso presentes na microbacia do Ribeirão Duas Águas, e depois foram realizadas as classificações que nos propomos em analisar. Primeiro, a classificação supervisionada MAXVER foi realizada (módulo *Maxlike*) com cinco amostras de cada uso como *training sites* (módulo *Makesig*), e também foi realizada uma divisão do uso “Reflorestamento” em 5 classes desse uso, a fim de conseguirmos uma maior acurácia em relação à classificação.

Esse procedimento foi feito pois, na área de estudo, havia diversos talhões de reflorestamento em diferentes estágios de crescimento. Segundo Novo (1992), a vegetação sofre alterações ao longo do tempo em relação ao seu comportamento espectral, como resultado de alterações sazonais ou em decorrência de seu estágio fenológico, dentre outros processos de modificação naturais. Logo depois de realizado o procedimento de classificação com diversas classes, foi feita uma reclassificação da imagem obtida para que a mesma tivesse exatamente o número de classes correspondente à verdade terrestre, resultando em cinco classes, sendo elas: Mata Nativa, Citrus, Culturas Anuais, Pastagem e Reflorestamento.

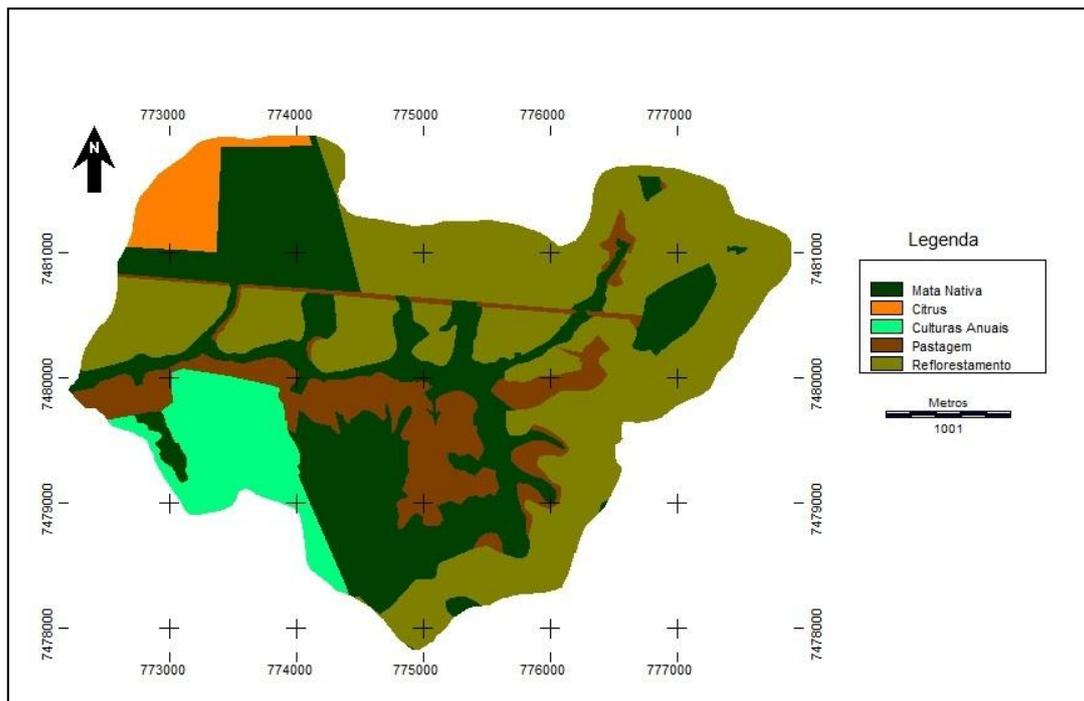
A partir das classes obtidas na classificação supervisionada, foi escolhido o mesmo número para que o *software* fizesse a classificação não supervisionada CLUSTER, com cinco classes referentes aos mesmos usos. Nesse método, o próprio *software* reconhece os padrões de agrupamento dos pixels coincidentes, e os agrupa por definição própria em classes distintas..

Ao fim dos processos de classificação, foi realizado o índice Kappa, processo realizado pelo Idrisi no módulo *Errmat*, inserindo-se o arquivo raster da verdade terrestre e as classificações realizadas, MAXVER e CLUSTER.

3. Resultados e discussões

Como resultados, obtivemos as imagens referentes à verdade terrestre presente na área de estudo, mostrada a seguir.

Figura 2. Verdade terrestre de microbacia do Ribeirão Duas Águas



Em relação aos usos presentes na microbacia, os mesmos se encontram quantificados na tabela a seguir.

Tabela 1. Área das classes de uso do solo da Microbacia do Ribeirão Duas Águas

Classes	Área (ha)	Área (%)
Mata Nativa	432.85	31,74
Citrus	52.09	3,82
Culturas Anuais	112.44	8,25
Pastagem	177.05	12,98
Reflorestamento	589.30	43,21
Total	1.363,73	100,00

Nota-se, a partir da tabela 1, que o principal uso é o reflorestamento, ocupando cerca de 43% do total da microbacia, enquanto que a mata nativa ocupa a segunda maior área, com cerca de 31% do total. Notamos que a área de mata nativa está presente principalmente nas áreas de proteção permanente seguindo os cursos que formam a rede de drenagem, mas

também em uma grande área no extremo norte da microbacia. As outras áreas em que está presente a mata nativa são em encostas no centro-sul da microbacia, onde se encontram algumas nascentes da drenagem.

A partir da separação da imagem real em três bandas, foram feitas as classificações MAXVER e CLUSTER, as quais demonstraram uma diferença considerável entre si em relação à localização das culturas presentes na área.

Figura 3. Classificação Supervisionada CLUSTER

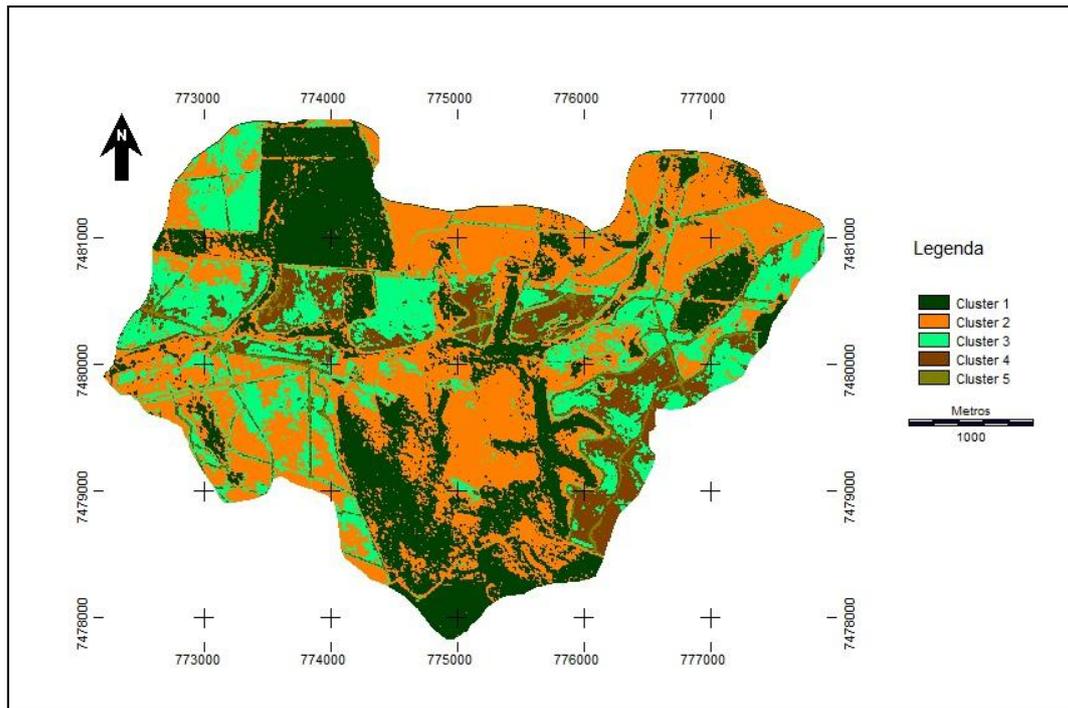
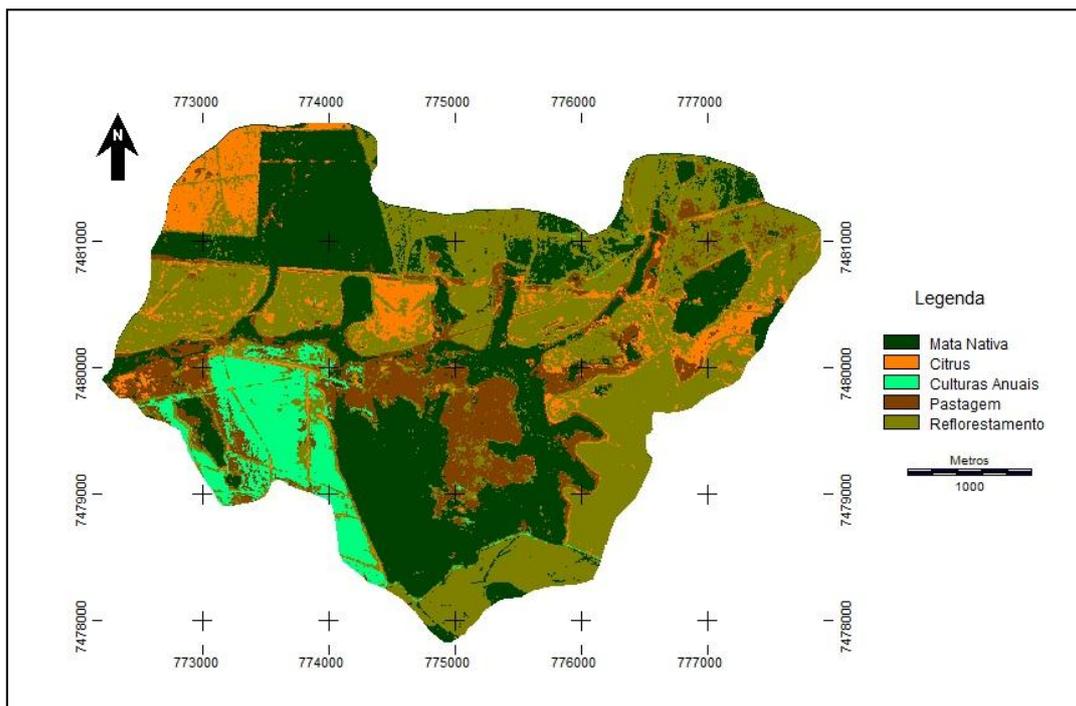


Figura 4. Classificação Supervisionada MAXVER



Podemos notar a diferença entre as classificações, sendo visivelmente mais apurada a MAXVER, porém somente observando-as não podemos inferir que uma ou outra são melhores para a representação do uso da terra da área de estudo. Para isso, foi utilizado o módulo *Errmat* do Idrisi para a confecção e cálculo do índice Kappa, um índice que mede a acurácia da classificação em relação à verdade terrestre. Quando realizado o índice para a comparação da verdade terrestre com a classificação CLUSTER, obtivemos um índice Kappa igual a 0,13, o que caracteriza, segundo Landis e Koch (1977 adaptado por Piroli, 2010b), a classificação como ruim, ou seja, que não atende às verdadeiras utilizações presentes na área.

Já quando realizamos o mesmo cálculo para a classificação MAXVER, obtivemos um índice Kappa igual a 0,55, considerada pelo mesmo autor como boa, atendendo em partes à realidade presente na área. Isso se deve ao método utilizado para a classificação MAXVER, que utiliza de intervalos definidos pelo usuário em *training sites* para a definição da classificação de um pixel em uma ou outra classe.

O intervalo de pixels analisados reduz a indução ao erro de confusão pelo *software*, apurando a classificação.

Na matriz de erro, o erro de omissão indica a probabilidade de uma unidade amostral de referência estar corretamente classificada e o erro de inclusão representa a probabilidade que um pixel classificado no mapa, represente a categoria no terreno. Erros de inclusão ocorrem quando um ponto é identificado como de uma classe, quando na realidade pertence a outra classe, enquanto que os erros de omissão ocorrem quando um ponto é identificado como pertencente a outra categoria, quando, na realidade, é membro de uma determinada classe (PIROLI, 2010b, p. 42)

4. Conclusões

Para a microbacia em estudo, o resultado foi que a classificação MAXVER é a mais indicada para realizarmos a análise do uso da terra por meio digitais, pela maior acurácia que esta demonstrou em relação à classificação CLUSTER. Entretanto, evidenciamos que ambas podem ter o mesmo resultado, se não considerado o método adequado para a realização das mesmas, mas a primeira, por possuir a interferência do usuário com a definição das áreas de treinamento para a classificação, sugere uma maior qualidade do que a segunda, que se baseia na reflectância simplesmente, e que está a mercê de confusões em relação aos valores dos pixels que são afetados seja pelo estado fenológico da planta, ou pelo tipo de cultura/uso ali presente.

Referências Bibliográficas

ALKIMIM, A. F. de; CASTRO, A. C.; DEMATTÊ, J. A. M.; FRANSCSCHINI, M. H. D.; GASPAR, J.. Uso de Imagens SPOT na identificação, quantificação e planejamento de Áreas de Proteção Permanente e Reserva Legal. *In: XV Simpósio Brasileiro de Sensoriamento Remoto*, Curitiba, 2011. **Anais do XV Simpósio Brasileiro de Sensoriamento Remoto – SBSR**. Curitiba: INPE, 2011. Disponível em: <<http://www.dsr.inpe.br/sbsr2011/files/p1386.pdf>>. Acesso em: 13 nov. 2012

CARREGA, E. F. B. **Delimitação de unidades ambientais na Bacia do Rio Capivara, Botucatu (SP)**. 2006. 98 f. Dissertação (Mestrado em Agronomia/ Energia na Agricultura) – Faculdade de Ciências Agrônomicas, Universidade Estadual Paulista, Botucatu, 2006.

- CARREGA, E. F. B. **Diagnóstico integrado de síntese da Bacia do Rio Capivara, Botucatu (SP)**. 2010. 260 f. Dissertação (Doutorado em Agronomia/ Energia na Agricultura) – Faculdade de Ciências Agrônomicas, Universidade Estadual Paulista, Botucaideratu, 2006.
- DEMARCHI, J. C.; PIROLI, E. L.; ZIMBACK, C. R. L.. Análise temporal do uso do solo e comparação entre os índices de vegetação NDVI e SAVI no município de Santa Cruz do Rio Pardo - SP usando imagens Landsat-5. **Ra'e ga** (UFPR), v. 21, p. 234-271, 2011.
- EASTMAN, J. R.. **Idrisi 15: The Andes Edition**. Worcester, MA: Clark University, 2006.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Superintendência de Cartografia. Plano Cartográfico do Estado de São Paulo. **Microbacia do Ribeirão Duas Águas. Folhas: Botucatu (SF-22-R-IV-3) e Barra Bonita (SF-22-Z-B-VI-1)**. São Paulo, Escala 1:50.000, 1973.
- INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Spring: Tutorial d Geoprocessamento: Classificação de Imagens**. Disponível em: <<http://www.dpi.inpe.br/spring/portugues/tutorial/classific.html>>. Acesso em 09 nov. 2012.
- JENSEN, J. R.. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. Tradução: José Carlos Neves Epiphanyo (coord.) et al. São José dos Campos: Parêntese, 2009.
- JORGE, L. A. B. Comportamento sazonal de fragmentos de vegetação natural na bacia do Rio Capivara, em Botucatu – SP. **Revista Árvore**, Viçosa – MGv. 24, n. 4, p.389 – 396, Viçosa, MG, 2000.
- LANDIS, R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v.33, n.1, p.159-174, Mar. 1977.
- NOVO, E. M. L. de M.. **Sensoriamento remoto: princípios e aplicações**. São Paulo: Edgard Blücher, 1992.
- PIROLI, E. L.. **Introdução ao geoprocessamento**. Ourinhos: Unesp/Campus Experimental de Ourinhos, 2010a. 46p.
- PIROLI, E. L.. **Disciplina de geoprocessamento: práticas em Idrisi – versão Taiga**. Ourinhos: Unesp/Campus Experimental de Ourinhos, 2010b. 56p.
- ROSA, R.. **Introdução ao sensoriamento remoto**. Uberlândia: EDUFU, 2009.
- ROSENFELD, G. H.; FITZPATRICK-LINS, K. A coefficient of agreement as a measure of thematic classification accuracy. **Photogrammetric Engineering and Remote Sensing**, v.52, n.2, p.223-227, Fev. 1986.