

DISCRIMINAÇÃO DE CLASSES DE COBERTURA VEGETAL UTILIZANDO TÉCNICAS DE CLASSIFICAÇÃO DIGITAL DE IMAGENS DE SENSORIAMENTO REMOTO¹

TATIANA MONTES CELINSKI² & CÉLIA REGINA LOPES ZIMBACK³

RESUMO: Este trabalho teve por objetivo a discriminação de classes de cobertura vegetal em imagens de sensoriamento remoto do satélite CBERS-2, relativas aos períodos de inverno e de verão da região dos Campos Gerais, no Estado do Paraná. A cobertura vegetal da região apresenta diferentes culturas de inverno e de verão, além de áreas de reflorestamento, pastagem e mata natural. Foram avaliadas as técnicas de classificação supervisionada Árvore de Decisão e Máxima Verossimilhança (MAXVER), a partir de um conjunto de atributos composto por: bandas do sensor CCD (1, 2, 3 e 4), índices de vegetação (CTVI, DVI, GEMI, NDVI, SR, SAVI, TVI), modelo de mistura (solo, sombra, vegetação) e componentes principais. A avaliação da acurácia dos classificadores foi feita com base na matriz de erros de classificação e no coeficiente *kappa*. Na definição das classes, buscou-se um alto nível discriminatório a fim de permitir a separação dos diferentes tipos de culturas presentes na região nas épocas de inverno e de verão. A classificação por Árvore de Decisão apresentou uma acurácia total de 94,5% e coeficiente *kappa* igual a 0,9389, para a cena 157/128; para a cena 158/127, apresentou os valores 88% e 0,8667, respectivamente. A classificação MAXVER apresentou uma acurácia total de 84,86% e coeficiente *kappa* igual a 0,8099, para a cena 157/128; para a cena 158/127, apresentou os valores 77,90% e 0,7476, respectivamente. Os resultados demonstraram que o desempenho do classificador Árvore de Decisão foi superior ao do MAXVER, especialmente para as classes relativas às culturas, indicando a sua aplicação para a obtenção do mapeamento da cobertura vegetal incluindo diferentes tipos de culturas.

Palavras-chave: Árvores de decisão, máxima verossimilhança, processamento de imagens digitais.

¹ Parte da tese de doutorado do 1.º autor intitulada: Classificação de cobertura do solo utilizando árvores de decisão e sensoriamento remoto

² Aluna do Programa de Pós-graduação em Agronomia – Energia na Agricultura – FCA/UNESP – Botucatu/SP e docente do Departamento de Informática – UEPG – Ponta Grossa/PR, tmontesc@uepg.br

³ 3Orientadora e docente do Departamento de Recursos Naturais /Ciência do Solo – FCA/UNESP – Botucatu/SP, rzimback@fca.unesp.br

DISCRIMINATION OF VEGETATION COVER CLASSES USING DIGITAL IMAGE TECHNIQUES AND REMOTE SENSING IMAGE

SUMMARY: *The aim of this work is to discriminate vegetation classes through remote sensing images from the satellite CBERS-2, related to winter and summer seasons in the Campos Gerais region Paraná State, Brazil. The vegetation cover of the region presents different kinds of vegetations: summer and winter cultures, reforestation areas, natural areas and pasture. Supervised classification techniques like Maximum Likelihood Classifier (MLC) and Decision Tree were evaluated, considering a set of attributes from images, composed by bands of the CCD sensor (1, 2, 3, 4), vegetation indices (CTVI, DVI, GEMI, NDVI, SR, SAVI, TVI), mixture models (soil, shadow, vegetation) and the two first main components. The evaluation of the classifications accuracy was made using the classification error matrix and the kappa coefficient. It was defined a high discriminatory level during the classes definition, in order to allow separation of different kinds of winter and summer crops. The classification accuracy by decision tree was 94.5% and the kappa coefficient was 0.9389 for the scene 157/128. For the scene 158/127, the values were 88% and 0.8667, respectively. The classification accuracy by MLC was 84.86% and the kappa coefficient was 0.8099 for the scene 157/128. For the scene 158/127, the values were 77.90% and 0.7476, respectively. The results showed a better performance of the Decision Tree classifier than MLC, especially to the classes related to cultivated crops, indicating the use of the Decision Tree classifier to the vegetation cover mapping including different kinds of crops.*

Keywords: *Decision trees, maximum likelihood, digital image processing.*

1 INTRODUÇÃO

O sensoriamento remoto juntamente com os sistemas de informações geográficas (SIG's) constituem-se em ferramentas essenciais para muitas aplicações na agricultura, como a identificação e o mapeamento de culturas. A identificação de um determinado tipo de cultura é fundamental para quantificação da biomassa, estimativa da área plantada, previsão de safras, dentre outras aplicações (CÂMARA et al., 1996).

Em aplicações voltadas à identificação e ao mapeamento de culturas, na etapa de classificação busca-se identificar cada *pixel* ou elemento da imagem segundo a ocupação do solo, obtendo-se um mapa de classes temáticas. A categorização é feita por meio de algoritmos de reconhecimento de padrões que utilizam os valores espectrais assim como outros parâmetros obtidos a partir da assinatura espectral. A classificação, portanto, requer o uso de algoritmos complexos e de técnicas de seleção e extração de características. Segundo Novo (1992), o objetivo das técnicas de classificação digital é tornar o processo de mapeamento ou reconhecimento de características da superfície terrestre menos subjetivo e com maior potencial de repetição em situações subseqüentes.

Os diversos métodos, em geral, quando aplicados à classificação de dados de sensores remotos de resolução média – como os sensores CCD/CBERS-2 – para fins de mapeamento da cobertura da terra, apresentam bons resultados quando o conjunto de classes reflete os tipos de cobertura, não fazendo distinção entre os diferentes tipos de culturas agrícolas.

Neste contexto, Silva e Pereira (2007) compararam o desempenho de três algoritmos: distância mínima até a média, paralelepípedo e máxima verossimilhança. Os algoritmos foram aplicados na classificação de uma imagem CCD/CBERS-2, utilizando como dados as bandas 2, 3 e 4 do sensor e considerou as classes mata, reflorestamento, pastagem, rio, lagoa e urbanização. Os resultados foram excelentes para os algoritmos distância mínima até a média e máxima verossimilhança, com valores *kappa* próximos a 0,90, e insatisfatórios para o método do paralelepípedo.

Da mesma forma, Freitas e Cruz (2005) avaliaram o desempenho de três diferentes combinações de atributos a partir de dados do sensor ETM+/Landsat 7 – somente bandas, bandas e imagens fração (solo, sombra e vegetação), imagens fração e componentes principais – na discriminação de classes de vegetação na mata atlântica. O procedimento de classificação constituiu-se pela seqüência: segmentação por crescimento de regiões, algoritmo ISOSEG e algoritmo Bhattacharya. Os resultados acusaram acurácias muito próximas para as três combinações de atributos, em torno de 64%, devido à dificuldade em separar as classes.

Figueiredo e Carvalho (2006) avaliaram a exatidão do mapeamento da cobertura da terra em Capixaba (AC) usando classificação por Árvore de Decisão. O conjunto de atributos utilizado incluiu as bandas do sensor TM/Landsat 5, as imagens fração derivadas do modelo de mistura espectral e o índice de vegetação NDVI. Para as classes temáticas consideradas neste estudo (floresta, capoeira, pasto alto, pasto baixo, solo e água) o desempenho da classificação por Árvore de Decisão apresentou valor *kappa* igual a 0,90. Comparativamente aos métodos MAXVER e ISODATA, os valores *kappa* foram de 0,89 e 0,79, respectivamente.

Em trabalho dos mesmos autores, Figueiredo e Carvalho (2007), outras comparações foram realizadas entre os algoritmos ISODATA, MAXVER e Árvore de Decisão, utilizando diferentes

composições de atributos a partir das bandas dos sensores TM/Landsat 5, ETM+/Landsat 7, das imagens fração do modelo de mistura espectral e do índice NDVI, para as mesmas classes de cobertura. Os resultados foram superiores para o algoritmo MAXVER com *kappa* próximo a 0,90, utilizando somente as bandas dos sensores ou somente as imagens fração. Entretanto, a classificação por Árvore de Decisão apresentou resultados próximos aos do classificador MAXVER, utilizando todos os atributos.

Quando a categorização é definida sobre classes específicas de cultivos agrícolas, os índices de acurácia apresentam-se relativamente menores, em função das semelhanças espectrais entre espécies vegetais. No trabalho de Fonseca et al. (2001), foram realizadas classificações pelos métodos MAXVER e ISOSEG, obtendo-se baixas acurácias de, respectivamente, 54,4% e 50%. O vetor característico usado foi composto das bandas 3, 4 e 5 do sensor TM/Landsat 5 e as classes consideradas foram milho, sorgo, cana-de-açúcar, pasto, sem vegetação, água e floresta.

Algoritmos baseados em árvores de decisão apresentam características interessantes, como boa acurácia na classificação, rapidez no treinamento e na execução. Acrescenta-se o fato de que não fazem suposições estatísticas sobre os dados e habilidade para manipular dados de diferentes escalas de medidas e relacionamentos não-lineares entre características e classes, o que permite a utilização de um conjunto característico amplo e variado.

Neste sentido, o objetivo desse trabalho foi: (i) avaliar a acurácia da classificação de imagens de sensoriamento remoto do satélite CBERS-2 utilizando os classificadores MAXVER e Árvore de Decisão, para fins de discriminação da vegetação da região dos Campos Gerais, no Estado do Paraná; (ii) avaliar o uso de um conjunto de atributos de imagem, para a obtenção do mapeamento da cobertura vegetal, composto por bandas do sensor CCD, índices de vegetação, imagens derivadas do modelo linear de mistura espectral e da transformação por componentes principais.

2 MATERIAL E MÉTODOS

2.1 Área de estudo

A área de estudo está localizada no Estado do Paraná, na região conhecida por Campos Gerais, concentrando-se nos municípios: Ponta Grossa, Carambeí, Castro, Imbaú, Ortigueira, Palmeira, Telêmaco Borba, Tibagi e Ventania. A região dos Campos Gerais recebeu originalmente este nome devido a sua caracterização fitogeográfica natural, com campos limpos e matas de galerias ou capões isolados de

floresta ombrófila mista, onde aparece o pinheiro araucária. O clima é do tipo subtropical úmido, com verão brando e, no inverno, geadas. As altitudes variam de 800m a 1200m, aproximadamente.

Quanto à geomorfologia, a região caracteriza-se por estar situada no segundo planalto paranaense, denominado Planalto de Ponta Grossa ou Planalto dos Campos Gerais. Este planalto está naturalmente limitado pela escarpa Devoniana, a leste, e pela escarpa da Esperança (Serra Geral), a oeste. A estrutura geológica e natureza das rochas são responsáveis pelos solos rasos e arenosos, pouco férteis.

Atualmente, é uma região de agricultura intensa, onde o sistema de plantio direto é amplamente difundido. Prevalece o plantio da soja e do milho no verão, e, em escala menor, do sorgo e do feijão. No inverno, predomina a cultura do trigo, e, em menor escala, culturas como triticale, centeio, cevada, girassol, canola, nabo forrageiro, aveia preta e aveia branca. Há também, na região, a ocorrência de pastagens utilizadas na criação de gado. A região possui, ainda, grandes áreas de reflorestamento de pinus e de eucalipto, principalmente no município de Telêmaco Borba e entorno.

2.2 Material

Foram utilizadas as cenas 157/128 e 158/127 do sensor CCD/CBERS-2, bandas 1, 2, 3 e 4, referentes às datas de passagem 07/03/2007 e 02/09/2007, respectivamente. No processamento dos dados, em todas as etapas foi utilizado o sistema de informações geográficas SPRING (CÂMARA et al., 1996), versão 4.3.2 e, para a classificação por Árvore de Decisão, o programa See5 (QUINLAN, 1993). Foi utilizado um aparelho GPS de navegação para verificação das amostras em campo.

2.3 Métodos

As imagens da região de estudo foram selecionadas por meio do catálogo disponível na internet, no sítio do INPE (2007), com correção geométrica de sistema, com datas de passagem em acordo com as épocas de cultura de inverno e de verão. Uma vez obtidas as imagens, estas foram convertidas para o formato GRIB pelo módulo IMPIMA do SPRING (CÂMARA et al., 1996).

As duas cenas foram recortadas com o objetivo de delimitar áreas de grande atividade agrícola e subsequente variedade de cobertura vegetal. A primeira, de 07/03/2007, tendo como referência o Município de Ponta Grossa, ficou delimitada pelas coordenadas UTM 570000 e 620000m E, 7180000 e 7282000m S, Datum SAD-69. A segunda, de 02/09/2007, tendo como referência o Município de

Telêmaco Borba, ficou delimitada pelas coordenadas UTM 506000 e 598000m E, 7280000 e 7348000m S, Datum SAD-69.

Para eliminar o erro de posicionamento das imagens com correção de sistema e refinar o erro interno, as imagens passaram pelo processo de registro. No procedimento de registro adotado neste trabalho, utilizou-se uma imagem georreferenciada e ortoretificada para servir de referência, a qual foi obtida na internet, por meio de sítio de distribuição de imagens Landsat (LANDSAT, 2007).

No processo de registro, pontos de controle da imagem de referência foram adquiridos e relacionados a pontos homólogos reconhecidos na imagem a ser registrada. Os pontos de controle adquiridos foram selecionados com base no erro calculado e apresentado, considerando-se o grau do polinômio usado no registro. A escolha do grau do polinômio e do número de pontos de controle, em função do erro apresentado, obedeceu às recomendações para o usuário do software SPRING.

Após o registro, as imagens correspondentes às duas cenas em estudo foram processadas a fim de realizar a correção radiométrica e também de gerar o conjunto de atributos a ser utilizado na etapa de classificação. Esta etapa incluiu os procedimentos de conversão dos números digitais (DN) para valores de reflectância aparente, e de correção atmosférica, os quais foram realizados por meio de programas escritos na linguagem de programação LEGAL (Linguagem Espacial de Geoprocessamento Algébrico) do SPRING.

A conversão para valores de reflectância aparente foi realizada de acordo com o método proposto por Markham e Barker (1986), usando os coeficientes de calibração absoluta da câmara CCD/CBERS-2 apresentados por Ponzoni et al. (2006). Esta conversão foi aplicada a cada uma das quatro bandas utilizadas.

Após a conversão para valores físicos, foi realizada a correção atmosférica pelo método de correção atmosférica pelo *pixel* escuro (CHAVEZ, 1988). Os valores subtraídos de cada banda foram obtidos observando-se os respectivos histogramas, no SPRING. Esta conversão foi realizada por meio de um programa escrito em LEGAL.

A partir das bandas corrigidas, foram também obtidos os seguintes atributos: índices de vegetação CTVI, DVI, GEMI, NDVI, SR, SAVI e TVI (PONZONI; SHIMABUKURO, 2007); imagens fração solo, vegetação e sombra/água a partir das bandas 2, 3 e 4; primeira e segunda componentes da transformação por componentes principais, a partir das bandas 2, 3 e 4. Os índices de vegetação CTVI, GEMI, SAVI e TVI foram obtidos por programas desenvolvidos na linguagem de programação LEGAL/SPRING. Os demais atributos foram obtidos diretamente no SPRING.

Para a definição das classes de cobertura vegetal assim como para a definição do conjunto de amostras de treinamento para a classificação, foram realizadas saídas de campo a fim de identificar as diferentes culturas existentes bem como sua distribuição nas áreas de estudo. A coleta foi realizada

utilizando-se um aparelho GPS de navegação para individualizar as amostras, buscando-se atingir a diversidade de culturas presente nas regiões de estudo. Os dados foram coletados em datas próximas às datas de passagem do satélite para as imagens selecionadas, a fim de garantir confiabilidade às amostras.

As classes de mapeamento e cobertura foram determinadas com base nas características da vegetação e das culturas presentes na região de estudo. Para o período de verão, foram estabelecidas as classes: corpos d'água, solo exposto, eucalipto, pinus, mata natural, sorgo-milho, pasto, soja, feijão e área urbana. Para o período de inverno, foram consideradas as classes: corpos d'água, solo exposto, eucalipto, pinus, araucária, mata natural, trigo, aveia, pasto e área urbana.

Antes da classificação propriamente dita, o treinamento supervisionado foi realizado no SPRING, com base na amostragem realizada em campo e nas classes de cobertura estabelecidas. Para a etapa de classificação foi considerado o conjunto característico com as bandas corrigidas 1, 2, 3 e 4.

A classificação foi realizada pelo método MAXVER, por meio do SPRING, a partir das amostras de treinamento selecionadas, utilizando-se diversas combinações de atributos e analisando-se a acurácia obtida a partir de cada combinação. Após a classificação, foi gerado o mapa de classes de cobertura no SPRING.

Para a classificação por Árvore de Decisão, foram gerados os conjuntos de amostras de treinamento e de teste, cada qual com duzentas amostras, sendo que cada amostra se constituiu em um vetor característico com dezesseis atributos da imagem. Essas amostras foram obtidas a partir do conjunto de treinamento gerado no SPRING para o classificador MAXVER, a fim de que as amostras utilizadas nos dois classificadores pertencessem à mesma localização na imagem, para fins de comparação.

Os arquivos de amostras de treinamento e de teste, juntamente com o arquivo contendo a definição dos dezesseis atributos e das classes, foram utilizados no programa See5 (QUINLAN, 1993), para a geração do classificador Árvore de Decisão, a partir dos valores padrão de todas as opções apresentadas na caixa de diálogo do programa. A partir do classificador obtido, foram implementadas rotinas em linguagem de programação LEGAL, no SPRING, para classificar a imagem e gerar o mapa de cobertura.

A análise estatística dos resultados de classificação obtidos pelos classificadores foi feita com base na matriz de erros de classificação ou matriz de confusão, obtida de cada classificação, que é o instrumento mais comumente usado para expressar a exatidão da classificação, de acordo com Lillesand e Kiefer (1994). A análise estendeu-se também à exatidão total, às exatidões individuais de cada classe e ao coeficiente *kappa*, obtidos a partir dos valores da matriz de confusão.

3 RESULTADOS E DISCUSSÃO

3.1 Discriminação da cena 157/128

Em relação à cena 157/128, a classificação realizada pelo algoritmo MAXVER usando combinações feitas a partir do conjunto característico contendo os dezesseis atributos (bandas corrigidas 1, 2, 3 e 4; índices de vegetação CTVI, DVI, GEMI, NDVI, SR, SAVI e TVI; imagens fração vegetação, solo e sombra/água; primeira e segunda componentes derivadas da transformação por componentes principais) resultou em diferentes valores para as acurácias. O melhor resultado, entretanto, foi obtido considerando-se somente as quatro bandas corrigidas, sendo a respectiva matriz de erros de classificação apresentada na Tabela 1.

Quanto à avaliação do resultado da classificação, a acurácia total foi igual a 84,86% e o coeficiente kappa foi 0,8099, sendo este último considerado excelente de acordo com a classificação de Landis e Koch (1977). Entretanto, os valores das acurácias das classes relativas às culturas – eucalipto, sorgo/milho, feijão e soja – e também à classe pasto, foram inferiores a 70%. Observou-se claramente confusão entre as classes mata e eucalipto, mata e pinus, pasto e sorgo/milho, soja e pasto, além de feijão, soja e pasto, expressa por meio da acurácia do produtor, que indica quão bem os *pixels* foram classificados para determinada classe. Este fato também pode ser observado por meio do erro de omissão das classes.

Tabela 1 - Matriz de erros de classificação (MAXVER, cena 157/128).

Classificação	Dados de referência										Total	Acurácia do usuário
	AG	EU	SM	PI	MA	SO	PA	AU	SJ	FE		
Água (AG)	278	0	0	0	0	0	0	0	0	0	278	100,00%
Eucalipto (EU)	0	12	0	2	0	0	0	0	0	0	14	85,71%
Sorgo/milho (SM)	0	0	8	0	0	0	12	0	0	4	24	33,33%
Pinus (PI)	0	0	0	108	24	0	0	0	0	0	132	81,82%
Mata (MA)	0	28	1	2	250	0	0	0	0	0	281	88,97%
Solo (SO)	0	0	0	0	0	42	0	0	0	0	42	100,00%
Pasto (PA)	0	0	43	0	0	0	22	0	21	4	90	24,44%
Área urbana (AU)	0	0	0	0	0	0	0	26	0	1	27	96,30%
Soja (SJ)	0	0	0	0	0	0	0	0	43	0	43	100,00%
Feijão (FE)	0	0	0	0	0	0	0	0	0	7	7	100,00%
Total	278	40	52	112	274	42	34	26	64	16	938	
Acurácia do produtor	100,00%	30,00%	15,38%	96,43%	91,24%	100,00%	64,71%	100,00%	67,19%	43,75%		
Acurácia total = 84,86%		Kappa = 80,99%										

O erro de inclusão ou comissão, assim como a acurácia do usuário, também são indicadores da qualidade da classificação. Estas medidas indicam a probabilidade de um *pixel* classificado para uma determinada classe de fato pertencer a esta classe. Com relação às classes eucalipto, mata, pasto, pinus e sorgo/milho, verificaram-se erros de inclusão superiores a 10%.

Analisando outras combinações de atributos verificaram-se acurácias totais superiores a 80% para diversas combinações, envolvendo as bandas corrigidas, as imagens fração, as componentes primeira e segunda, e os índices GEMI, NDVI, SAVI e TVI. Todavia, as acurácias individuais de cada classe, em especial das classes de interesse, em geral, não melhoraram. Quanto às combinações envolvendo os demais índices de vegetação em conjunto, não houve melhora nos resultados em termos de acurácia total nem individual de cada classe.

A Figura 1 exibe o mapa temático correspondente à classificação MAXVER realizada a partir das quatro bandas corrigidas, a qual apresentou o maior valor para a acurácia total assim como, para as acurácias relativas às culturas agrícolas.

A análise visual do mapa temático, com base no trabalho de campo realizado, confirma a presença de áreas de reflorestamento de pinus em diversos pontos e também a mata natural da região de Itaiacoca, distrito do Município de Ponta Grossa, que se destaca à direita. Da mesma forma, visualiza-se a confusão entre diversas classes, especialmente envolvendo as classes sorgo/milho e pasto.

As classes eucalipto, sorgo/milho e feijão apresentaram um número de amostras de referência menor em relação às outras classes porque ocorreram em menor quantidade na cena 157/128. Este fato pode explicar a confusão e os erros relativos a essas classes.

À época da aquisição da cena 157/128, o pasto apresentava-se heterogêneo, por tratar-se de final da estação de verão, fato que foi observado durante as visitas a campo. Na matriz de erros de classificação correspondente, a classe pasto apresentou erro de inclusão igual a 75,56% e confusão com a classe sorgo/milho.

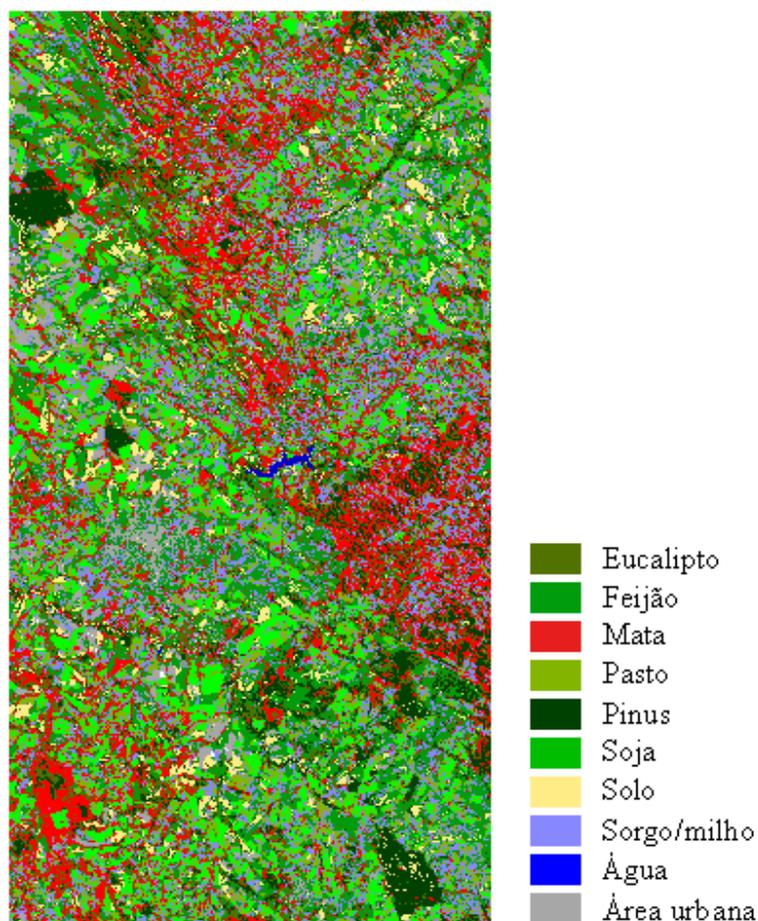


Figura 1 – Mapa temático resultante da classificação MAXVER, relativo à cena 157/128.

As saídas resultantes da aplicação do algoritmo Árvore de Decisão, a partir do conjunto de treinamento, correspondem à árvore de decisão ou classificador e à matriz de erros de classificação relativa aos dados de teste. A Tabela 2 apresenta a matriz de erros de classificação relativa ao conjunto de teste, que corresponde a casos não usados no treinamento, para a cena 157/128.

Ressalta-se que estes resultados foram obtidos considerando-se os valores padrão para as opções da caixa de diálogo do programa See5, visto que os resultados foram satisfatórios para os mesmos. Nesta situação, fica habilitado o recurso de poda da árvore em até 25%, e o número mínimo de casos de treinamento estabelecido para os ramos igual a 2.

A acurácia total obtida a partir do classificador Árvore de Decisão foi de 94,50%, superior à acurácia obtida pelo classificador MAXVER para a mesma cena (84,86%). Da mesma forma, o valor do coeficiente *kappa*, igual a 0,9389, foi superior. De acordo com Landis e Koch (1977), este valor é

considerado excelente. Algumas acurácias individuais de cada classe foram igualmente superiores, permitindo uma melhor discriminação entre as classes. Para as classes feijão, soja, pasto, solo, mata, sorgo/milho e eucalipto, verificou-se uma melhora importante. Os erros de omissão e inclusão ficaram abaixo de 10%, com exceção das classes mata (omissão: 13,64%), eucalipto e pasto (inclusão: 15,00%).

Tabela 2 – Matriz de erros de classificação relativa ao classificador Árvore de Decisão, cena 157/128.

Classificação	Dados de referência										Total	Acurácia do usuário
	AG	EU	SM	PI	MA	SO	PA	AU	SJ	FE		
Água (AG)	19	0	0	0	0	1	0	0	0	0	20	95,00%
Eucalipto (EU)	0	17	0	1	2	0	0	0	0	0	20	85,00%
Sorgo/milho (SM)	0	0	19	0	0	0	1	0	0	0	20	95,00%
Pinus (PI)	0	0	0	19	1	0	0	0	0	0	20	95,00%
Mata (MA)	0	1	0	0	19	0	0	0	0	0	20	95,00%
Solo (SO)	0	0	0	0	0	19	0	0	0	0	20	95,00%
Pasto (PA)	0	0	1	0	0	0	17	0	1	1	20	85,00%
Área urbana (AU)	0	0	0	0	0	0	0	20	0	0	20	100,00%
Soja (SJ)	0	0	0	0	0	0	0	0	20	0	20	100,00%
Feijão (FE)	0	0	0	0	0	0	0	0	0	20	20	100,00%
Total	21	18	18	21	21	19	20	22	20	20	200	
Acurácia do produtor	100,00%	94,44%	95,00%	90,48%	86,36%	95,00%	94,44%	100,00%	95,24%	95,24%		
Acurácia total = 94,50%		<i>Kappa</i> = 93,89%										

Comparada aos resultados obtidos por Fonseca et al. (2001) com os classificadores MAXVER e ISOSEG, que buscaram a categorização nas classes milho, sorgo, cana-de-açúcar, pasto, sem vegetação, água e floresta, as acurácias total e individual das classes foram superiores. Os autores utilizaram como atributos as bandas 3, 4 e 5 do sensor TM/Landsat 5, obtendo acurácias totais de 54,4% e 50%, respectivamente, para os classificadores MAXVER e ISOSEG.

A árvore de decisão obtida foi utilizada para classificar a imagem, por meio de um programa desenvolvido em linguagem de programação LEGAL no SPRING. A Figura 2 exibe o mapa temático correspondente à classificação utilizando o classificador Árvore de Decisão, para a cena 157/128, de 07/03/2007.

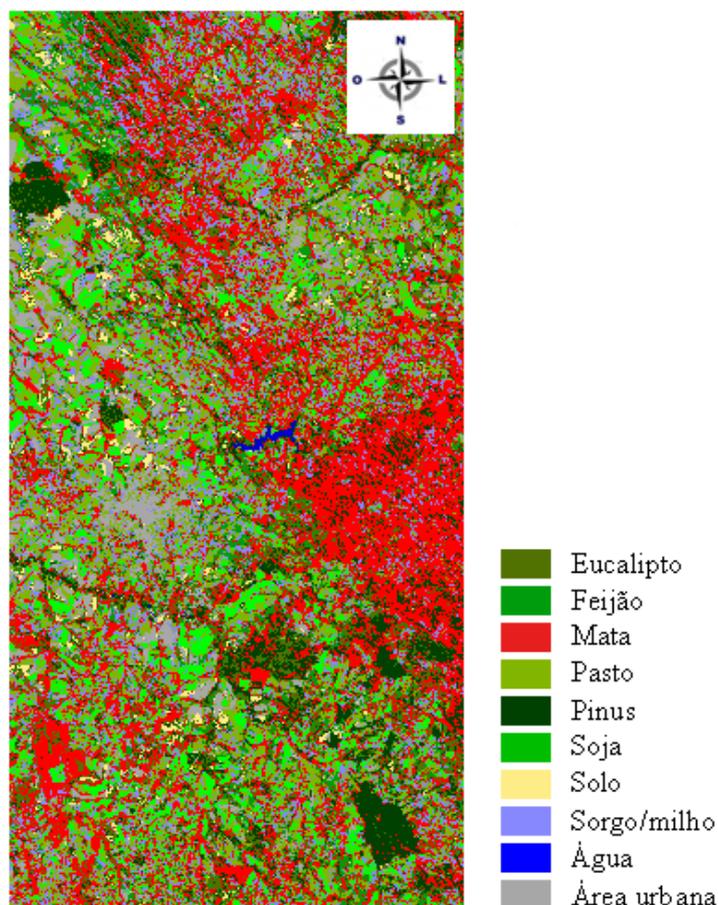


Figura 2 – Mapa temático resultante da classificação por Árvore de Decisão, relativo à cena 157/128.

3.2 Discriminação da cena 158/127

O mesmo procedimento metodológico foi aplicado à cena 158/127, de 02/09/2007, considerados os mesmos dezesseis atributos, resultando na matriz de erros de classificação exibida na Tabela 3. O mapa temático correspondente é exibido na Figura 3. Da mesma forma que para a cena 157/128, os melhores resultados também ocorreram para a classificação realizada a partir das quatro bandas corrigidas. Outras combinações de atributos não resultaram em melhores resultados quanto à acurácia total.

O valor da acurácia total obtido a partir das bandas foi igual a 77,90%. Entretanto, as classes aveia, araucária e área urbana apresentaram acurácias individuais inferiores a 60%. Ficaram evidentes as confusões entre as classes araucária e pinus, eucalipto e pinus, assim como aveia e trigo. Os erros de

omissão das classes aveia, trigo, eucalipto, pinus, araucária e área urbana ficaram acima de 10%, sendo os mais elevados relativos às classes aveia e araucária. Os menores valores para a acurácia do usuário ocorreram para as classes pinus e araucária, com respectivos erros de inclusão maiores, indicando uma menor probabilidade de um *pixel* pertencer a estas classes, de fato.

Tabela 3 – Matriz de erros de classificação relativa ao Classificador MAXVER, cena 158/127.

Classificação	Dados de referência										Total	Acurácia do usuário
	AV	TR	EU	PI	AR	MA	SO	PA	AU	AG		
Aveia (AV)	16	1	2	0	0	3	0	0	0	0	22	72,73%
Trigo (TR)	21	73	7	0	0	0	0	0	0	0	101	72,28%
Eucalipto (EU)	0	8	53	6	0	2	0	0	0	0	69	76,81%
Pinus (PI)	0	0	8	34	18	1	0	0	0	0	61	55,74%
Araucária (AR)	0	0	0	9	11	0	0	0	0	0	20	55,00%
Mata (MA)	0	0	6	1	0	88	0	1	0	0	96	91,67%
Solo (SO)	0	0	0	0	0	0	64	3	11	0	78	82,05%
Pasto (PA)	6	0	0	0	0	0	0	42	0	0	48	87,50%
Área urbana (AU)	0	0	0	0	0	0	4	0	13	0	17	76,47%
Água (AG)	0	0	0	0	0	0	0	0	0	22	22	100,00%
Total	43	82	76	50	29	94	68	46	24	22	534	
Acurácia do produtor	37,21%	89,02%	69,74%	68,00%	37,93%	93,62%	94,12%	91,30%	54,17%	100,00%		
Acurácia total = 77,90%		Kappa = 74,76%										

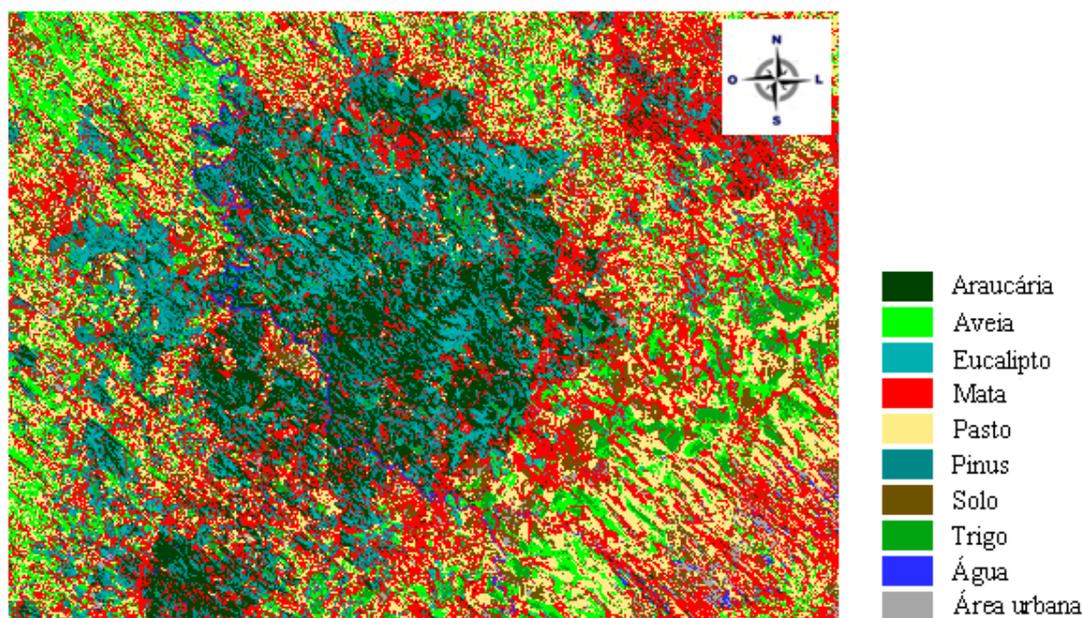


Figura 3 – Mapa temático resultante da classificação MAXVER, relativo à cena 158/127.

Da análise visual do mapa temático exibido na Figura 3, é possível perceber a área de reflorestamento no entorno do Município de Telêmaco Borba. Entretanto, nesta área, a espécie predominante era o pinus, seguido do eucalipto e em menor escala a araucária, o que não ocorre no mapa resultante. De fato, a matriz de erros de classificação expressa confusão entre essas classes.

Para a cena 158/127, de 02/09/2007, que corresponde ao período de inverno, os resultados obtidos da aplicação da classificação por Árvore de Decisão foram superiores em relação ao classificador MAXVER, da mesma forma que ocorreu para a cena 157/128. A matriz de erros de classificação correspondente é exibida na Tabela 4. Para a obtenção desses resultados, foram utilizados os valores padrão do programa See5.

Tabela 4 – Matriz de erros de classificação relativa ao classificador Árvore de Decisão, cena 158/127.

Classificação	Dados de referência										Total	Acurácia do usuário
	AV	TR	EU	PI	AR	MA	SO	PA	AU	AG		
Aveia (AV)	19	1	0	0	0	0	0	0	0	0	20	95,00%
Trigo (TR)	1	19	0	0	0	0	0	0	0	0	20	95,00%
Eucalipto (EU)	0	0	20	0	0	0	0	0	0	0	20	100,00%
Pinus (PI)	0	0	0	13	6	1	0	0	0	0	20	65,00%
Araucária (AR)	0	0	0	2	18	0	0	0	0	0	20	90,00%
Mata (MA)	0	0	0	3	0	16	0	1	0	0	20	80,00%
Solo (SO)	0	0	0	0	0	0	15	2	3	0	20	75,00%
Pasto (PA)	1	0	0	0	0	0	0	19	0	0	20	95,00%
Área urbana (AU)	0	0	0	0	0	0	2	1	17	0	20	85,00%
Água (AG)	0	0	0	0	0	0	0	0	0	20	20	100,00%
Total	21	20	20	18	24	17	17	23	20	20	200	
Acurácia do produtor	90,48%	95,00%	100,00%	72,22%	75,00%	94,12%	88,24%	82,61%	85,00%	100,00%		
Acurácia total = 88,00%		<i>Kappa</i> = 86,67%										

A acurácia total foi de 88%, superior àquela obtida pelo classificador MAXVER para a mesma cena (77,9%). Quanto às acurácias individuais de cada classe, os valores obtidos apresentaram valores superiores em relação ao Classificador MAXVER para a mesma cena, para as classes aveia, trigo, eucalipto pinus, araucária, mata e área urbana. Do ponto de vista dos erros de omissão e inclusão, estes apresentaram valores relativamente menores para a mesma cena, em relação à classificação MAXVER. As classes pinus e araucária foram as que apresentaram maior confusão.

A Figura 4 exibe o mapa temático resultante do classificador gerado. Para a migração da imagem classificada para o mapa, foi feito um programa em linguagem de programação LEGAL no SPRING.

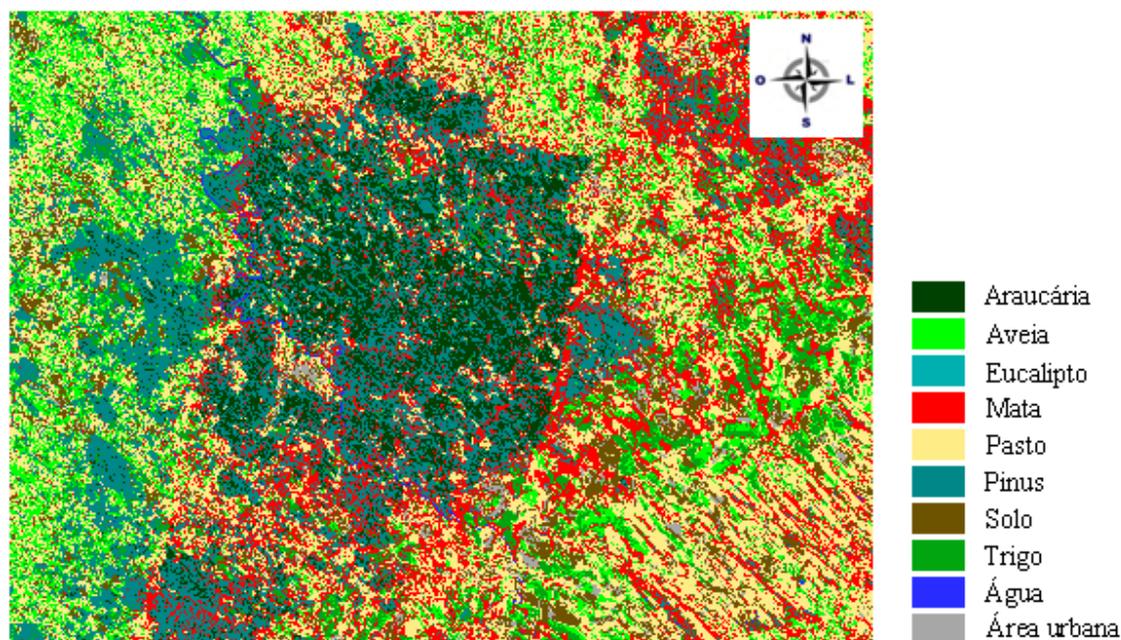


Figura 4 – Mapa temático resultante da classificação por Árvore de Decisão, relativo à cena 158/127.

3.3 Análise dos resultados

A Tabela 5 resume os resultados relacionados à exatidão obtidos da aplicação dos métodos de classificação MAXVER e Árvore de Decisão em relação às duas cenas estudadas. Para as duas as cenas, o classificador Árvore de Decisão resultou em uma maior acurácia total. O coeficiente *kappa*, que leva em conta também os valores da matriz de confusão que estão fora da diagonal principal, foi igualmente superior, indicando uma concordância maior em relação aos dados usados como referência.

Para o classificador Árvore de Decisão, o uso de um amplo e variado conjunto de atributos demonstrou ser conveniente, enquanto que para o Classificador MAXVER isto não se verificou. Com base nas matrizes de erros de classificação, observou-se que os resultados das acurácias individuais de cada classe apresentaram resultados melhores, na sua maioria, para a classificação por Árvore de Decisão.

Tabela 5 – Resultados obtidos (MAXVER X Árvore de Decisão).

Classificador	Cena 157/128 (07/03/2007)		Cena 158/127 (02/09/2007)	
	Acurácia total	<i>Kappa</i>	Acurácia total	<i>Kappa</i>
MAXVER	84,86%	80,99%	77,90%	74,76%
Árvore de Decisão	94,50%	93,89%	88,00%	86,67%

4 CONCLUSÕES

O estudo mostrou a possibilidade de obtenção de um mapeamento da cobertura vegetal incluindo a discriminação entre culturas espectralmente próximas, a partir de dados de sensoriamento remoto de nível médio como o sensor CCD/CBERS-2, utilizando classificação por Árvore de Decisão.

Os resultados demonstraram que a utilização de um grande número de atributos obtidos a partir das bandas do sensor é vantajosa para o classificador Árvore de Decisão, que seleciona aqueles que melhor separam as classes durante o processo de geração da árvore de decisão. Desta forma, é possível incorporar outros atributos, os quais podem ser adquiridos de outras fontes, de forma a garantir uma discriminação que mais se aproxime da realidade.

A metodologia utilizada para o mapeamento a partir de classificação por Árvore de Decisão incluiu dados de sensoriamento remoto e softwares disponíveis na internet, provendo uma forma simples para utilização e de resultados excelentes para a discriminação das classes presentes nas cenas de estudo.

4 REFERÊNCIAS

CÂMARA, G. et al. SPRING: integrating remote sensing and GIS by object-oriented data modelling. *Computers & Graphics*, New York, v. 20, n. 3, p. 395-403, May/June 1996.

CHAVEZ, P. S. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. **Remote Sensing of Environment**, New York, v. 24, n. 3, p. 459-479, 1988.

FIGUEIREDO, S. M. de M.; CARVALHO, L. M. T. de. Análise comparativa entre técnicas de classificação digital em imagens Landsat no mapeamento do uso e ocupação do solo em Capixaba, Acre. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., 2007, Florianópolis. **Anais eletrônicos...** São José dos Campos: INPE, 2007. Disponível em: <<http://www.dsr.inpe.br/sbsr2007/biblioteca/>>. Acesso em: 23 ago. 2007.

FIGUEIREDO, S. M. de M.; CARVALHO, L. M. T. de. Avaliação da exatidão do mapeamento da cobertura da terra em Capixaba, Acre, utilizando classificação por árvore de decisão. **Cerne**, Lavras, v. 12, n. 1, p. 38-47, jan./mar. 2006.

FONSECA, E. L. da; OLIVEIRA, J. C. de; RIZZI, R. Avaliação da identificação de áreas agrícolas para fins de previsão de safras utilizando procedimentos de classificação digital de imagens do sensor TM/Landsat 5. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 10., 2001, Foz do Iguaçu. **Anais eletrônicos...** São José dos Campos: INPE, 2001. Disponível em: <<http://www.dsr.inpe.br/sbsr2007/biblioteca/>>. Acesso em: 16 nov. 2006.

FREITAS, S. R.; CRUZ, C. B. M. Análise de componentes principais e modelo linear de mistura na discriminação de classes de vegetação na Mata Atlântica. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., 2005, Goiânia. **Anais eletrônicos...** São José dos Campos: INPE, 2005. Disponível em: <<http://www.dsr.inpe.br/sbsr2007/biblioteca/>>. Acesso em: 9 abr. 2007.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. Satélite sino-brasileiro de recursos terrestres. Disponível em: <<http://www.cbcrs.inpe.br/?content=index>>. Acesso em: 15 dez. 2007.

LANDIS, J.; KOCH, G. G. The measurements of agreement for categorical data. **Biometrics**, Washington, DC, v. 33, n. 3, p. 159-179, Mar. 1977.

LANDSAT. Disponível em: <<http://www.landsat.org>>. Acesso em: 10 mar. 2007.

LILLESAND, T. M.; KIEFER, R. W. **Remote sensing and image interpretation**. 3rd ed. New York: John Wiley, 1994. 1040 p.

MARKHAM, B. L.; BARKER, J. L. Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures. **EOSAT Landsat Technical Notes**, Lanham, MD, n. 1, p. 3-8, Aug. 1986.

NOVO, E. M. L. de. **Sensoriamento remoto: princípios e aplicações**. 2. ed. São Paulo: Edgard Blücher, 1992. 328 p.

PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento remoto no estudo da vegetação**. São José dos Campos: Parêntese, 2007. 136 p.

PONZONI, F. J.; ZULLO JUNIOR, J.; LAMPARELLI, R. A. C. Transformações radiométricas dos dados dos sensores do satélite CBERS-2. In: SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 1., 2006, Campo Grande. **Anais eletrônicos...** Campinas: EMBRAPA Informática Agropecuária; São José dos Campos: INPE, 2006. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m17@80/2006/12.08.13.51.31>>. Acesso em: 27 ago. 2007.

QUINLAN, J. R. **C4.5: programs for machine learning**. San Francisco, CA: Morgan Kaufmann, 1993. 302 p.

SILVA, M. L. M. de; PEREIRA, G. P. M. Avaliação de algoritmos de classificação supervisionada para imagens do Cbers-2 da Região do Parque Estadual do Rio Doce-MG. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., 2007, Florianópolis. **Anais eletrônicos...** São José dos Campos: INPE, 2007. Disponível em: <<http://www.dsr.inpe.br/sbsr2007/biblioteca/>>. Acesso em: 17 abr. 2007.