

Avaliação da acurácia de classificação digital de imagens no mapeamento de áreas de pastagens degradadas em Rio Branco, Acre

Symone Maria de Melo Figueiredo¹
Luciana Mendes Cavalcante²
Judson Ferreira Valentim³
Evandro Orfanó Figueiredo³

¹ Universidade Federal do Acre - UFAC
Campus Universitário, Rodovia BR-364, Km 04, 69915-900 – Rio Branco - AC, Brasil
symone_ac@yahoo.com.br

² Empresa Brasileira de Pesquisa Agropecuária Amazônia Oriental - Embrapa Amazônia Oriental
Trav. Dr. Enéas Pinheiro s/nº, Caixa Postal 48 - 66095-100 – Belém - PA - Brasil
luciana@cpatu.embrapa.br

³ Empresa Brasileira de Pesquisa Agropecuária - Embrapa Acre
Rodovia BR-364, Km 14, Caixa Postal 321 – 69914-220 – Rio Branco - AC, Brasil
(judson, orfano)@cpafac.embrapa.br

Abstract. The purpose of this work was to evaluate the effect of using several image classification schemes for mapping of degraded pasture in the municipality of Rio Branco-AC, Brazil. Multi-spectral Landsat data, fraction images and images representing the Normalized Difference Vegetation Index (NDVI) were used as attributes in the classification procedures in order to map forest, degraded pasture, high and low pasture, water and bare soil. Four classification schemes were used: a) maximum likelihood using Landsat images; b) maximum likelihood using fraction images; c) isodata using Landsat images; and d) decision tree using Landsat images, fraction images and NDVI. Results showed that all classification methods used were efficient ranging from very good to excellent classification according to the Kappa coefficient. The scheme using the maximum likelihood classification algorithm performed better than the others the mapping of degraded pasture.

Palavras-chave: análise de mistura espectral, NDVI, mineração de dados, exatidão do mapeamento, Amazônia, linear spectral unmixing, NDVI, data mining, mapping accuracy, Amazon.

1. Introdução

O Estado do Acre possui aproximadamente 12% de seu território desmatado e nessa área, cerca de 90% representam áreas de pastagem sob produção pecuária. Em termos econômicos, a importância da atividade pecuária tem crescido significativamente no setor primário de produção, representando cerca de 46% do valor bruto da produção agropecuária do Estado (Ibge, 2005).

Apesar dessa importância, a degradação de pastagens constitui um dos principais problemas agrônômicos para a atividade pecuária no Acre. Nos últimos quinze anos, a Embrapa Acre em parceria com outras instituições do Acre (Emater, Imac, Senar, Sebrae) promoveu um intensivo processo de transferência e adoção de tecnologias (Acre, 2000), visando aumento da produtividade e eficiência das atividades agropecuárias, além de estimular aplicação de metodologias que mitiguem a degradação das pastagens. Avaliações em campo visando caracterizar o grau de degradação de pastagens vêm sendo realizadas a contento (Dias-Filho e Andrade, 2005). Por outro lado, face à disponibilidade e possibilidades de aplicação das técnicas e dos produtos de sensoriamento remoto e geoprocessamento, faz-se

necessário uma avaliação sobre a aplicação de tais técnicas e produtos na caracterização e avaliação da dinâmica de áreas de pastagem no Acre.

Nesse contexto, o objetivo desse estudo é avaliar as metodologias de classificação de imagens de sensoriamento remoto (máxima verossimilhança, isodata e árvore de decisão) no mapeamento das áreas de pastagens degradadas. Entendendo-se a relação de interdependência entre o conhecimento de uma área, um bom mapeamento e uma boa classificação, a execução deste projeto poderá contribuir com subsídios a futuras aplicações por parte de instituições ligadas ao meio ambiente e de fomento às atividades agropecuárias em áreas geográficas semelhantes.

2. Área de estudo

A área de estudo está localizada na região sudeste do estado do Acre, situada em torno de $10^{\circ}03'33''$ de latitude sul e $68^{\circ}15'40''$ de longitude oeste (**Figura 1**) no trecho de aproximadamente 40 km ao longo da Rodovia AC-90 no município de Rio Branco. Foi gerado um *buffer* ao longo da rodovia de 3 km, com superfície territorial de 26.205,86 hectares. A área é caracterizada por paisagens com pastagens cultivadas em propriedades que desenvolvem atividade de pecuária bovina.

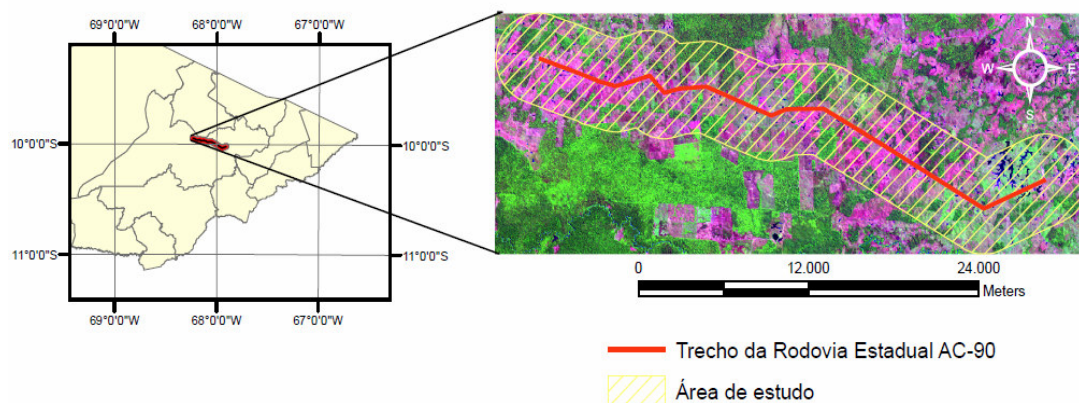


FIGURA 1 Localização da área de estudo no município de Rio Branco, estado do Acre

O clima é do tipo equatorial quente e úmido, caracterizado por altas temperaturas, elevados índices de precipitação pluviométrica e alta umidade relativa do ar. Os solos predominantes são argissolos com horizonte B textural e baixa atividade da argila, muitos deles com alta saturação de alumínio, bem drenados e com relevo plano a suave ondulado. A vegetação predominante é floresta ombrófila aberta com bambu com fasciações de floresta ombrófila aberta com palmeiras.

3. Fonte de dados e classes de mapeamento

Para a realização deste estudo, as fontes de dados foram as imagens multiespectrais do satélite Landsat TM (*Thematic Mapper*), bandas 1 ($0,45 - 0,52 \mu m$), 2 ($0,53 - 0,61 \mu m$), 3 ($0,63 - 0,69 \mu m$), 4 ($0,76 - 0,90 \mu m$), 5 ($1,55 - 1,75 \mu m$) e 7 ($2,08 - 2,35 \mu m$), do ano de 2006, e as imagens derivadas de técnicas de extração de informações, como imagens-fração geradas pela aplicação do modelo linear de mistura espectral e índice de vegetação por diferença normalizada (NDVI), utilizadas na classificação digital do uso e ocupação do solo.

Para a definição das classes de mapeamento do uso e ocupação do solo foram levados em consideração vários fatores: levantamento bibliográfico sobre o mapeamento do uso e

ocupação do solo na Amazônia, atividades econômicas predominantes na área de estudo, dados e sazonalidade das imagens de satélite Landsat. Dessa forma foram definidas sete classes temáticas, cujas principais características são:

Floresta: compreende o domínio da floresta ombrófila aberta, com palmeiras, cipós ou bambus, dependendo das feições fitoecológicas de floresta densa ou de floresta aberta, em função das características das formas de relevo. Nessa classe foi agrupado a sub-classe capoeira caracterizada pela ocorrência de vegetação em vários estágios de sucessão secundária, plantios florestais com seringueira (*Hevea brasiliensis* Muel. Arg.) ou mistos com várias espécies florestais e sistemas agroflorestais com consórcio de culturas perenes e espécies florestais e frutíferas. O agrupamento das subclasses (estágios de sucessão secundária, plantios florestais e agroflorestais) ocorreu devido à heterogeneidade da estrutura da vegetação e à limitação da resolução espacial dos sensores remotos.

Pasto alto: composta por as áreas de pasto limpo com gramíneas exóticas de porte alto e textura mais homogênea.

Pasto degradado: ou pastos sujos, caracterizados por áreas de pastagem com plantas invasoras. Nessa classe temática estão incluídas as pastagens com degradação leve (11%-35% de plantas invasoras ou solos descoberto; pastagem com degradação moderada (36%-60%); e, pastagem degradada (>60%).

Pasto baixo: encontrado nas grandes fazendas de pecuária bovina, formado principalmente pelas gramíneas *Brachiaria brizantha*, *Brachiaria decumbens* e *Panicum maximum*. A pressão de pastejo maior que a capacidade de produção forrageira, principalmente na época da seca, confere o porte mais baixo ao pasto neste período, além da verificação de solo exposto.

Solo: composta de áreas construídas de uso residencial, comercial, industrial e público, e ainda as edificações nas propriedades rurais, as estradas e ramais.

Água: representa os corpos d'água da área de estudo, composta pelo leito dos rios, vários igarapés permanentes e temporários e açudes das propriedades rurais.

4. Técnicas de pré-processamento de imagens digitais

As imagens Landsat foram processadas de acordo com as seguintes etapas:

Correção geométrica: o método de transformação usado foi polinomial de 1º grau e o de reamostragem dos pixels a técnica de alocação do vizinho mais próximo.

Conversão radiométrica: consiste na transformação dos valores digitais brutos (número digital DN) das imagens Landsat para reflectância aparente, com o objetivo exclusivo de viabilizar a aplicação do modelo linear de mistura espectral. A reflectância aparente foi calculada utilizando-se as relações propostas por Markham & Barker (1987), usando as seguintes equações:

$$L(\lambda) = L_{\min \lambda} + \frac{(L_{\max \lambda} - L_{\min \lambda})}{QCAL_{\max \lambda}} QCAL, \text{ em que:}$$

$L(\lambda)$ = radiância espectral, em $\text{mW cm}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$;

$L_{\min \lambda}$ = radiância espectral equivalente ao número digital mínimo;

$L_{\max \lambda}$ = radiância espectral equivalente ao número digital máximo;

$QCAL_{\max \lambda}$ = amplitude de variação dos números digitais, equivalente a 255 para todos os dados TM e ETM+ ; e

$QCAL$ = número digital.

Os valores de radiância obtidos nesta equação são convertidos em reflectância aparente pela seguinte equação:

$$\rho_p = \frac{\pi L_\lambda d^2}{E_{sol\lambda} \cos \theta_s}, \text{ em que:}$$

ρ_p = reflectância aparente, adimensional;

L_λ = radiância espectral ($\text{mW cm}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$);

d = distância Terra-Sol em unidades astronômicas;

$E_{sol\lambda}$ = irradiância solar exoatmosférica média ($\text{mW cm}^{-2} \mu\text{m}^{-1}$); e

θ_s = ângulo zenital solar, em graus.

5. Técnicas de processamento de imagens digitais

A extração de informações das imagens Landsat foi realizada por meio da geração de imagens fração ou sintéticas oriundas da análise de mistura espectral e do índice de vegetação por diferença normalizada (NDVI).

A hipótese básica da mistura espectral é que a variação espectral numa imagem é causada por misturas de um número limitado de materiais na superfície e a sua análise envolve técnicas em que as respostas espectrais misturadas são comparadas com um conjunto de referência espectral pura. O resultado é uma estimativa de proporções aproximadas das áreas em cada pixel que são ocupadas pelas classes de referência.

Os componentes puros solo, sombra e vegetação foram selecionados utilizando a teoria do simplex (Aguiar, 1991; Bajjouk et al., 1998). Posteriormente, foi realizada a decomposição espectral linear, cuja formulação matemática pode ser expressa por Lu et al., 2003:

$$R_i = \sum_{k=1}^n f_k R_{ik} + \varepsilon_i$$

em que:

$i = 1, \dots, m$ (número de bandas espectrais);

$k = 1, \dots, n$ (número de componentes puros);

R_i = reflectância espectral na banda i de um pixel, contendo um ou mais componente puros;

f_k = proporção do endmember k dentro de um pixel;

R_{ik} = reflectância espectral do endmember k dentro do pixel na banda i ; e,

ε_i = erro da banda i .

O NDVI foi usado para transformar os dados multiespectrais do sensor em uma imagem com única banda representando a distribuição da vegetação. Os valores do NDVI indicam a quantidade de vegetação verde presente no pixel, ou seja, quanto maior o valor do NDVI, mais vegetação verde possui o pixel. O NDVI foi gerado por meio do seguinte algoritmo padrão: $\text{NDVI} = (\text{NIR} - \text{R}) / (\text{NIR} + \text{R})$. Em que: NIR = valor do pixel na faixa do espectro eletromagnético correspondente à região do infravermelho próximo (0,7 a 0,9 μm); R = valor do pixel na faixa do espectro eletromagnético correspondente a região do vermelho (0,6 a 0,7 μm).

Na classificação digital foram testadas quatro modalidades, listadas na **Tabela 1**.

TABELA 1 – Modalidades de classificação digital

Modalidade	Fonte de dados	Classificador
1	Imagens do sensor TM do satélite Landsat	Máxima verossimilhança
2	Imagens do sensor TM do satélite Landsat	Isodata
3	Imagens fração solo, sombra e vegetação	Máxima verossimilhança
4	Imagens do Landsat; imagens fração solo, sombra, vegetação; e índice de vegetação NDVI	Árvore de decisão

4. Comparação entre os algoritmos de classificação

Para quantificar os erros gerados pelas diferentes técnicas de classificação utilizadas neste estudo, foram construídas as matrizes de erro ou de confusão, que permitem a comparação entre os dados provenientes das amostras de validação e o resultado da classificação.

Em todas as modalidades de classificação, as matrizes de erros foram calculadas com o mesmo conjunto de amostras de validação. Além da exatidão global ou acurácia do mapeamento, por meio da matriz de erro, foram calculados também os erros de inclusão e de omissão por classe temática e o coeficiente Kappa, que auxilia na avaliação do classificador.

No levantamento do uso e ocupação do solo computou-se a estatística das classes temáticas para todos os métodos de classificação, permitindo dessa forma avaliar o melhor desempenho da classe temática pasto degradado.

Para comparar o desempenho entre os diferentes métodos de classificação foi usada a variância de Kappa e estatística Z, comumente utilizada na literatura (Fidalgo, 1995; Lu et al., 2003). Segundo Skidmore (1999), este teste estatístico pode ser aplicado em combinações pares de matrizes de erro, determinando se as matrizes são significativamente diferentes.

Para se fazer a inferência estatística, utilizou-se o teste de hipótese. Consideraram-se K_1 e K_2 como as estimativas de Kappa e σ_{k1}^2 e σ_{k2}^2 suas respectivas variâncias, representando dois classificadores, respectivamente, para testar a igualdade de duas matrizes de confusão e a estatística Z, expressa por:

$$Z = \frac{|K_1 - K_2|}{\sqrt{\sigma_{k1}^2 + \sigma_{k2}^2}}$$

As hipóteses para o teste de significância foram: $H_0: K_1 = K_2$ (hipótese nula) e $H_1: K_1 > K_2$ (hipótese alternativa). A hipótese H_0 foi testada para os níveis de significância (α) de 0,01; 0,05 e 0,10. O valor tabelado de Z foi o da hipótese alternativa unilateral à direita. Assim, à significância α , H_0 será rejeitada se $Z > Z_\alpha$.

4. Resultados e Discussão

Em relação à exatidão global do mapeamento geral, a classificação supervisionada apresentou maior acurácia em relação à técnica de classificação não supervisionada, alcançando 83,17% na classificação por máxima verossimilhança usando as imagens derivadas do modelo linear de mistura espectral.

Na aplicação do teste Z, ao nível de confiança de 90%, as classificações por máxima verossimilhança (com imagens Landsat e com imagens fração) apresentaram diferença significava em relação ao algoritmo isodata.

Esse resultado é evidenciado na **Figura 2**, na qual se verifica que a classificação por máxima verossimilhança usando as imagens Landsat e as imagens fração geradas pelo análise de mistura espectral, e por árvore de decisão, apresentam os maiores valores do coeficiente de concordância Kappa, variando de 0,78 a 0,80, consideradas uma qualidade de classificação muito boa, de acordo com Landis & Koch (1977). Na modalidade de classificação usando o algoritmo isodata, o valor de Kappa foi de 0,74.

A matriz de erro ou de confusão foi usada para determinar a exatidão para cada categoria de uso e ocupação do solo mapeada, por meio da estimativa dos erros de omissão e de inclusão, e a exatidão do produtor e do usuário, respectivamente

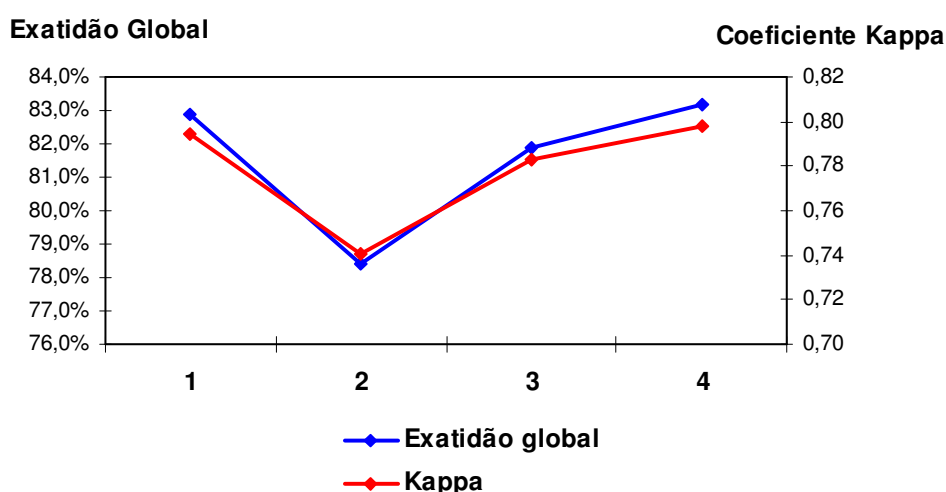


FIGURA 2 Exatidão Global e Índice de concordância Kappa das quatro técnicas de classificação digital para o mapeamento do uso e ocupação do solo em Rio Branco, Acre, no ano de 2006: (1) máxima verossimilhança usando imagens Landsat, (2) isodata usando imagens Landsat, (3) árvore de decisão, (4) máxima verossimilhança usando imagens fração solo, sombra e de vegetação.

As classes temáticas floresta e água apresentaram excelente acurácia em todas as modalidades de classificação, com percentual de exatidão variando no intervalo de 89% a 100%. A classe solo exposto também apresentou bom desempenho na sua exatidão, a maior confusão foi com a classe de pasto baixo que é caracterizada por áreas de pós-pastejo aliado ao período seco na região. Essas condições resultam numa maior quantidade de solo exposto em áreas de pastagens cultivadas.

O resultado global de todas as classes mapeadas pode ser verificado na **Tabela 2**.

TABELA 2 Acurácia do usuário e acurácia do produtor das classes de mapeamento do uso e ocupação do solo em Rio Branco, Acre, no ano de 2006, segundo a classificação por : (1) isodata usando imagens Landsat; (2) máxima verossimilhança usando imagens Landsat, (3) máxima verossimilhança usando imagens fração solo, sombra e de vegetação, e (4) árvore de decisão.

Classe temática / Acurácia	Metodologia de Classificação							
	1		2		3		4	
	Acurácia							
	Usuário	Produtor	Usuário	Produtor	Usuário	Produtor	Usuário	Produtor
Água	100,00%	96,00%	100,00%	98,00%	100,00%	98,00%	100,00%	100,00%
Floresta	89,23%	100,00%	92,06%	100,00%	92,06%	100,00%	98,31%	100,00%
Pasto degradado	90,91%	19,23%	91,67%	42,31%	88,46%	44,23%	100,00%	25,00%
Pasto Alto	60,53%	88,46%	62,67%	90,38%	63,01%	88,46%	57,30%	98,08%
Pasto Baixo	68,18%	84,91%	84,31%	81,13%	86,00%	81,13%	74,65%	100,00%
Solo Exposto	81,63%	80,00%	79,25%	84,00%	79,63%	86,00%	100,00%	100,00%

A acurácia do produtor do mapeamento das pastagens degradadas, cujo resultado percentual indica a probabilidade da amostra de referência ter sido classificada corretamente, os melhores resultados foram obtidos na classificação por máxima verossimilhança, em média, 43,3%. Na classificação não supervisionada o resultado obtido com o algoritmo isodata a acurácia foi de apenas 19,2%. A classificação por árvore de decisão melhorou o resultado somente em relação à classificação não supervisionada.

A variação na exatidão da classe de pastagem degradada ocorreu devido à semelhança espectral na fronteira de decisão entre as classes de pasto degradado, pasto alto e pasto baixo, podendo ocorrer a sobreposição em alguma delas, e a habilidade de cada classificador em estabelecer o limite de decisão no espaço de feições.

A confusão entre as classes de mapeamento, causada pelos erros de inclusão ou omissão nas diferentes técnicas de classificação, constituem num dos fatores que levam as diferentes estimativas da proporção de área ocupada por essas classes. Na área de estudo, o percentual de área ocupada por pastagem degradada foi de 11,23%, que totaliza 2.957,85 hectares. A maioria das pastagens degradadas são de pequena dimensão até 1 hectare (75%) que é explicado pela presença de pequenos produtores rurais localizados em projetos de assentamento e colonização do INCRA. Cerca de 23% dos polígonos de pastagem degradada mapeados estão entre 1 a 10 hectares e as maiores áreas, chegando até 99,36 hectares representam apenas 2% do total de polígonos mapeados nessa classe temática.

Considerando-se os resultados, a diferença entre as estimativas de áreas das classes de mapeamento de uso e ocupação do solo é consequência das diferentes técnicas de classificação e da habilidade e experiência do analista. Dessa forma, é importante considerar os desempenhos da exatidão global e por classe temática entre as diferentes técnicas de classificação, bem como a importância das categorias de uso e ocupação do solo que se deseja mapear para selecionar o método mais adequado ao objetivo do trabalho a ser realizado.

5. Conclusões

Os resultados obtidos neste trabalho de comparação entre técnicas de classificação de imagens Landsat no mapeamento de pastagens degradadas em Rio Branco, Acre, permitem as seguintes conclusões:

- os diferentes métodos de classificação com os algoritmos de máxima verossimilhança, isodata e árvore de decisão, são eficientes para o mapeamento de uso e ocupação do solo em Rio Branco, pois apresentaram uma classificação muito boa a excelente;
- entre os métodos testados, a classificação por máxima verossimilhança foi superior às demais técnicas de classificação;
- a utilização das imagens derivadas do modelo linear de mistura espectral na classificação supervisionada por máxima verossimilhança melhorou a exatidão global do mapeamento;
- as classes floresta, água, solo, pasto alto e pasto baixo obtiverem bom desempenho em todas as técnicas de classificação aplicadas;
- a classe pasto degradado apresentou confusão entre os classificadores principalmente com as classes de pasto alto e pasto baixo, devido à semelhança espectral entre elas;
- para o mapeamento de pastagem degradada, considerando a acurácia do produtor, a técnica de classificação supervisionada por máxima verossimilhança representa o melhor desempenho.

4. Citações e Referências

ACRE. Secretaria de Estado Planejamento e Coordenação. Programa Estadual de Zoneamento Ecológico-Econômico do Estado do Acre. Zoneamento **ecológico-econômico: aspectos sócio-econômicos e ocupação territorial – documento final**. Rio Branco: SECTAM, 2000. 3 v.

AGUIAR, A.P.D. **Utilização de atributos derivados de proporções de classes dentro de um elemento de resolução de imagem (“pixel”) na classificação multispectral de imagens de sensoriamento remoto**. 1991. 227p. Dissertação (Mestrado em Sensoriamento Remoto e Aplicações)-Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP. 1991.

BAJJOUK, T.; POPULUS J.; GUILLAUMONT, B. Quantification of subpixel cover fractions using principal component analysis and a linear programming method: application to the Coastal Zone of Roscoff (France). **Remote Sensing of Environmental**, v.64, p.153-165, 1998.

DIAS-FILHO, M.B.; ANDRADE, C.M.S. Pastagens no ecossistema do trópico úmido. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 42., 2005, Goiânia. **Anais...** Goiânia: SBZ, 2005. p. 95-104.

FIDALGO, L.C.C. **Exatidão no processo de mapeamento temático da vegetação de uma área de mata atlântica no estado de São Paulo, a partir de imagens TM-Landsat**. 1995. 184p. Dissertação (Mestrado em Sensoriamento Remoto e Aplicações)-Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP. 1995.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Sidra – Banco de dados agregados. Brasília: IBGE, 2005. Disponível em: <<http://www.sidra.ibge.gov.br/bda/>>. Acesso em: 10/08/2005.

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

LU, D.; MORAN, E. BATISTELLA, M. Linear mixture model applied to Amazonian vegetation classification. **Remote Sensing of Environment**. v.87. p.456-469. 2003.

MARKHAM, B.L.; BARKER, J.L. **Radiometric properties of U.S. processes Landsat MSS data**. *Remote Sensing of Environment*, v.17, p.39-71, 1987.

SKIDMORE, A.K. Accuracy assessment of spatial information. In: STEIN, A.; MEER, F. van der; GORTE, B. (Ed.). **Spatial statistics for remote sensing**. Dordrecht: Kluwer, 1999. p.197-209.