

# CLASSIFICAÇÃO DA COBERTURA DA TERRA NA REGIÃO AMAZÔNICA ATRAVÉS DE CLASSIFICAÇÃO DE REGIÕES EM IMAGENS DE PROPORÇÃO DE COMPONENTES

JORGE LUIS GAVINA PEREIRA<sup>1</sup>

GETULIO T. BATISTA<sup>1</sup>

DAR A. ROBERTS<sup>2</sup>

<sup>1</sup>Divisão de Sensoriamento Remoto (DSR)

Instituto Nacional de Pesquisas Espaciais (INPE)

Caixa Postal 515, 12201-097 – São José dos Campos, SP, Brasil

gavina@ltid.inpe.br getulio@ltid.inpe.br

<sup>2</sup>UCSB - University of California

Department of Geography, EH3611

Santa Barbara, CA, 93106, USA

dar@geog.ucsb.edu

**Abstract:** The Amazon rain forest plays an important role in the global climate. The dynamics of land use and cover change (LUCC) has a major impact on the global carbon and water cycles. However, current figures on carbon budget associated with LUCC are not well known for the Amazon, except for annual total deforested area. Remote sensing associated with geographic information systems has the potential to improve the estimates of the areal extent of the different cover classes and therefore, improve the knowledge about carbon stocks. This paper classifies land covers of a region in Eastern Amazon using a supervised classification algorithm applied to regions obtained by segmentation of fraction images derived from Landsat TM sensor. It was found that 61% of the study area (~2.000 km<sup>2</sup>) is in forest, 37% is deforested and the remaining area (2%) is dominated by *babaçu* and water bodies. Primarily pasture (80%) and secondary growth forest of different ages (20%) composed the deforested area.

**Keywords:** Amazon, Land Use, Remote Sensing, Spectral Mixture Model, Segmentation and Classification.

## 1 Introdução

A Floresta Amazônica, com suas dimensões continentais, tem um importante papel no sistema climático regional, exercendo uma função essencial nas trocas de massa e energia com a atmosfera. Por este motivo, são crescentes os interesses da comunidade científica nacional e internacional na modelagem dos ciclos hidrológico e de carbono da Amazônia. As emissões líquidas de carbono associadas às mudanças de cobertura da terra são motivos de grandes especulações, pois dependem do tamanho da área desmatada, da biomassa da vegetação original, da biomassa da nova cobertura e do destino do carbono da área desmatada. Na Região Amazônica, devido às suas dimensões e a dificuldade de trabalhos de campo, o sensoriamento

remoto associado a sistemas de informação geográfica (SIG's) apresentam-se como ferramentas valiosas para o monitoramento das alterações da cobertura da terra.

Este trabalho tem como objetivo a classificação da cobertura da terra, de uma região ao sul da cidade de Marabá-PA (**Figura 1**), aplicando-se um algoritmo de classificação supervisionado por regiões a imagens de Proporções de Componentes (Vegetação, Solo, Sombra e Material vegetal não-fotossintético) geradas a partir de um modelo linear de mistura espectral, bem como o cálculo e análise da área de cada uma das classes definidas *a priori*.

A área abrangida por este estudo localiza-se ao sul da cidade de Marabá, centro-leste do Estado do Pará. Esta região está inserida na fronteira agrícola, apresentando desta forma uma rápida mudança da sua paisagem natural. A conversão da floresta para pastagens tem sido a causa principal de tal mudança, principalmente a partir do final da década de 1960. A área estudada possui aproximadamente 2000 km<sup>2</sup>, e está localizada entre as coordenadas geográficas 48°55'13" - 49°16'31" W e 5°35'13" - 6°03'14" S. Esta área é cortada no sentido norte - sul pela rodovia PA - 150, e pelos rios Vermelho e Sororó (**Figura 1**). Na área de influência desta rodovia, mais notadamente de Marabá até a divisa do Estado do Pará com Mato Grosso, instalaram-se grandes projetos agropecuários, com incentivos fiscais do governo. Logo, a área de estudo apresenta duas classes principais de cobertura da terra: florestas e pastagens. Contudo, são encontradas também pastagens degradadas (pasto sujo), agricultura itinerante e capoeiras em áreas de pastagem abandonadas e nas roças durante o estágio de pousio.

## **2 Materiais e métodos**

Neste trabalho foi utilizado o quadrante C da imagem Landsat TM órbita-ponto 223/64, do dia 20 de julho de 1993, com deslocamento de 27 Km para o norte, bandas 1 a 5, e 7. A partir das 6 bandas Landsat TM originais foram geradas as imagens proporções dos seguintes componentes: vegetação, sombra, solo e material vegetal não-fotossintético. Estas imagens foram posteriormente segmentadas e sobre as regiões geradas foi aplicado um algoritmo de classificação supervisionado.

### **2.1 Proporção de Componentes**

São inúmeros os trabalhos encontrados na literatura abordando aplicações e metodologia de geração de imagens de proporção de componentes decompostas por um modelo linear de mistura espectral (Adams et al., 1989 e 1993; Roberts, 1991; Aguiar, 1991; Pereira, 1996).

O procedimento para a conversão dos dados da imagem original para reflectância e, posteriormente, para o cálculo da proporção de componentes, para cada pixel, foi baseado no método proposto por Roberts (1991). Este método baseia-se no uso de componentes puros (*endmembers*) de referência (biblioteca espectral), relacionando-os aos componentes puros da imagem, e não apenas na escolha de elementos puros selecionados na própria imagem como frequentemente utilizado em vários trabalhos (Shimabukuro e Smith, 1991; Shimabukuro et al., 1997).

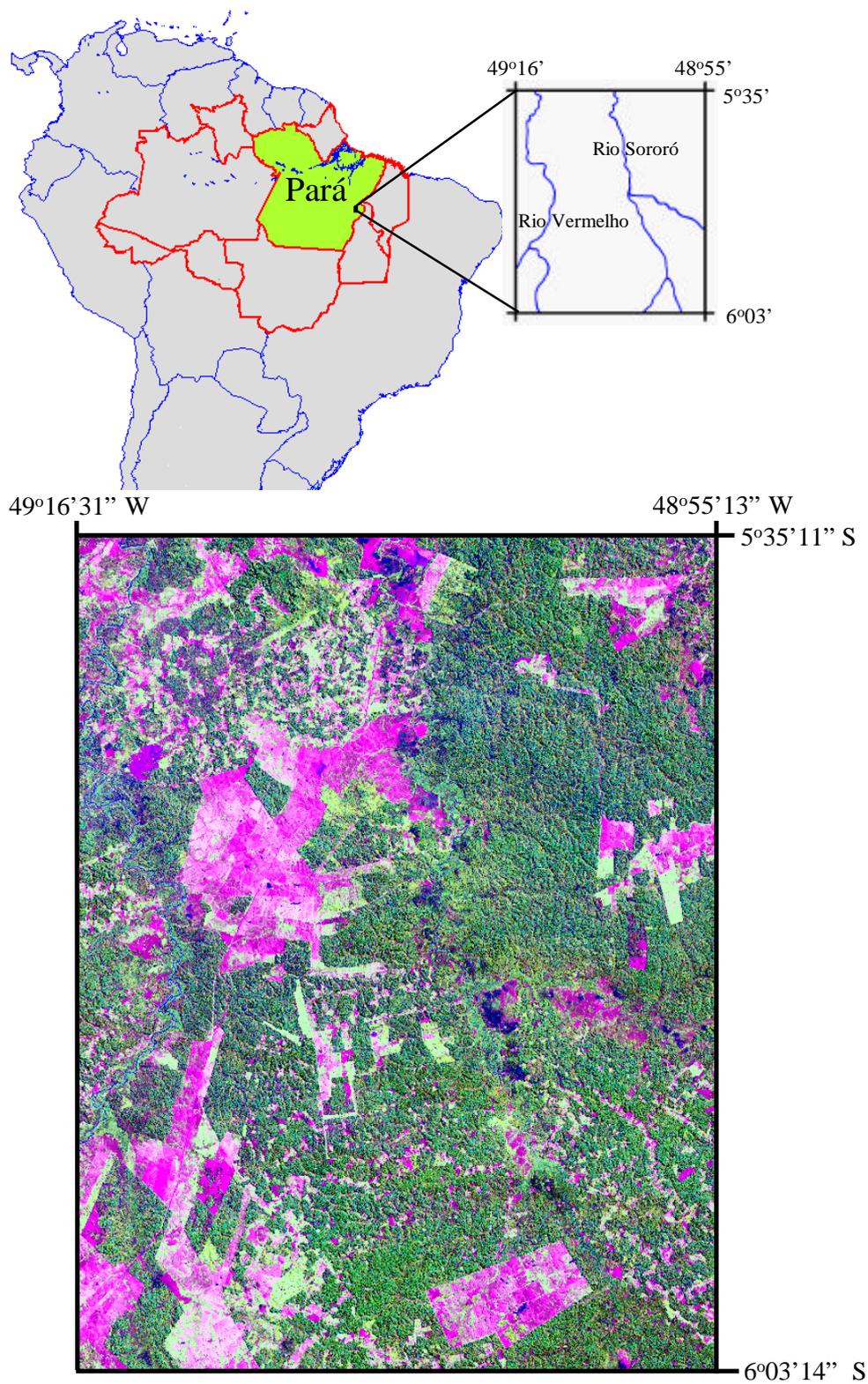


Fig. 1 - Área de estudo - bandas do Landsat TM: TM3 (azul), TM4 (verde) e TM5 (vermelho).

Após a aplicação do modelo, foram estimadas as proporções  $F_i$ ,  $i = 1, \dots, N$  (número de componentes) correspondentes a cada pixel e para cada componente é gerada uma banda sintética linearmente relacionada às proporções estimadas. Neste trabalho foram geradas imagens proporções para os componentes vegetação, sombra, solo e material vegetal não-fotossintético.

## 2.2 Segmentação de Imagens

Segundo Gonzalez e Wintz (1987), a segmentação de imagens é um processo que subdivide uma imagem em suas partes constituintes ou objetos. Os algoritmos de segmentação geralmente são baseados em uma das duas propriedades básicas dos valores de níveis de cinza: descontinuidade e similaridade. Na primeira categoria, a partição de uma imagem baseia-se em mudanças abruptas dos níveis de cinza. As principais áreas de interesse dentro desta categoria são a detecção de pontos isolados, de linhas e de bordas numa imagem. Os principais acessos da segunda categoria são baseados em limiares, crescimento de regiões, partição e fusão de regiões. Como subentendido pelo seu nome, crescimento de regiões é um procedimento que agrupa pixels ou sub-regiões em regiões maiores. O mais simples destes procedimentos é a agregação de pixels, iniciando-se com uma série de pontos "sementes", e destes crescem regiões pela anexação de pixels vizinhos com propriedades como nível de cinza e textura semelhantes (Gonzalez e Wintz, 1987).

No presente trabalho foi utilizado o procedimento descrito por Batista et al. (1994) e por Bins et al. (1996) para a segmentação das imagens proporções. Este procedimento, implementado no SPRING, baseia-se num algoritmo de crescimento de regiões que agrega pixels com propriedades similares. Desta forma, uma região é uma coleção de pixels conectados de acordo com suas propriedades.

O algoritmo de segmentação por crescimento de regiões em questão, necessita da definição pelo usuário dos seguintes parâmetros:

- a) **Similaridade:** limiar sob o qual duas regiões são consideradas similares e portanto agrupadas. Ele representa a distância euclidiana entre as médias das duas regiões; e
- b) **Área:** área mínima a ser considerada como uma região, definida como o número de pixels.

Após uma exaustiva experimentação com vários valores de limiares (similaridade e área), foram selecionados para esta área de estudo os valores 8 para o limiar de similaridade e 25 para o de área.

A utilização de um limiar de área de 25 pixels resulta em regiões com no mínimo 22.500 m<sup>2</sup>, o que é menos de 0,5 alqueire da região (o alqueire da região corresponde a 48.400 m<sup>2</sup>). Este tamanho mínimo parece ser bastante apropriado para a região, pois as pequenas roças (roças de posseiros) raramente possuem menos de 8 alqueires, sendo comum o manejo anual de áreas de 2 alqueires. A utilização do limiar de similaridade 8, teve como base a adequação das bordas (regiões) das capoeiras visitadas em campo.

### 2.3 Classificação de imagens

A análise de imagens digitais tem sido realizada, em muitos casos, pelo método convencional de classificação (pixel-a-pixel), ou seja, a identificação do objeto é feita levando-se em conta apenas características pontuais, não levando em consideração o contexto onde o pixel está inserido (vizinhança). Um procedimento alternativo à classificação pontual (pixel-a-pixel) é aquele em que a imagem é subdividida em regiões de onde são extraídos atributos espectrais, geométricos e contextuais, os quais são usados na classificação de cada região. Esta abordagem é uma tentativa de incorporar outros tipos de informação, além das espectrais, usualmente utilizadas, para melhorar os resultados da classificação (Bins et al., 1992). Um classificador de pixel usa apenas a média e a variância da intensidade espectral (níveis de cinza), ao passo que um classificador de segmento pode usar, além da média e variância espectrais, parâmetros que descrevem a forma, o tamanho, e o contexto do segmento (Belaid et al., 1992).

A classificação de uma imagem segmentada pode ser implementada através de métodos supervisionados, não-supervisionados e/ou híbridos. No SPRING, encontram-se implementados dois algoritmos de classificação por regiões: não-supervisionado - ISOSEG; e supervisionado – Bhattacharyya.

Richards (1986) cita que a classificação supervisionada é o procedimento mais utilizado para análises quantitativas de dados de imagens de sensoriamento remoto. Nesta técnica de classificação há uma grande interação do analista com o sistema de processamento das imagens, sendo necessário ainda que o analista tenha conhecimento da área a ser classificada.

Independente do algoritmo escolhido, a seqüência essencial para a classificação supervisionada envolve:

- a) definir o número de classes de cobertura do solo que a imagem será particionada;
- b) escolher pixels representativos de cada uma das classes desejadas, denominados de amostra de treinamento;
- c) usar dados de treinamento para estimar os parâmetros do algoritmo de classificação; estes parâmetros são as propriedades do modelo de probabilidade a ser usado para a partição da imagem no espaço multiespectral;
- d) usar o classificador treinado para associar todos os pixels da imagem a uma das classes definidas *a priori*; e
- e) produzir resumos tabulares ou mapas temáticos que sintetizem os resultados da classificação.

Na fase de treinamento do algoritmo para a classificação supervisionada, definem-se dois tipos de conjuntos de amostras de treinamento: amostras de aquisição e amostras de teste. As amostras de aquisição definirão os parâmetros estatísticos das classes: vetor de médias e matriz de covariância. As amostras de teste servem para verificar o desempenho do classificador, antes de utilizá-lo na classificação da imagem. A partir da classificação das amostras de teste, com base nos parâmetros gerados pelas amostras de aquisição, é gerada uma matriz de confusão, com todas as classes definidas *a priori*. Os valores da diagonal principal da matriz representam os pixels corretamente classificados, e os pixels fora desta diagonal, representam confusões entre classes.

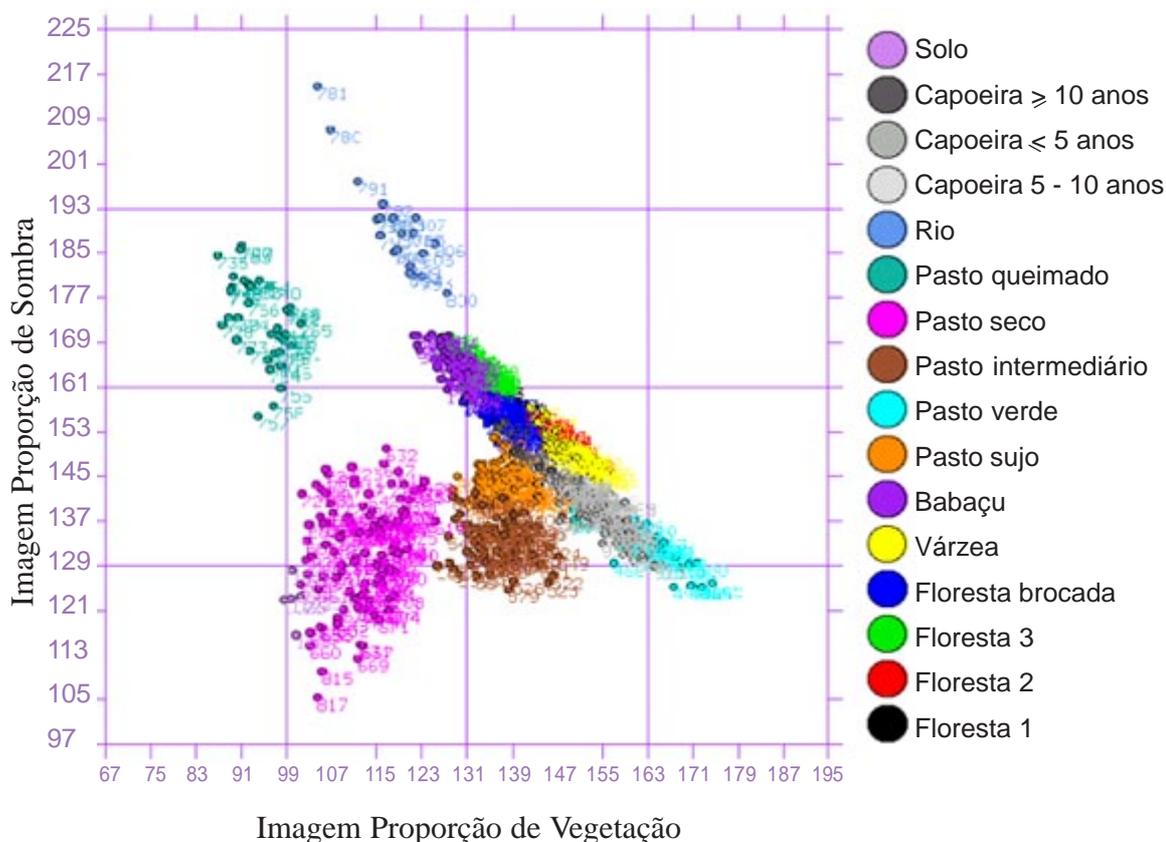


Fig. 2 - Distribuição das amostras de treinamento das classes de cobertura do solo, em um espaço de atributos bidimensional: Imagem Proporção de Vegetação x Imagem Proporção de Sombra.

A **Tabela 1** mostra a matriz de confusão entre a classificação da imagem e a referência, definida por verdade terrestre (trabalho de campo), referente às amostras de treinamento, após o agrupamento das subclasses espectrais de floresta, pasto limpo e capoeiras. As percentagens das classes corretamente classificadas (diagonal da matriz de confusão) correspondem à área das regiões das amostras de treinamento corretamente classificadas em relação a área total das amostras de treinamento de cada classe.

Com o agrupamento das subclasses de floresta, pasto limpo e capoeiras, o desempenho médio da classificação sobe para 98,6%, ficando a confusão média em 1,4%. O cálculo do desempenho médio para as classes agrupadas foi realizado ponderando-se a área de cada classe. A elevada percentagem de área corretamente classificada, deve-se ao elevado percentual de classificação correta de classes com áreas grandes, como é o caso das classes floresta e pasto limpo com 99,8 e 98,3%, respectivamente, da área corretamente classificada, que representam juntas 92% do total da área das amostras de teste.

O algoritmo de classificação supervisionada por regiões implementado no SPRING, utiliza como critério de decisão estatística a distância de Bhattacharyya, frequentemente utilizada na seleção de atributos para medir a distância entre duas distribuições. Mascarenhas e Velasco (1984) descrevem em detalhe o algoritmo de classificação com base na distância de Bhattacharyya.

A definição das classes de cobertura da terra de interesse para a área de estudo, baseou-se no trabalho de campo. Foram definidas as seguintes classes:

- a) Floresta: áreas de floresta sem antropização - reserva florestal da Companhia Vale do Rio Doce;
- b) Babaçu: áreas de floresta com dominância de babaçu;
- c) Pasto limpo:
  - seco - áreas de pasto com elevada resposta de solo e/ou material vegetal não-fotossintético;
  - verde - pasto vigoroso com elevada resposta de vegetação;
- d) Pasto sujo: pastagens com presença significativa de espécies invasoras, em geral ervas e arbustos;
- e) Capoeiras com menos 10 anos;
- f) Capoeiras com mais de 10 anos;
- g) Rio.

As amostras de treinamento foram obtidas a partir das regiões criadas pela segmentação das bandas de proporções dos componentes citados anteriormente.

### 3 Resultados e discussão

Das 8 classes definidas *a priori* (floresta, babaçu, pasto limpo, pasto sujo, capoeiras com menos de 10 anos, capoeiras com mais de 10 anos - Capoeira  $\geq 10$  anos, e rio) as classes floresta, pasto limpo e capoeira com menos de 10 anos, devido às suas heterogeneidades e para satisfazer a exigência do algoritmo de classificação de que a distribuição de cada classe seja *gaussiana*, foram divididas em subclasses espectrais, posteriormente agrupadas (floresta: Floresta 1, Floresta 2, Floresta 3, Floresta brocada e Várzea; pasto limpo: Pasto verde, Pasto intermediário, Pasto seco, Pasto queimado e Solo; capoeiras com menos de 10 anos; Capoeira  $\leq 5$  anos e Capoeira 5-10 anos). Pela análise dos histogramas de distribuição dessas classes verificou-se que o critério de distribuição *gaussiana* foi satisfeito.

A distribuição das amostras de treinamento (aquisição e teste) num espaço de atributos bidimensional pode ser observada na **Figura 2**. Algumas classes encontram-se tão próximas que é impossível observá-las neste gráfico. É interessante também observar a correlação negativa para a maioria das classes entre as imagens vegetação e sombra.

TABELA 1 - MATRIZ DE CONFUSÃO DA CLASSIFICAÇÃO DA IMAGEM

Classificação Referência	Floresta	Babaçu	Pasto Sujo	Pasto Limpo	Rio	Capoeiras com menos de 10 anos	Capoeiras com mais de 10 anos	Pixels	Área (%)
Floresta	<b>99,8</b>						0,2	129.515	79,9
Babaçu	18,7	<b>81,3</b>						3.178	2,0
Pasto Sujo			<b>86,6</b>	8,2		5,2		3.712	2,3
Pasto Limpo			1,7	<b>98,3</b>				19.686	12,1
Rio					<b>100,0</b>			399	0,2
Capoeiras com menos de 10 anos			5,6	0,6		<b>93,0</b>	0,8	4.640	2,9
Capoeiras com mais de 10 anos						23,8	<b>76,2</b>	980	0,6

Com exceção da classe Rio, as demais apresentam algumas confusões. Classes como floresta e pasto limpo apresentam confusões pequenas: a primeira com capoeira com mais de 10 anos - 0,2%, uma vez que capoeiras mais velhas desenvolvem uma estrutura complexa, assemelhando-se espectralmente à floresta; e a segunda com pasto sujo - 1,7%, devido exclusivamente à confusão entre pasto sujo e a subclasse pasto intermediário. Uma outra classe que possui uma elevada percentagem de acerto (93%) é a classe de capoeiras com menos de 10 anos. Esta classe possui uma confusão maior com a classe pasto sujo (5,6%), uma vez que as áreas de pasto sujo são caracterizadas pela presença de indivíduos arbustivos e arbóreos, iniciando o processo de regeneração da vegetação; e confusões pequenas com as classes capoeira com mais de 10 anos - 0,8%, e com pasto limpo - 0,6%, dada a semelhança espectral de capoeiras com predominância de *Cecropia* com pastagens vigorosas (**Tabela 1**).

As classes que possuem percentagem de áreas corretamente classificadas menores que 90% são: Pasto sujo - 86,6%; Babaçu - 81,3%; e Capoeira com mais de 10 anos - 76,2%.

As áreas de babaçu confundem-se unicamente com áreas de floresta (18,7%), sobretudo com a subclasse floresta brocada. Esta subclasse também apresenta respostas espectrais baixas (bandas TM), associadas à ação do fogo; ou seja, ambas são áreas escuras (elevada proporção de sombra). O babaçu (*Orbignya phalerata*) é uma palmeira comum nas florestas da região de estudo, mas que devido a algum distúrbio, especialmente queimadas, passam a dominar uma área. O fato das áreas de babaçu aparecerem mais escuras pode estar relacionado com a arquitetura do dossel da palmeira (predominância de folhas verticais - reflexão especular), e com sombreamento que as palmeiras emergentes exercem sobre o dossel da floresta.

O pasto sujo confunde-se tanto com pasto limpo (8,2%), devido à confusão com a subclasse pasto intermediário, quanto com capoeiras com menos de 10 anos (5,2%), dada a sua semelhança com as capoeiras mais novas. Após os agrupamentos, a classe com maior percentual de confusão é a capoeira com mais de 10 anos, que apresenta 23,8% da área das amostras de treinamento

classificada como capoeiras com menos de 10 anos. Esta confusão é de certa forma esperada, pois ambas são classes de capoeiras.

As florestas respondem por aproximadamente 61% da área estudada, enquanto que a área desflorestada (pastos e capoeiras) correspondem a cerca de 37%. As áreas com dominância de babaçu representam 1,7% do total da área e os rios respondem apenas por 0,3%. Com relação a área desmatada, as áreas de pasto limpo são predominantes, correspondendo a 56,8%, sendo as áreas de pasto sujo a segunda em importância, com 23%, e por fim as capoeiras representam apenas 20,2% da área desmatada (**Figuras 3 e 4**).

#### 4 Conclusões

A classificação da cobertura da terra usando o algoritmo de classificação supervisionado por regiões a partir de imagens de proporção de componentes (vegetação, solo, sombra e material vegetal não-fotossintético) derivadas do Landsat TM, segmentadas pelo processo de crescimento de regiões, resultou em exatidão elevada (98,6% de acerto), indicando a adequação deste método para classificação do uso da terra nesta região. Essa exatidão elevada se deve também ao fato de ter sido estimada com base nas amostras de treinamento e não em áreas testes. O resultado do mapeamento com validação de campo revelou que 37% da área de estudo foi desmatada até o ano de 1993. Cerca de 20% dessa área desmatada foi abandonada e se encontra em capoeiras de diferentes idades. Do restante da área desmatada, cerca de 57% encontram-se preservados como pasto limpo com boa capacidade de suporte.

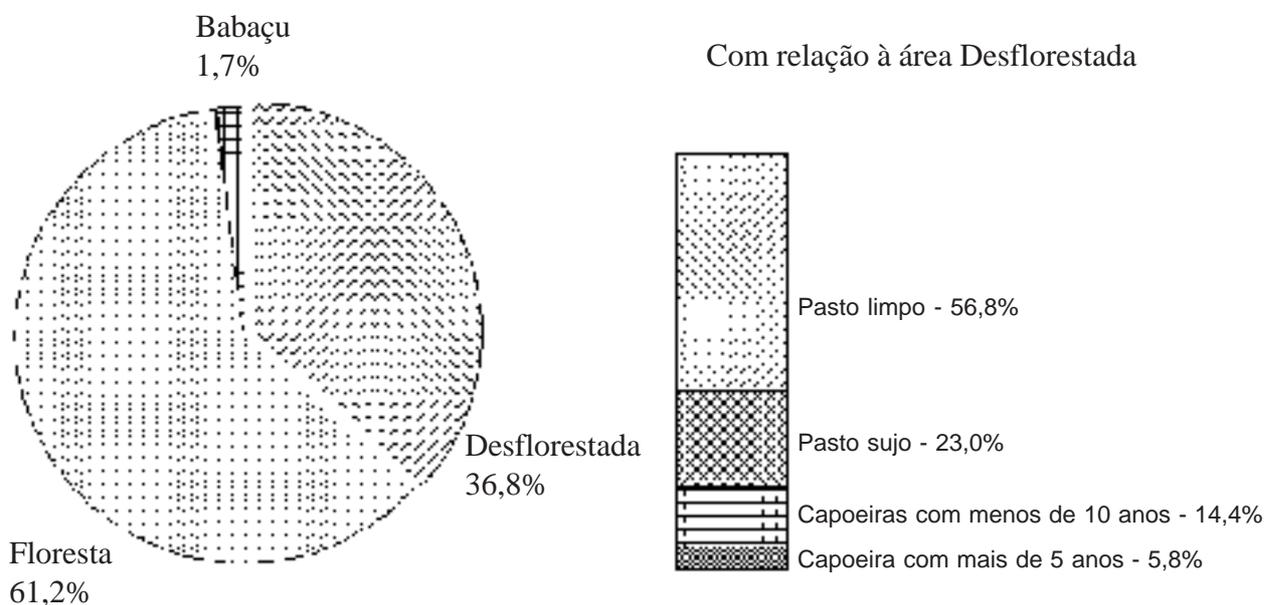


Fig. 3 - Distribuição da área de estudo entre as principais classes de cobertura.

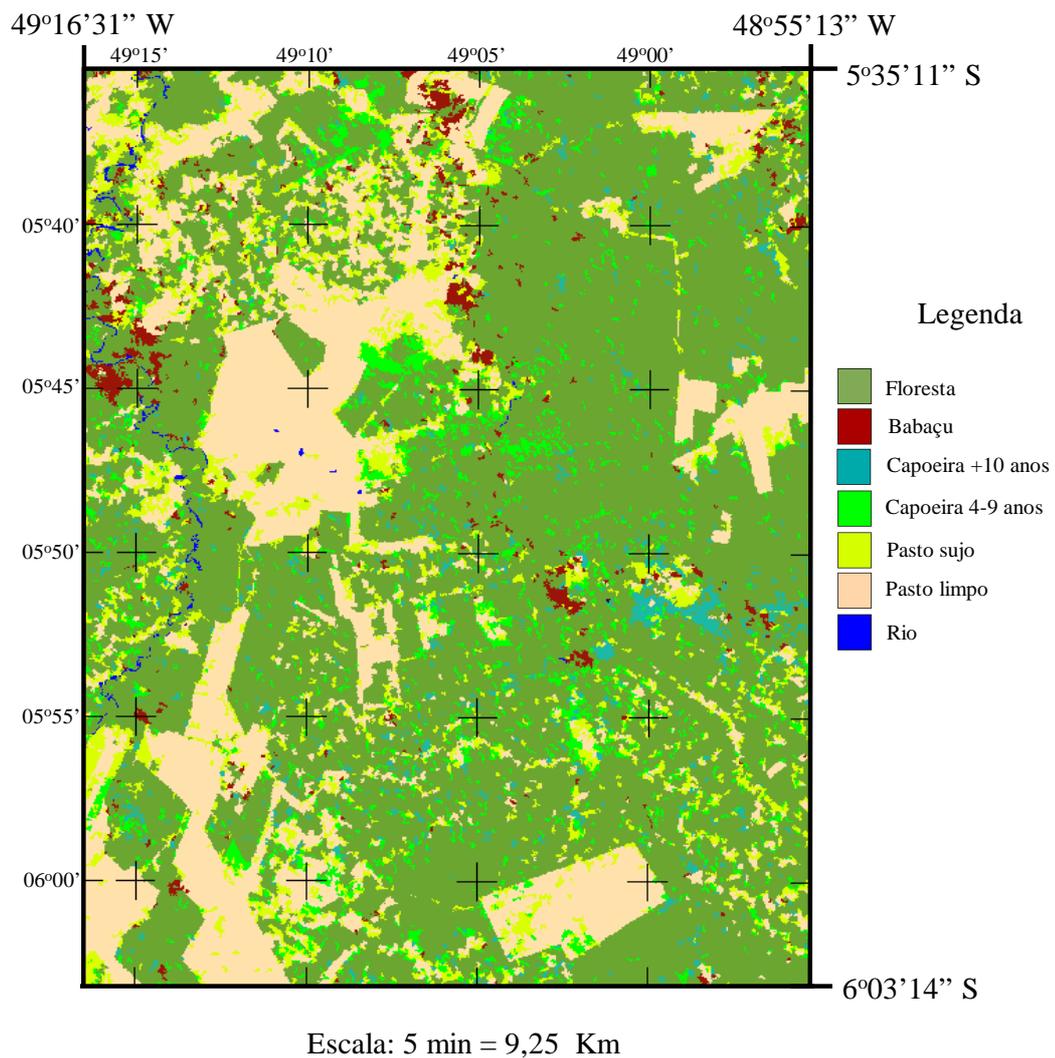


Fig. 4 - Mapa da cobertura da terra na área de estudo resultante da classificação supervisionada por regiões.

### Agradecimentos

Os autores deste trabalho são gratos a FAPESP pelo apoio financeiro desta instituição (Projeto FAPESP – N<sup>o</sup> 94/2706-5).

## Referências

- Adams J.B.; Smith, M.O.; Gillespie, A.R. Simple models for complex natural surfaces: a strategy for the hyperspectral era of remote sensing. In: IGARSS'89: Canadian Symposium on Remote Sensing, 12., Vancouver, 1989. **Remote Sensing: an economic tool for the nineties**. Vancouver, IEEE, 1989, v. 1, p. 16-21.
- Adams J.B.; Smith, M.O.; Gillespie, A.R. Imaging spectroscopy: interpretation based on spectral mixture analysis. In: Pieters, C.M.; Englert, P., eds. **Remote Geochemical Analysis: elemental and mineralogical composition**. New York, Cambridge University Press, 1993. cap. 7, p. 145-166.
- Aguiar, A.P.D. **Utilização de atributos derivados de proporções de classes dentro de um elemento de resolução de imagem (“pixel”) na classificação multiespectral de imagens de sensoriamento remoto**. (Dissertação de Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1991. 202 p. (INPE-5541-TDI/529).
- Batista, G.T.; Medeiros, J.S.; Mello, E.M.K; Moreira, J.C.; Bins, L.S. A new approach for deforestation assessment. In: International Symposium on Resource and environmental monitoring, Rio de Janeiro, 1994. **Proceedings**. Rio de Janeiro, ISPRS Commission VII, 1994, v. 30, p. 7a, p.170-174.
- Belaid, M.A.; Edwards, G.; Jatou, A.; Thomson, K.P.B.; Beaulieu, J.M. Post-segmentation classification of images containing small agricultural fields. **Geocarto International**, 3:53-60, 1992.
- Bins, L.S.; Erthal, G.J.; Fonseca, L.M.G. Um método de classificação não-supervisionado por regiões. In: Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens, 5., Águas de Lindóia, 1992. **Anais**. São José dos Campos, INPE, 1992, v. 2, p. 65-68.
- Bins, L.S.; Fonseca, L.M.G.; Erthal, G.J.; Ii, F.M. Satellite imagery segmentation: a region growing approach. In: Simpósio Brasileiro de Sensoriamento Remoto, 8., Salvador, 1996. **CD-ROM - Seção de Comunicações**. São Paulo, Imagem@multimídia, 1996.
- Gonzalez, R.C.; Wintz, P. **Digital image processing**. 2.ed. Reading, Addison-Wesley, 1987. 503 p.
- Mascarenhas, N.D.D.; Velasco, F.R.D. **Processamento Digital de Imagens**. São Paulo, Instituto de Matemática e Estatística - USP, 1984.
- Pereira, J.L.G. **Estudos de áreas de florestas em regeneração através de imagens Landsat TM**. (Dissertação de Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1996. 137p. (INPE-5987-TDI/578).
- Richards, J.A. **Remote sensing digital image analysis: an introduction**. Berlin, Springer-Verlag, 1986. 281 p.
- Roberts, D.A. **Separating spectral mixtures of vegetation and soils**. (Ph.D. Thesis in Geological Sciences) - University of Washington, Seattle, 1991.
- Shimabukuro, Y.E.; Smith J.A. The least-squares mixing models to generate fractions images derived from remote sensing multispectral data. **IEEE Transactions on Geoscience and Remote Sensing**. 29(1):16-20, Jan., 1991.

Shimabukuro, Y.E.; Batista, G.T.; Mello, E.M.K.; Moreira, J.C.; Duarte, V. Using shade fraction image segmentation to evaluate deforestation in Landsat Thematic Mapper images of the Amazon Region. **Int. J. Remote Sensing**, 19(3):535-541, 1998.