

Utilização de classificadores automáticos no mapeamento temático de uso da terra

Maria Angélica Petrini
Christiano Luna Arraes
Jansle Vieira Rocha

Universidade Estadual de Campinas – UNICAMP/FEAGRI
Av. Candido Rondon, 501 – 13083-875 - Campinas - SP, Brasil
{maria.petrini; christiano.arraes; jansle.rocha}@feagri.unicamp.br

Abstract. Remote sensing is widely used to generate thematic maps, based mostly on image classification. The main objective of this study was to apply different methods of automatic classification, unsupervised and supervised, and to analyze which one showed better results in the creation of a land use map for the region of Luis Eduardo Magalhães municipality, located in the western part of Bahia State, which is characterized as an important expanding agricultural frontier. A Landsat 5 / TM image, path/row 220/68, was used to obtain the thematic maps. For the automatic classification it were used the classifiers Cluster, Parallelepiped, Minimum Distance and Maximum Likelihood. The latter gave best response. The main thematic classes present in the image were rough pasture, agriculture and wet bare soil. The classification accuracy was evaluated through an error matrix, with kappa coefficient of 0.62 and overall accuracy of 72,3%.

Palavras-chave: remote sensing, image processing, error matrix, sensoriamento remoto, processamento de imagens, matriz de erros, Luís Eduardo Magalhães-BA.

1. Introdução

A produção de mapas temáticos, como aqueles que representam a cobertura da terra, utilizando classificações de imagens de satélite é uma das aplicações mais frequentes do sensoriamento remoto. Isto pode ser feito tanto por análise visual quanto com auxílio do computador (FOODY, 2002). Neste último caso, há dois tipos de classificação automática: não supervisionada e supervisionada.

Segundo Richards (1986), a classificação não supervisionada é realizada na maioria das vezes através de métodos de agrupamento. Nesse tipo de classificação, o analista não precisa necessariamente ter conhecimento prévio sobre as classes existentes na imagem. Contudo, ele pode fornecer ao sistema alguns parâmetros que funcionam como critérios de decisão no momento de agrupar os *pixels* em classes distintas. Ou seja, o algoritmo identifica as classes e separa os *pixels* pertencentes a cada uma delas sem a interferência do analista. De acordo com o autor, a classificação não supervisionada é útil para determinar a composição das classes espectrais dos dados antes de uma análise mais detalhada que pode ser feita com os métodos de classificação supervisionada.

Para Richards (1986), a classificação supervisionada é o procedimento mais frequentemente utilizado para análise quantitativa de dados de sensoriamento remoto. Baseia-se no uso de algoritmos para rotular os pixels da imagem como representante de determinada cobertura da terra, ou classes temáticas. Um pressuposto importante nessa classificação é que cada classe espectral pode ser descrita por uma distribuição de probabilidade no espaço multiespectral, sendo uma distribuição multivariada com tantas variáveis quantas dimensões do espaço. Adami *et al.* (2002) destacam que, nesse método, é exigido o conhecimento prévio do analista sobre a área que se pretende classificar, bem como é fundamental o conhecimento do comportamento espectral dos alvos para poder coletar amostras significativas de cada classe. Com essas amostras são extraídos os parâmetros para treinamento do classificador, o qual vai identificar e separar cada classe temática selecionada por meio de regras estatísticas pré-estabelecidas.

A área de estudo na qual serão utilizados classificadores automáticos para mapeamento do uso da terra é a região de Luís Eduardo Magalhães, no oeste da Bahia. Esta região se

tornou a principal fronteira agrícola do estado após a expansão da agropecuária nas décadas de 1980 e 1990, com acréscimo significativo nas áreas de grãos, cultivos perenes e agricultura irrigada. Tal fato acarretou importantes transformações na região, sobretudo no que se refere ao uso e ocupação do solo (MENDONÇA, 2006). O município de Luís Eduardo Magalhães, antigo distrito de Barreiras, se destaca como um dos pólos agropecuários do país, com grandes áreas destinadas ao cultivo de soja, milho, algodão e frutas tipo exportação.

Nesse contexto, o objetivo principal desta pesquisa foi aplicar diferentes classificadores automáticos (supervisionado e não supervisionado) e verificar qual deles apresenta melhor resultado na geração de mapa temático de uso da terra da região de Luís Eduardo Magalhães, na Bahia.

2. Metodologia de Trabalho

A área de estudo selecionada abrange parte do município de Luís Eduardo Magalhães, na região oeste do Estado da Bahia, que se caracteriza como uma importante fronteira agrícola em expansão.

Foi utilizada uma imagem da órbita/ponto 220/68, datada de 04/02/2010, do satélite Landsat 5 sensor TM. Para o processamento digital das imagens foram utilizados os *softwares* ENVI 4.5, IDRISI 15.0 (*The Andes Edition*) e ArcMap 9.3.

A correção geométrica das imagens foi realizada com pontos de controles coletados na interface do Google Earth, utilizando sistema de projeção UTM, *datum* WGS-84, zona 23 Sul. O erro RMS de 0,15 ficou abaixo do erro máximo permitido de 0,5, recomendado por Eastman (1998).

Depois, foi realizado um recorte de 621 x 632 *pixels* da cena, correspondente a 353,22 km² que abrange parte do município de Luís Eduardo Magalhães (Figura 1).

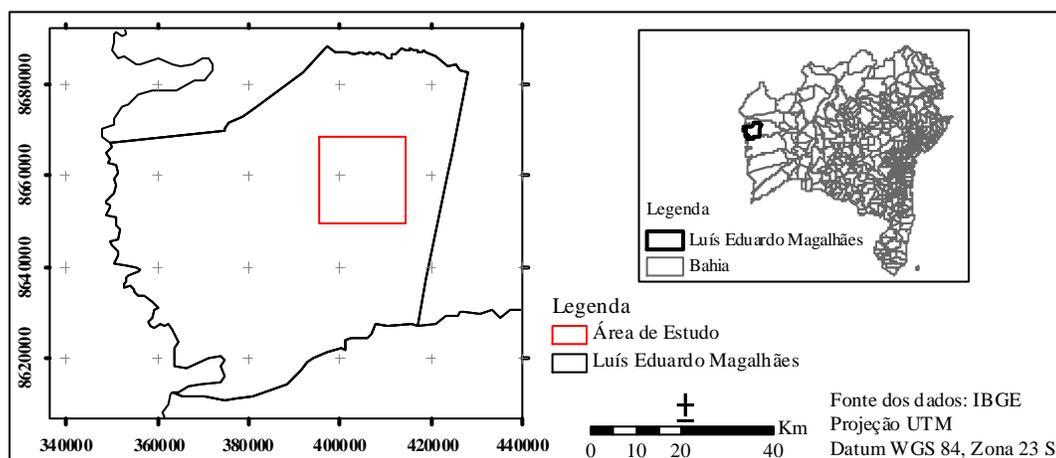


Figura 1. Localização do recorte da cena 220/68 do Landsat 5, correspondente a área de estudo no município de Luís Eduardo Magalhães-Ba.

Para a identificação visual dos temas presentes na imagem foi utilizada a composição colorida RGB 453. Em uma composição colorida, a combinação de bandas que contém as informações espectrais desejadas deve ser alocada nos canais de cores que traduzam para o sistema visual humano tais informações da melhor maneira possível, levando-se também em consideração a percepção visual de cada intérprete (CROSTA, 1993). Assim, foram identificados os alvos: água, vegetação natural, campo sujo, agricultura, solo seco, solo úmido e área urbana, a partir de padrões de cores, texturas e formas.

Para a classificação automática do uso da terra na imagem recortada, utilizaram-se algoritmos classificadores não supervisionado (*Cluster*) e supervisionados (Paralelepípedo, Distância Mínima e Máxima Verossimilhança). Vale ressaltar que a área urbana foi vetorizada manualmente e sobreposta às classificações finais, uma vez que esse tipo de alvo apresenta elevada mistura espectral e, portanto, não é separado por classificação digital de maneira satisfatória.

A fim de avaliar a classificação realizada, foi gerada a matriz de erros entre a referência (imagens Landsat) e o mapa temático final para verificar a acurácia do mapeamento. De acordo com Foody (2002), a matriz de erros é um método que vem sendo amplamente utilizado na avaliação da exatidão de dados derivados de sensores remotos. Nela, é descrito o padrão de atribuição das classes em relação aos dados de referência, ou seja, o quanto a situação retratada está classificada em conformidade com a ‘realidade’.

A partir da matriz de erros, foi calculada a Exatidão Global (Equação 1), que significa o total de acertos em relação ao total de amostras da imagem classificada. Também foi determinado o Coeficiente Kappa (Equação 2), que é uma técnica estatística para avaliação da concordância ou discordância em duas situações de interesse, cuja variação é de 0 a 1 (CONGALTON, 1991).

$$EG = \frac{A}{n} * 100 \quad (1)$$

onde: EG = exatidão global;

A = pontos amostrais com acerto;

n = número de pontos amostrais.

$$K = \frac{\left[n * \sum_{i=1}^r x_{ij} - \sum_{i=1}^r (x_i * x_j) \right]}{\left[n^2 - \sum_{i=1}^r (x_i * x_j) \right]} \quad (2)$$

onde: K = coeficiente Kappa de concordância;

r = número de linhas da matriz de erro;

x_{ij} = observação na linha i e coluna j;

x_i = total marginal da linha i;

x_j = total marginal da coluna j.

Neste trabalho, a amostragem de *pixels* foi feita de forma estratificada-aleatória, sendo que o número de amostras para cada classe temática foi calculado proporcionalmente à área ocupada pela mesma no recorte da imagem, a partir de um total definido de 195 amostras (Tabela 1). A distribuição espacial dessas amostras é mostrada na Figura 2.

Tabela 1. Número de amostras para avaliação da classificação.

Classe temática	Nº amostras
Água	1
Solo seco	2
Solo úmido	44
Campo sujo	84
Vegetação natural	4
Agricultura	54

Área urbana	6
Total	195

3. Resultados e Discussão

O mapeamento do uso da terra através de classificações automáticas foi realizado com base na composição RGB 453 da imagem Landsat, mostrada na Figura 2.

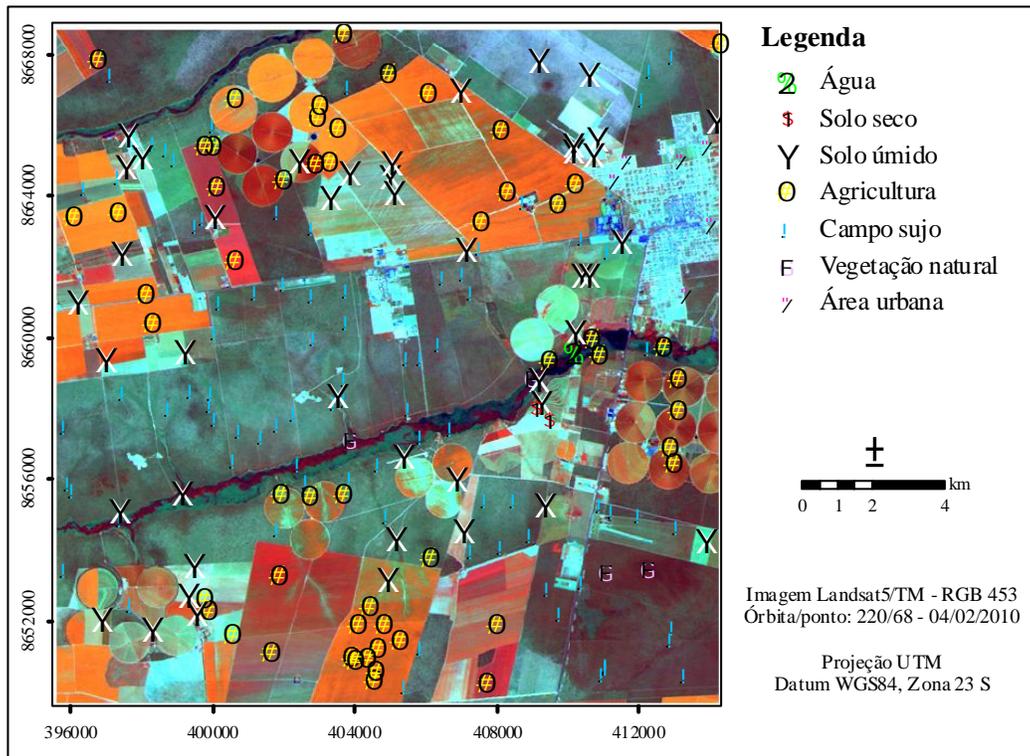


Figura 2. Composição colorida RGB 453, com as amostras de cada classe temática para avaliação da classificação.

3.1. Classificação não supervisionada

Primeiramente, foi feita uma classificação não supervisionada através da análise *Cluster* para verificar quantos agrupamentos o classificador identificaria. Houve 18 classes iniciais e passou-se, então, para o reagrupamento das classes que eram mais semelhantes entre si, resultando, no final, cinco agrupamentos por classificação não supervisionada, que são: solo seco, solo úmido, agricultura, campo sujo e vegetação natural. Pode-se perceber pela Figura 3 que esse resultado não foi satisfatório, uma vez que não foi identificada a classe água e também houve confusão principalmente entre agricultura e vegetação natural.

3.2. Classificação supervisionada

Paralelepípedo – A classificação pelo método Paralelepípedo deixou de classificar cerca de 15% da área total da imagem, principalmente os *pixels* pertencentes às classes solo úmido e solo seco. Portanto, não pode ser considerada boa classificação (Figura 4).

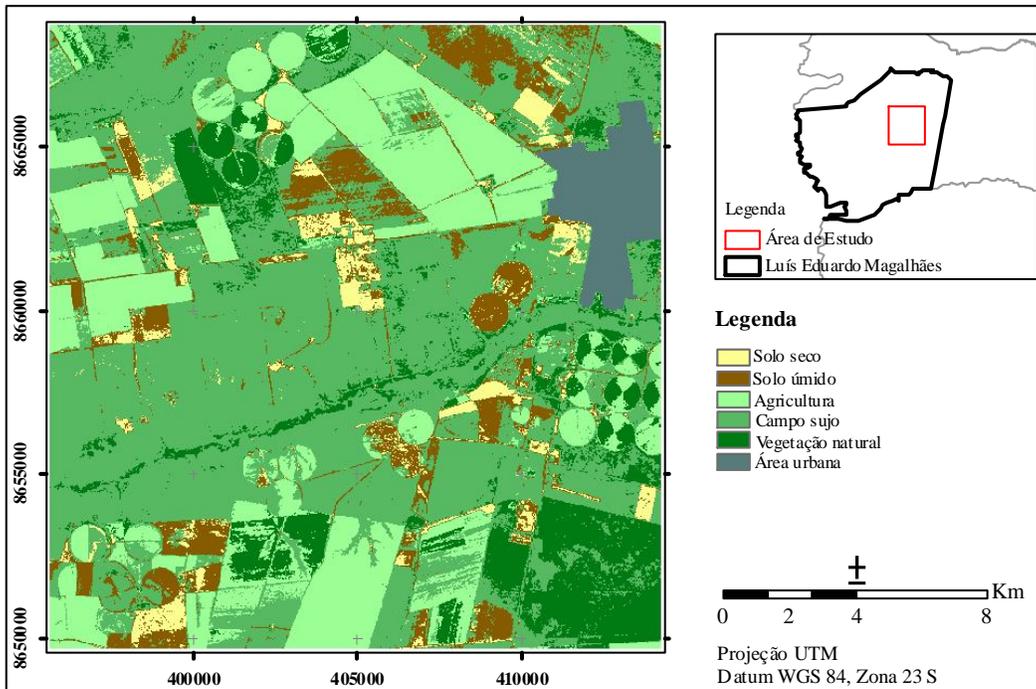


Figura 3. Classificação não supervisionada através da análise *Cluster*.

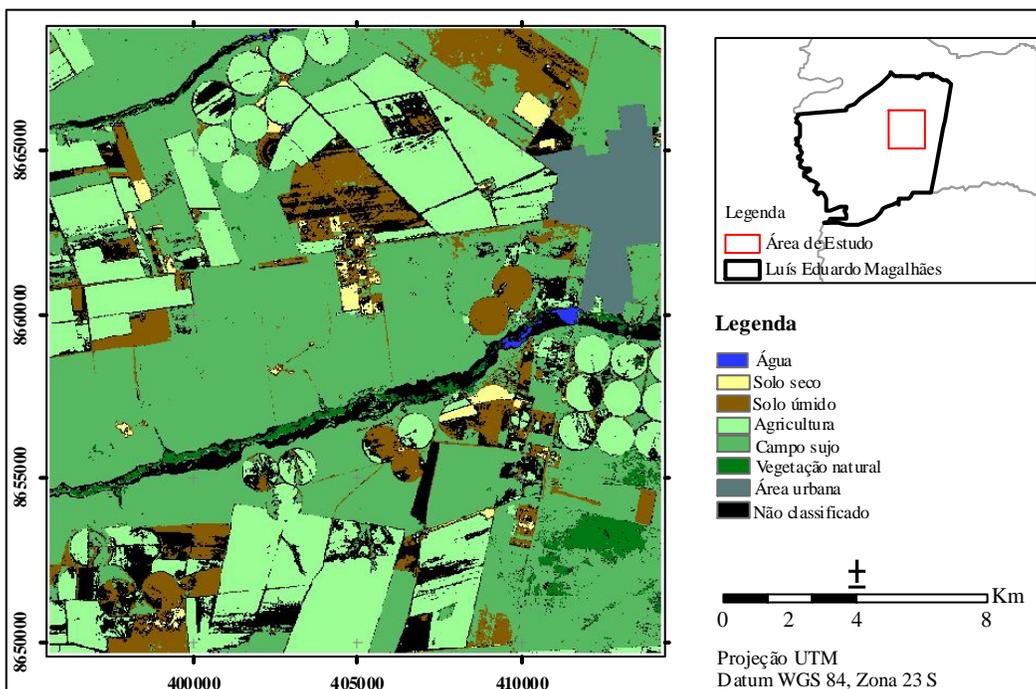


Figura 4. Classificação supervisionada pelo método Paralelepípedo.

Distância Mínima – A classificação supervisionada pelo método da Distância Mínima classificou todos os *pixels* presentes na imagem, porém apresentou confusão espectral entre agricultura e vegetação natural e entre campo sujo e solo úmido (Figura 5).

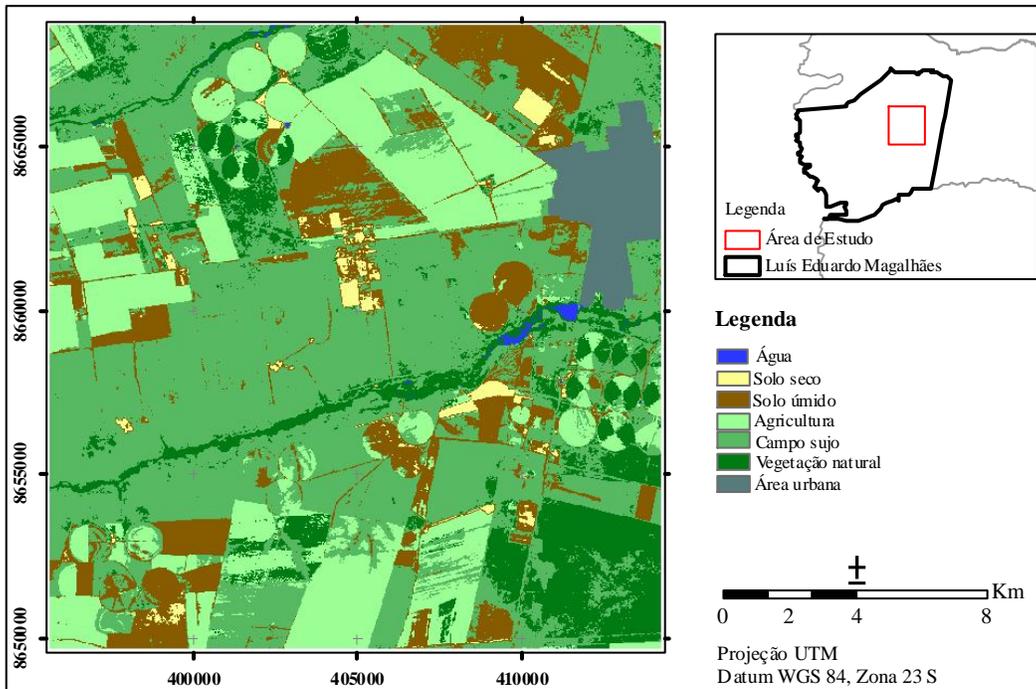


Figura 5. Classificação supervisionada pelo método Distância Mínima.

Máxima Verossimilhança – A melhor classificação supervisionada da imagem da área de estudo foi obtida pelo método da Máxima Verossimilhança, em que se puderam separar satisfatoriamente as seis classes pré-definidas de uso da terra, conforme mostra a Figura 6.

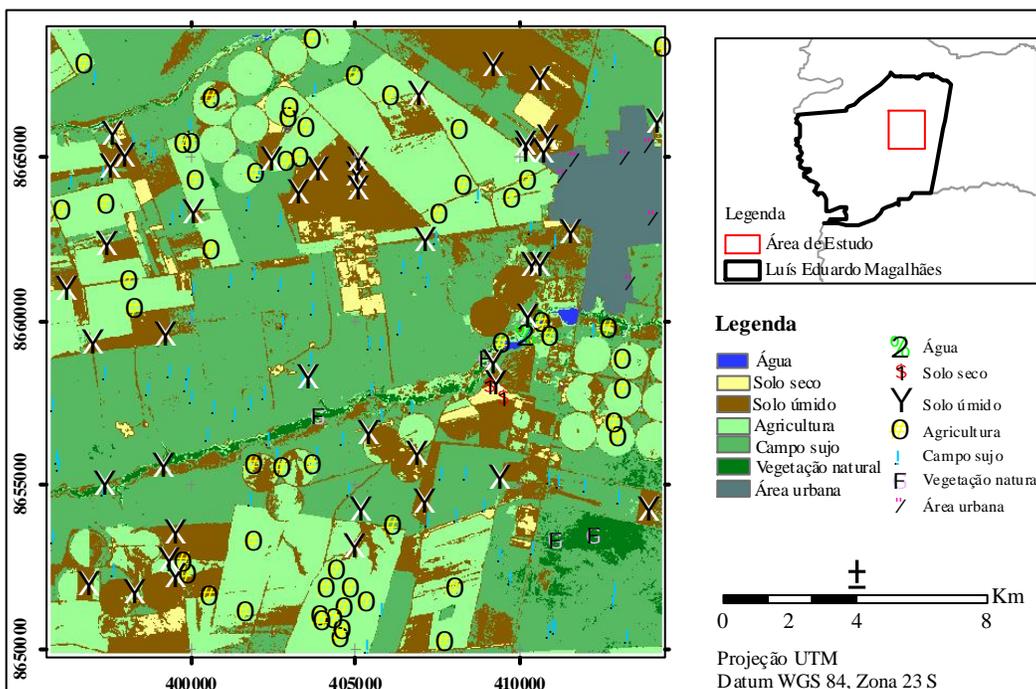


Figura 6. Classificação supervisionada pelo método Máxima Verossimilhança, com as amostras de cada classe temática para avaliação da classificação.

A Tabela 2 mostra a área absoluta e relativa ocupada por cada alvo obtida pelos diferentes classificadores. Observa-se que, pelo classificador de Máxima Verossimilhança, o qual apresentou melhor desempenho, as classes predominantes são campo sujo, agricultura e solo úmido, ocupando 42,67%, 27,04% e 22,22% da área total da imagem, respectivamente.

Tabela 2. Área absoluta e relativa ocupada por cada classe temática na imagem, por classificador utilizado.

Classe temática	Não supervisionado		Paralelepípedo		Distância Mínima		Máxima Verossimilhança	
	Área (ha)	%	Área (ha)	%	Área (ha)	%	Área (ha)	%
Água	-	-	50,94	0,14	70,11	0,20	56,88	0,16
Solo seco	1408,14	3,99	392,85	1,11	529,56	1,50	683,64	1,94
Solo úmido	3172,14	8,98	3418,47	9,68	4930,56	13,96	7849,44	22,22
Campo sujo	17922,96	50,74	15877,80	44,95	16984,62	48,08	15070,59	42,67
Vegetação natural	4327,02	12,25	686,16	1,94	4381,65	12,40	805,32	2,28
Agricultura	7186,68	20,35	8526,42	24,14	7120,44	20,16	9551,07	27,04
Área urbana	1305,54	3,70	1305,54	3,70	1305,54	3,70	1305,54	3,70
Não classificado	-	-	5064,3	14,34	-	-	-	-
Total	35322,48	100	35322,48	100	35322,48	100	35322,48	100

3.3. Avaliação da classificação

Analisando a matriz de erros montada entre a referência (imagem Landsat 5/TM, RGB 453) e o mapa temático obtido pelo classificador Máxima Verossimilhança, através dos pontos amostrais das Figuras 2 e 6, verifica-se que o erro de inclusão foi nulo somente para água, enquanto que o erro de omissão foi nulo para água e área urbana. Os erros de inclusão foram mais representativos para a classe solo seco, na qual foi incluída uma amostra de solo úmido, e para a classe solo úmido, onde foram incluídas principalmente amostras da classe agricultura. Já os maiores erros de omissão referem-se a solo seco e vegetação natural, pois o classificador confundiu solo seco e vegetação natural com campo sujo, incluindo este nessas duas classes.

Os dados da diagonal principal que estão em destaque na Tabela 3 indicam o número de amostras corretamente classificadas. Dessa forma, tem-se que, das 195 amostras, 141 delas foram corretamente classificadas, representando uma Exatidão Global de 72,3%.

O Coeficiente Kappa para o mapa temático obtido foi de 0,62, indicando qualidade muito boa da classificação por Máxima Verossimilhança, conforme proposto por Landis e Koch (1977).

Tabela 3. Matriz de erros.

		Referência – Imagem RGB 453							Σ	Erro de Inclusão
		Água	Solo seco	Solo úmido	Campo sujo	Vegetação natural	Agricultura	Área urbana		
Mapa – Max Ver	Água	1	0	0	0	0	0	0	1	0,0%
	Solo seco	0	1	1	0	0	0	0	2	50,0%
	Solo úmido	0	2	29	3	1	9	0	44	34,1%
	Campo sujo	0	4	10	57	9	4	0	84	32,1%
	Vegetação natural	0	0	1	0	3	0	0	4	25,0%
	Agricultura	0	2	4	3	0	45	0	54	16,7%
	Área urbana	0	0	0	1	0	0	5	6	16,7%
	Σ	1	9	45	64	13	58	5	195	
Erro de Omissão		0,0%	88,9%	35,6%	10,9%	76,9%	22,4%	0,0%		

4. Conclusões

Dentre os classificadores automáticos utilizados para realizar o mapeamento do uso da terra da região de Luís Eduardo Magalhães-BA tendo como base uma imagem de satélite Landsat 5 / TM, RGB 453, aquele que apresentou melhor resultado foi o Máxima Verossimilhança. Na avaliação da classificação por meio da matriz de erros, obteve-se Exatidão Global de 72,3% e Coeficiente Kappa de 62%, evidenciando bom índice de acerto.

5. Agradecimentos

A CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – pela bolsa de mestrado concedida à primeira autora; e ao CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico – pela bolsa de doutorado concedida ao segundo autor.

A Faculdade de Engenharia Agrícola da UNICAMP.

Referências Bibliográficas

Adami, M.; Pinheiro, E. S.; Moreira, M. A.; Fonseca, L. M. G. Aplicação de diferentes algoritmos para a classificação de imagens ETM+/Landsat-7 no mapeamento agrícola. In: Simpósio Latinoamericano de Percepción Remota, 10, 2002, Cochabamba. **Anais...** Cochabamba: SELPER, 2002. Disponível em: <<http://www.selper.org/selper2002/selper/articulos/T079.pdf>> Acesso em: 17 nov 2010.

Congalton, R.G. A review of assessing the accuracy classifications of remotely sensed data. **Remote Sensing of Environment**, New York, v. 37, p.35-46, 1991.

Crosta, A. P. **Processamento Digital de Imagens de Sensoriamento Remoto**. Ed. rev. Campinas, SP: IG/UNICAMP, 1993. 170 p.

Eastman, J. R. **IDRISI for Windows: Introdução e Exercícios Tutoriais**. Tradução para o português de H. Hasenack e E. Weber. Porto Alegre, UFGRS: Centro de Recursos Idrisi, 1998. 240p.

Foody, G. M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, New York, v. 80, p.185-201, 2002.

Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159-174, 1977.

Mendonça, J. O. O potencial de crescimento da produção de grãos no Oeste da Bahia. **Bahia Agríc.**, Bahia, v.7, n.2, p. 38-46, 2006.

Richads, J. A. **Remote sensing digital image analysis: an introduction**. Berlin: Springer-Verlag, 1986. 281 p.