

Avaliação do potencial de classificadores automáticos para mapeamento de uso e cobertura do solo sob manejo agroecológico

Raquel de Castro Portes ¹

Alice Azevedo Scudeller ¹

Ivo Jusksch ¹

Elpídio Inácio Fernandes Filho ¹

Irene Maria Cardoso ¹

José Marinaldo Gleriani ¹

¹ Universidade Federal de Viçosa – UFV

Avenida Peter Henry Rolfs, s/n, Campus Universitário, Viçosa – MG, Brasil

raquel_portes@yahoo.com.br; ascudeller@gmail.com; ivo@ufv.br; elpidio@ufv.br;
irene@ufv.br; gleriani@ufv.br

Abstract: The production areas based in agroecology systems are being implemented in Brazil and present on small farms intercropping different species of plants and making the diverse agricultural landscape. The classification of land cover and soil of these areas requires the use of images with high spatial resolution for detailed mapping and to identify the best method to rank areas with heterogeneous patterns of use. This study aimed to evaluate the potential of automatic classifiers for mapping land cover and soil under agro-ecological management in the São Joaquim River basin in Araçuaia, MG - Brazil. In the methodology were performed field expeditions to collect the training samples and validation using GPS. In the laboratory, supervised classifications were performed on IKONOS image by the algorithms of Maximum Likelihood and Artificial Neural Networks (Backpropagation Error) and regions (Bhattacharya). Among the classifiers evaluated in this experiment, the classification by regions had the best result, with Kappa of 0.76. The ratings by Maximum Likelihood and Artificial Neural Networks were respectively 0.48 and Kappa 0.51. This demonstrates the great potential that the supervised classification by segmentation have on classifying areas with many classes of land cover and soil and heterogeneous intra-class patterns. Thus, the findings of this study besides being useful for future planning in the watershed, will serve as universal knowledge to use classification and land cover in other areas with agro-ecological management.

Keywords: Remote Sensing, Maximum Likelihood, Artificial Neural Networks, SAFs, IKONOS, sensoriamento remoto, Máxima verossimilhança, redes neurais artificiais e Bhattacharya

1. Introdução

O planejamento dos espaços agrários é uma peça chave para elaboração de estratégias para o seu desenvolvimento. Nos dias atuais, a falta de informações espaciais em grandes escalas tem causado prejuízos nos processos de tomada de decisão, tornando-os demorados e onerosos.

Os Sistemas de Informações Geográficas (SIG's) tem contribuído para subsidiar diversos estudos na tomada de decisões e conseqüente intervenção no espaço geográfico, devido sua grande capacidade de integração de dados convencionais e georreferenciados, através do Banco

de Dados Georreferenciados (BDG). Aliado aos SIGs, os produtos do Sensoriamento Remoto, como as imagens de satélite proporcionam uma rica fonte de dados variados da superfície terrestre que podem ser extraídos, como o uso e cobertura do solo através da classificação automática.

O uso de imagens de alta resolução espacial permite a identificação e mapeamento de pequenos objetos na superfície terrestre. Isto torna o uso dessas imagens úteis no mapeamento de uso e cobertura do solo em paisagens diversificadas, como aquelas sob manejo agroecológico.

De acordo com Altieri (1999) os sistemas de produção com base agroecológica vêm sendo implantados na agricultura familiar, geralmente em pequenas propriedades e apresentam alta diversificação genética decorrentes da consorciação de espécies, tornando a paisagem agrícola mais diversa e com padrões heterogêneos de uso e cobertura do solo.

Contudo, a evolução tecnológica das imagens de satélite não foi acompanhada pelas técnicas de classificação e o aumento da variação interna das classes de uso do solo dos sistemas agroecológicos gera sobreposições das classes no espaço, que acabam dificultando a classificação.

Classificadores automáticos supervisionados de imagens de satélite têm sido utilizados amplamente na comunidade técnica e científica com o propósito de mapear uso e cobertura do solo de diferentes ambientes, como os classificadores Máxima Verossimilhança (MaxVer), Redes Neurais Artificiais (RNA) e Região. Porém, não é encontrada na literatura pesquisas utilizando estes classificadores para mapeamento de uso e cobertura do solo de sistemas agroecológicos.

De acordo com Liu (2006) o classificador MaxVer é o método mais usado no processo de identificação e delineamento das classes. O algoritmo utiliza a média e a covariância das amostras de treinamento para ser computada a probabilidade estatística de um *pixel* desconhecido pertencer a uma classe ou outra classe. Funções de densidade de probabilidade são utilizadas para classificar um *pixel* desconhecido computando a probabilidade de que aquele *pixel* venha a pertencer a uma dada categoria. Depois de avaliar a probabilidade para cada categoria, o *pixel* é assinalado à classe cuja probabilidade tenha sido a mais alta, ou a nenhuma delas no caso de os valores de probabilidade estar abaixo de um determinado valor estipulado pelo analista.

As RNAs são definidas por Ribeiro (2005) como um conjunto de unidades computacionais conectadas segundo uma topologia específica. Cada unidade representa um número de conexões de entrada, uma função de ativação e uma conexão de saída. Os produtos das entradas e dos pesos são resumidos aos neurônios, intermediários para derivar uma rede de entrada para aquela unidade. Essa rede é transformada por uma função de ativação, a fim de produzir uma saída para a unidade. O treinamento ou calibração do comportamento de uma RNA pode ser sintetizado como sendo a modificação do comportamento de saída em resposta ao estímulo de entrada. A rede aprende pela modificação dos pesos associados com as conexões entre neurônios.

Segundo Kury e Vieira (2002) a classificação por Região é um método alternativo, e que diferentemente da classificação por *pixel*, provê um meio, dentro do qual a variabilidade espacial

e a textura natural podem ser usadas no processo de classificação. Ao invés de considerar as características espectrais de um determinado *pixel* no processo de decisão, um agrupamento de *pixels* é selecionado e usado como unidade de classificação. Os classificadores por região utilizam como critério de decisão, a informação espectral de cada *pixel* da imagem e a informação espacial que envolve agrupamentos de *pixels*.

Deste modo, a pesquisa teve por objetivo avaliar o potencial de classificadores automáticos para o mapeamento do uso e cobertura do solo sob manejo agroecológico, em sistemas agroflorestais (SAF's) com cultivo de café, com intuito de gerar uma metodologia eficiente e universal para distinção de padrões heterogêneos de uso do solo para estas áreas.

2. Metodologia do trabalho

2.1. Área de estudo

A área de estudo é a bacia do Rio São Joaquim, com aproximadamente 1.089 hectares, dentro do município Araponga, na mesorregião da Zona da Mata Mineira, Brasil (Figura 1). A bacia é composta, na sua grande maioria, por pequenas propriedades rurais. Algumas propriedades há cerca de 20 anos vem adotando princípios agroecológicos no cultivo de café. Existem SAFs de diferentes idades com variadas espécies, sendo a principal uma espécie não lenhosa, a bananeira (*Musa spp.*).

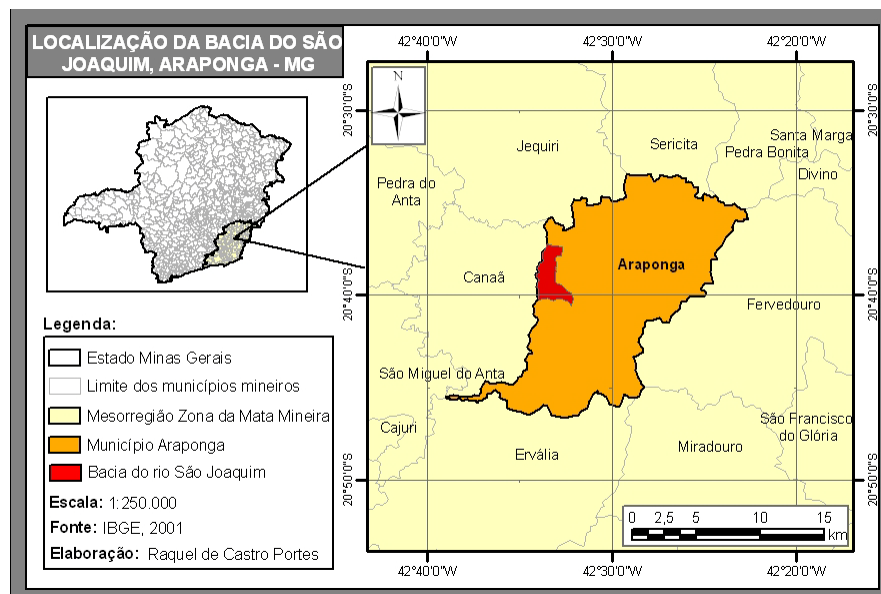




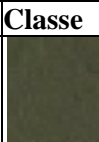




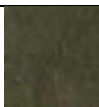


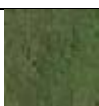
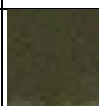

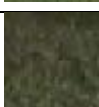


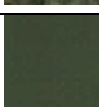
Figura 1. Mapa de localização da bacia do Rio São Joaquim, município Araponga – MG.

2.2. Procedimentos metodológicos

A primeira etapa do trabalho foi a coleta das amostras de treinamento e validação em campo, com apoio de GPS. Foram detectadas 17 classes representativas ao longo da bacia (Tabela 1.). Em laboratório com o *software* ArcGis 9.3 foram gerados dois arquivos, um para as amostras de treinamento e outro para as amostras de validação. A imagem de satélite utilizada foi a IKONOS com resolução espacial de 4 metros. As amostras não se sobrepuseram no espaço,

sendo vetorizadas em áreas distintas. O número de amostras vetorizadas em polígonos seguiu o critério de ocorrência na bacia e distribuição espacial. No geral coletou-se 6 amostras de treinamento e validação para cada padrão. Entretanto, algumas classes como pouca representatividade em área, o número de amostras foi menor. No caso da pastagem, foram coletadas mais amostras (15) devido à grande variedade que ela apresentava.

Tabela 1. Chave de interpretação de uso e cobertura do solo e número de amostragem da classe entre parênteses.

Nº	Classe	Descrição	Nº	Classe	Descrição	Nº	Classe	Descrição
1		Mata (6)	7		Solo exposto (6)	13		Taboa (6)
2		Café solteiro adulto (6)	8		Capoeira (4)	14		Pastagem (15)
3		Café solteiro jovem (6)	9		Milharal (6)	15		Café com árvore (6)
4		Café pequeno (6)	10		Cana (6)	16		Sombra (6)
5		Eucalipto (4)	11		Cana queimada (2)	17		Afloramento de rocha (3)
6		Eucalipto clonado (3)	12		Água (6)			

Foram testados três classificadores supervisionados. A classificação pelo MaxVer foi realizada *software* IDRISI Andes. No módulo MAKESIG, com as amostras de treinamento e as bandas 1 2 3 e infravermelho próximo foram criadas as assinaturas para cada classe de uso e cobertura do solo para posterior classificação no módulo MAXLIKE. Foi adotada a equi-probabilidade para cada assinatura.

A classificação por RNA ocorreu no *software* IDRISI Andes, com o módulo *Multi-Layer-Perceptron* (MLP) que utiliza o algoritmo *Error Backpropagation* para treinamento e posterior classificação. Este algoritmo segundo Vieira (2000) minimiza uma função de erro interagindo os dados de saída da rede com conjuntos de saídas desejadas, treinadas com base em uma seleção de dados amostrais. O processo continua até o erro convergir para um nível baixo e aceitável.

Para o treinamento das RNAs, foram inseridas as bandas 1 2 3 e infravermelho próximo e montadas diferentes arquiteturas, que variaram de 10; 15; 20; 25; 30 e 35 o número de neurônios, taxa de aprendizagem de 0,005; 0,007 e 0,001 e números de interações 10.000, 15.000 e 20.000. A metodologia utilizada foi proposta por Vieira (2000), que utiliza baixos valores de taxa de aprendizagem e números de neurônios e aumenta estes valores até que a taxa de aprendizagem

seja satisfatória. Assim, a arquitetura que apresentou melhor resultado foi usada para classificação.

A classificação por regiões ocorreu no *software* Spring 5.1 através do algoritmo Bhattacharya que teve por finalidade medir a separabilidade estatística entre um par de classes espectrais, através das distâncias entre as distribuições de probabilidades das classes.

Na primeira etapa, foram testadas várias combinações de similaridade e número de *pixels* para segmentação da imagem, que separou em regiões a imagem de acordo com os níveis de cinza. Assim foi definido um grau de similaridade de 10 e 10 *pixels* de área, devido à heterogeneidade espectral da imagem. O passo seguinte foi a criação de um arquivo de contexto onde foram inseridas as bandas 1 2 3 e infravermelho próximo. Este arquivo teve por objetivo armazenar as bandas utilizadas no processo de classificação, bem as amostras e a escolha do método a ser utilizado, se é *pixel a pixel* ou por região.

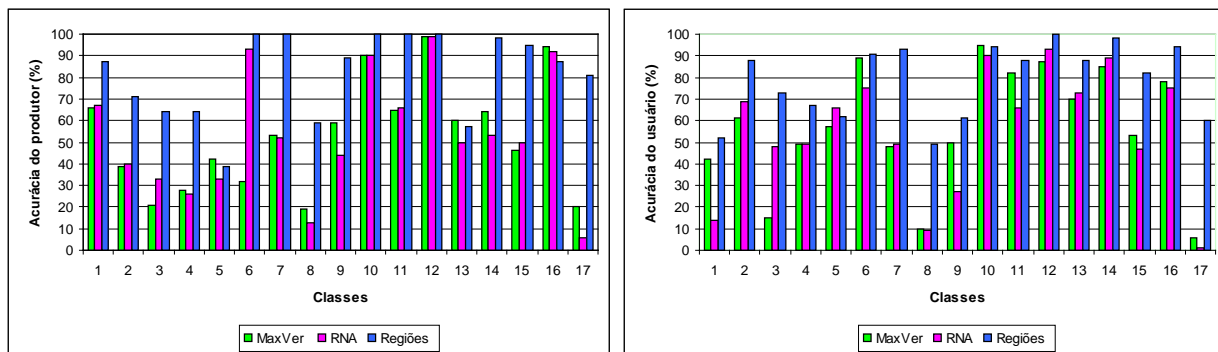
Os mapas temáticos gerados por cada classificador foram cruzados com o arquivo das amostras de validação para obtenção da matriz de erros. A partir da matriz de erros de cada mapa classificado foram derivados o índice de Exatidão Total (IET), acurácia do produtor e do usuário e índice Kappa. Os resultados encontrados para as classificações foram analisados a partir da metodologia proposta por Colgaton e Green (2008) (Equação 1 e 2) considerando que acurácia do produtor e do usuário para representar a precisão de cada classe de forma individual, ao invés de considerar apenas a exatidão total e o índice Kappa.

$$\text{Acurácia do produtor} = 1 - \text{Erro de Comissão} \tag{1}$$

$$\text{Acurácia do usuário} = 1 - \text{Erro de Omissão} \tag{2}$$

3. Resultados e discussões

A Figura 2. demonstra através de gráficos, os resultados de acurácia do produtor e usuário dos classificadores (maxVer, RNA e Regiões) para avaliação de cada classe contribuinte para geração do IET e índice Kappa.



1- Mata 2- Café solteiro adulto 3- Café solteiro jovem 4- Café pequeno 5- Eucalipto adulto 6- Eucalipto jovem 7- Solo exposto 8- Capoeira 9- Milharal 10- Cana 11- Cana queimada 12- Água 13- Taboa 14- Sombra 15- Afloramento de rocha 16- Pastagem 17- Café com árvore - SAF.

Figura 2. Acurácia do produtor e usuário para cada classificador.

3.1. Classificação pelo MaxVer

A classificação pelo MaxVer apresentou índice Kappa de 0,51, considerado “Bom” de acordo com Galparsoro e Fernández (1999).

Analisando a Figura 2, observa-se que a classe Mata, Café solteiro adulto, Café solteiro jovem, Café pequeno, Capoeira e Café com árvore apresentaram maior confusão na classificação. Portanto, o classificador não conseguiu discriminar os vários estágios de crescimento do cultivo de café presente na bacia (Café solteiro adulto, Café solteiro jovem e Café pequeno). A classe Capoeira também não foi bem classificada, sendo confundida com outros tipos de uso, como a Mata. Para a classe Café com árvore - SAF, o classificador não teve bom desempenho ao classificá-la, visto que apenas acertou 20%. Já as classes que apresentam menor heterogeneidade espectral intra-classe como Cana, Água e Pastagem, o classificador foi excelente, apresentando resultados acima de 80% para acurácia do produtor e do usuário.

3.2. Classificação pelo RNA - *Error Backpropagation*

Das arquiteturas geradas pelo Algoritmo *Error Backpropagation* o melhor resultado foi o que apresentou taxa de acurácia de 65,04% para 20 neurônios, taxa de aprendizagem de 0.007 e 20.000 interações, como demonstra a Figura 3.

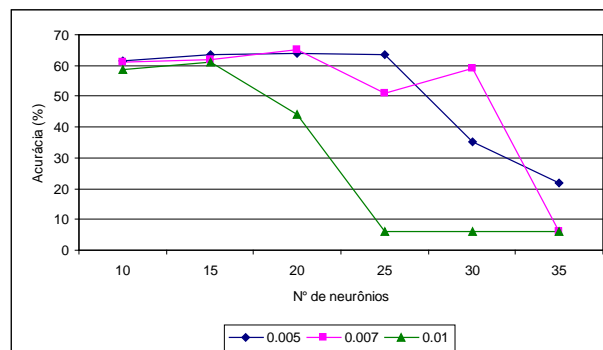


Figura 3. Treinamento de arquiteturas para 20.000 interações.

Os resultados revelaram que com o aumento do número de interações houve aumento pouco significativo da taxa de acurácia das arquiteturas geradas. O valor ótimo de números de neurônios variou de 20 a 25 neurônios na camada intermediária. Para as taxas de aprendizagem não houve uma taxa específica que apresentasse os melhores resultados. Estes dados da pesquisa demonstram que é sempre necessário realizar o treinamento das arquiteturas das RNAs de acordo com a metodologia proposta por Vieira (2000) por não existir uma regra de números de neurônios, taxas de aprendizagem e número de interações.

A arquitetura com taxa de acurácia de 65,04% apresentou índice Kappa de 0,48. Quando comparado ao MaxVer, as RNA's tiveram uma exatidão maior na classificação do eucalipto jovem (Figura 2.). Porém, a classe Café com árvore – SAF que era a mais importante para ser

discriminada obteve resultados inferiores e considerados ruins, sendo a classe que mais se confundiu com as outras.

3.3. Classificação por regiões – Bhattacharya

Dentre os classificadores analisados nesta pesquisa, o classificador por regiões Bhattacharya apresentou os melhores resultados da classificação, com índice Kappa de 0,76. Comparando com os outros classificadores utilizados nesta pesquisa (Figura 2.), o Bhattacharya teve desempenho superior ao classificar todas as 17 classes e obteve acurácia de produtor de 0,80 e usuário de 0,60 para a classe Café com árvore - SAF. Esta classe foi a que mais confusão gerou nos outros classificadores.

Como pôde ser visto pelos dados gerados na pesquisa, o algoritmo por regiões apresentou melhores resultados de acurácia de produtor e usuário para classificação de uso e cobertura do solo numa bacia hidrográfica que possui muitos padrões distintos na imagem e classes que possui alta variabilidade espectral entre os *pixels*, demonstrando que o processo de segmentação da imagem, anterior ao processo de classificação foi essencial para alcançar tais resultados.

4. Conclusões

Dentre os classificadores avaliados nesta pesquisa, a classificação por região pelo algoritmo Bhattacharya obteve o melhor resultado, com Índice Kappa de 0,76, considerado muito bom. Isto demonstra o grande potencial para classificação de áreas com muitas classes de uso e cobertura do solo e padrões heterogêneos intra-classe. Sendo assim, os resultados encontrados nesta pesquisa além de serem úteis para futuros planejamentos na bacia hidrográfica, servirão como conhecimento universal para classificação do uso e cobertura do solo em outras áreas com manejo agroecológico.

Referências bibliográficas

- Altieri, M.A. **Bases científicas para uma agricultura sustentável**. Montevideo: Nordan-Comunidad, 1999. 325 p.
- Colgaton, R. G. e Green, K. **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices**. Nova iorque: CRC Press, 2008.
- Galparsoro, L. U. e Fernandès, P. S. Medidas de concordância: el índice Kappa. **CaD Aten Primaria**, Corunã, n.6. p. 169-171, 1999.
- Kury, E. B.; Vieira, C. A. O. **Classificação automática de imagens de alta resolução** In: UFVGEO 2002 – I Workshop sobre Geoprocessamento da UFV, 2002.
- Liu, W. T. H. **Aplicações de sensoriamento remoto**. Campo Grande: UNIDERP, 2006. 822 p.
- Ribeiro, R. M. P; Soares, V. P; Vieira, C. A. O. **Avaliação de métodos de classificação de imagens IKONOS para mapeamento da cobertura terrestre**. In: Anais XII Simpósio Brasileiro de Sensoriamento Remoto, Goiânia, 2005.
- Vieira, C. A. O. **Accuracy of remotely sensing classification of agricultural crops: a comparative study**, 2000. 325 p. Tese (Ph.D) – University of Nottingham, Reino Unido, 2000.

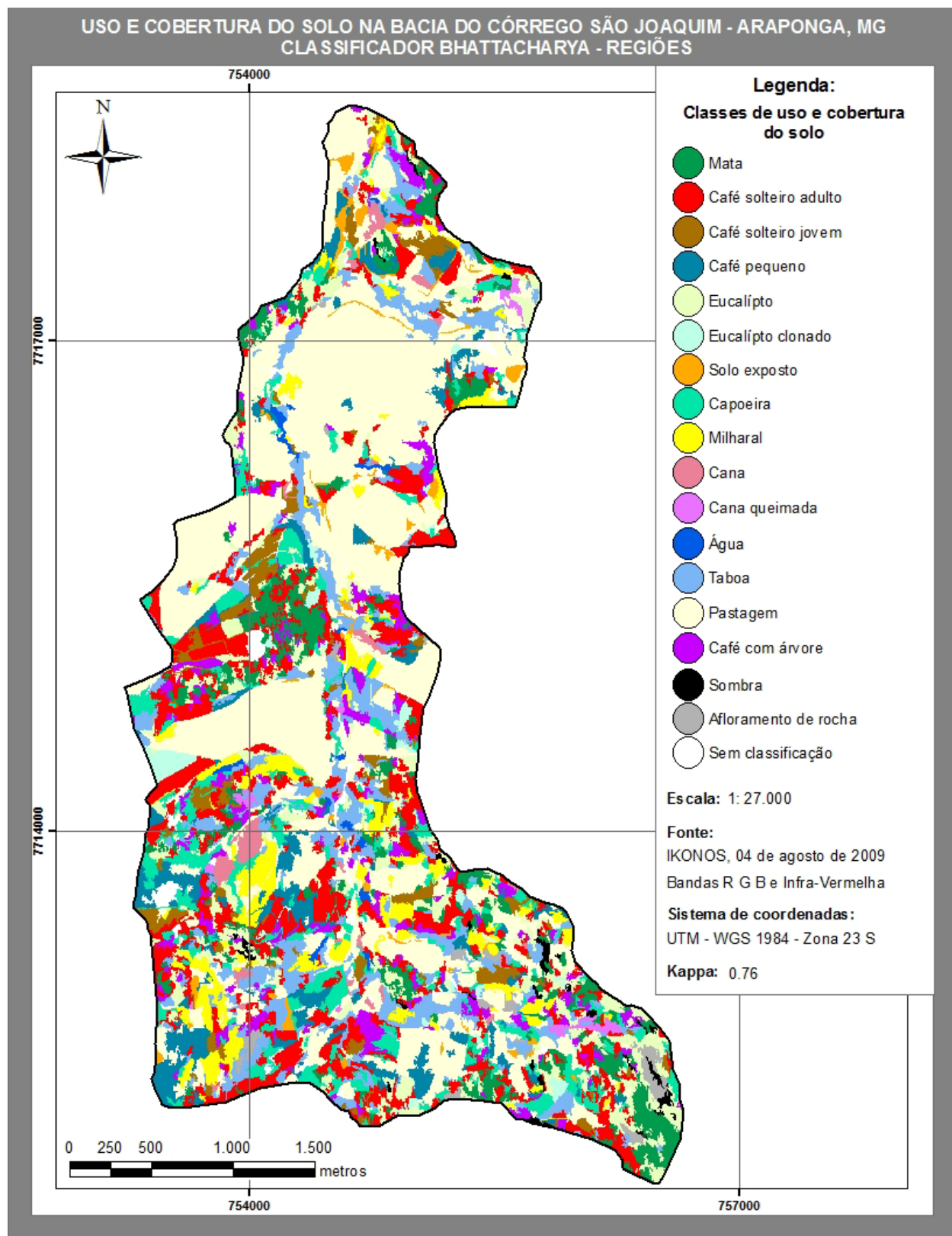


Figura 3. Mapa de uso e cobertura do solo da bacia do Rio São Joaquim – MG, Brasil por classificação por regiões – Bhattacharya.