

## Utilização de algoritmos de classificação supervisionada no mapeamento do uso e cobertura da terra no aplicativo computacional Spring 5.1.6

Telma Machado de Oliveira Peluzio<sup>1,2</sup>

Nathália Suemi Saito<sup>2</sup>

Valéria Hollunder Klippel<sup>2</sup>

Samira Murelli de Souza<sup>2</sup>

Alexandre Rosa dos Santos<sup>2</sup>

<sup>1</sup> Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina – IFSC

Av: dos Expedicionários, S/N -89460-000 – Canoinhas - SC, Brasil

tmpeluzio@hotmail.com

<sup>2</sup> Universidade Federal do Espírito Santo -UFES

Rua: Alto Universitário S/N – 295000-000 - Alegre - ES, Brasil

tmpeluzio@hotmail.com

nssaito@gmail.com

valeria.h.klippel@hotmail.com

samira\_murelli@yahoo.com.br

mundogeomatica@yahoo.com.br

**Abstract.** This study aimed to evaluate different algorithms for supervised classification Bhattacharya, maximum likelihood and minimum Euclidean distance in the mapping of land cover and land use in river basin Alegre, Alegre-ES. From the georeferenced image sensor and fused Ikonos II, 4 m resolution panchromatic band with a resolution of 1 meter, computer application in Spring 5.1.6. The classes in the study area were water, paved road, forest fragmentation, bare soil, pasture and urbanization. The resulting images of classificatory processes were evaluated by Kappa index accuracy, overall performance and test t. The results for using these indices demonstrated that the three classification algorithms were used reasonable performance, and the algorithm Battacharya who showed better discrimination of the land use and land cover present in the study area by 100% of present values for the indexes studied.

**Palavras-chave:** Remote sensing, supervised classification algorithms, sensoriamento remoto, algoritmo de classificação supervisionada.

### 1. Introdução

O conhecimento do uso da terra em uma microbacia é muito importante para o planejamento ambiental. O planejamento de manejo e conservação dos recursos naturais, ou qualquer atividade que envolva georreferenciamento pode ser executado de forma mais precisa, rápida e econômica com a utilização de sistemas de informações geográficas e sensoriamento remoto.

O sensoriamento remoto, com suas características sinópticas, multiespectrais e temporais, é uma importante ferramenta no mapeamento, identificação e monitoramento de feições terrestres. De acordo com Santos (2007), os sistemas de sensoriamento remoto, hoje disponíveis, permitem a aquisição de dados de forma global, confiável, rápida e repetitiva, sendo estes dados de grande importância para o levantamento, mapeamento e utilização das informações de uso e ocupação do solo de uma dada região.

A classificação apresenta-se de dois tipos: supervisionada e não supervisionada. A classificação supervisionada depende de amostras de treinamento que sejam representativas das classes presentes na imagem. Assim, o algoritmo classifica os pixels para cada classe. As amostras de treinamento devem ser em número, representativas e uniformes, ficando ao próprio manipulador definir quantas e quais classes serão utilizadas, enquanto que a

classificação não supervisionada dispensa a definição do número de classes e das amostras de treinamento, uma vez que o próprio Software por meio de algoritmos realizará a extração, identificação e classificação dos pixels, e o intérprete identificará as classes geradas pelo algoritmo (SILVA e PEREIRA, 2007).

Classificação, em Sensoriamento Remoto, significa a associação de pontos de uma imagem em uma classe ou grupo, ou ainda processo de reconhecimento de classes ou grupos cujos membros exibem características comuns. Como produto final tem-se um mapa temático, o qual representa a distribuição geográfica das classes temáticas.

O desempenho da classificação depende da capacidade de reconhecer as feições e da experiência do intérprete com a área de estudo. Porém, qualquer tipo de classificação está sujeita a erros. Uma vez que o método simplifica a complexidade da imagem, muitas vezes as classes predefinidas são incompletas e apenas as características espectrais são consideradas, deixando de lado as texturais.

A avaliação do desempenho dos algoritmos de classificação digital baseia-se comumente na matriz de erros e do índice Kappa. Neste processo, há a comparação entre a verdade terrestre e os resultados obtidos com a classificação digital. Na matriz de erros podem-se avaliar os pixels corretamente classificados, bem como os erros de comissão e de omissão (Matsukuma, 2002).

O objetivo principal deste estudo foi avaliar diferentes algoritmos de classificação, supervisionada, para o mapeamento da cobertura e uso da terra na bacia hidrográfica do Rio Alegre mediante a utilização do aplicativo computacional SPRING 5.1.6.

## 2. Metodologia de Trabalho

A região de estudo está localizada entre as coordenadas geográficas de 20°46'S, 41°37'O e 20°55'S, 41°28'O. Esta área foi escolhida, por ser representativa do bioma de mata atlântica e importante fornecedora de água para a cidade de Alegre-ES.

Os dados de estudo foram obtidos a partir da imagem do sensor IKONOS II (2002), composta por um mosaico georreferenciado, com resolução de 4 m e uma banda pancromática com resolução de 1 m.

Para geração e análise da imagem temática com o propósito de mapeamento do uso da terra utilizaram-se algoritmos de classificação supervisionada: Battacharya, Máxima Verossimilhança (MaxVer) e Distância Euclidiana.

Para Moreira (2005), o algoritmo Bhattacharya: requer a seleção de áreas de treinamento, podendo utilizar as regiões separadas durante o processo de segmentação ou polígonos. Em seguida, avalia, em cada região, a distância de Battacharya entre as classes (Equação 1).

$$B(p_i, p_j) = \frac{1}{2}(m_1 - m_2)^T \sum (m_i - m_j) + \frac{1}{2} \ln \frac{|\sum (m_i - m_j)|}{|\sum i|^{1/2} |\sum j|^{1/2}} \quad (1)$$

Em que,

$B$  : distância de Bhattacharya;

$p_i$  e  $p_j$  : pixels nas classes  $i$  e  $j$  ;

$m_i$  e  $m_j$  : médias das classes  $i$  e  $j$  ;

$T$  : matriz transposta;

$\ln$  : logaritmo neperiano;

$i$  e  $j$  : classes dentro do contexto.

O algoritmo Máxima Verossimilhança (MaxVer) avalia a variância e a covariância das categorias de padrões de resposta espectral quando classifica um *pixel* desconhecido. Para isto, assume-se que a distribuição da nuvem de pontos que forma a categoria dos dados de

treinamento é Gaussiana (LILLESAND e KIEFFER, 2004). Assumida essa suposição, a distribuição de um padrão de resposta espectral da categoria pode ser complementemente descrita por um valor de média e a matriz de covariância. Tornando-se possível o cálculo da probabilidade estatística de um dado valor de *pixel* pertencente a uma classe particular de cobertura da terra (Equação 2).

$$P_c = \left[ -0,5 \log_e(Det(V_c)) \right] - \left[ 0,5(X - M_c)^T (V_c)^{-1} (X - M_c) \right] \quad (2)$$

Em que,

$P_c$  : probabilidade de o vetor  $X$  ser assinalado;

$V_c$  : matriz de covariância da classe  $c$  contemplando todas as bandas (K,..., L);

$Det(V_c)$  : determinante da matriz de covariância  $V_c$ ;

$X$  : vetor de medidas dos pixels desconhecidos;

$M_c$  : vetor das médias para cada classe  $c$ ;

$T$  : matriz transposta.

O algoritmo da distância Euclidiana é calculado entre determinado pixel ( $x$ ) apresentado ao sistema e o pixel médio de cada uma das classes. Como o próprio nome sugere, o pixel apresentado é classificado na categoria para a qual a distância Euclidiana resultou num valor mínimo. Há casos, no entanto, em que este mínimo valor da distância Euclidiana ainda é muito elevado e mesmo o pixel estando muito distante de todas as categorias ele é classificado. Para evitar este inconveniente, é necessário que seja estabelecido uma distância máxima que permita ao pixel ainda ser classificado. Caso este limite seja ultrapassado, o pixel é atribuído a uma categoria não classificada (Equação 3).

$$DIST = \sqrt{\sum (VB_{ijk} - \bar{x}_{ck})^2} \quad (3)$$

Em que,

$DIST$  : Distância Euclidiana;

$VB_{ijk}$  : valor de brilho na linha  $i$ , coluna  $j$  e banda  $k$ ;

$\bar{x}_{ck}$  : média da classe  $c$  a banda  $k$ ;

$C$  : número de classes envolvidas na classificação;

$K$  : número de bandas envolvidas na classificação.

Os dados foram processados no aplicativo computacional Spring 5.1.6; para a geração das imagens temáticas para cada método de classificação. Foram identificadas após várias visitas à área de estudo, seis classes predominantes dentro da imagem, cujos temas encontram-se descritos na Figura 1.

Item	Classe	Cor	Descrição
1	Água		Corpos d'água
2	Estrada Pavimentada		Áreas com manta asfáltica
3	Fragmento Florestal		Áreas de reflorestamento e vegetação nativa (mata e capoeira)
4	Solo Exposto		Áreas de solo sem cobertura vegetal
5	Pastagem		Áreas cobertas por gramíneas utilizadas para o pastoril
6	Urbanização		Áreas urbanas

Figura 1. Quadro da relação de classes temáticas utilizadas nas classificações.

O limiar de aceitação para todos os métodos de classificação utilizados foi de 100%. O limiar dado em porcentagem define uma distância de Mahalanobis, de forma que todas as regiões pertencentes a uma dada classe por uma distância inferior a esta (INPE, 2004). Quanto maior o limiar, maior esta distância e conseqüentemente menor será o número de classes detectadas pelo algoritmo. Os dados de similaridade, área e limiar de aceitação foram escolhidos com base nas imagens geradas, pela técnica visual.

Para validação do processo de classificação supervisionada, foi utilizada a matriz de erro para determinação do desempenho global (DG) e o índice Kappa(K) e o teste de estatística T.

O exame da matriz de erros revela, para cada categoria, erros de omissão e de comissão (BRITES, 1996), sugerindo que o índice Kappa como um dos procedimentos mais utilizados para mensurar a exatidão das classificações temáticas por representar inteiramente a matriz de confusão.

Colgaton (1991) relata que o uso do coeficiente Kappa (K) é satisfatório na avaliação da precisão de uma classificação temática, pelo fato de levar em consideração toda a matriz de confusão no seu cálculo, inclusive os elementos de fora da diagonal principal, os quais representam as discordâncias na classificação, diferentemente do DG (Equação 4).

$$DG = \frac{D}{T} \quad (4)$$

Em que,

*DG*: desempenho Global;

*D* : elementos da diagonal principal;

*T* : número total de amostras.

Para Congalton e Green (1998), o cálculo de K (Equação 5), é assim descrita:

$$K = \frac{D - Q}{T - Q} \quad (5)$$

Onde Q (Equação 6).

$$Q = \frac{\sum_{i=1}^r X_{i+} X_{+i}}{T} \quad (6)$$

Em que,

*X<sub>i+</sub>* : total da linha para uma dada categoria informacional;

*X<sub>+i</sub>* : total da coluna para uma dada categoria informacional;

*T* : número total de amostras.

Todos os dados e equações foram processados no Spring 5.1.6.

### 3. Resultados e Discussão

Após o processamento da imagem do sensor IKONOS II, utilizando os algoritmos de classificação supervisionada de Battacharya, MaxVer e Distância Euclidiana, foram elaborados os mapas de uso e ocupação da terra na bacia hidrográfica do rio Alegre-ES, (Figura 2), bem com a quantificação das áreas e porcentagens do uso e ocupação da terra na bacia hidrográfica do rio Alegre-ES (Tabela1) .

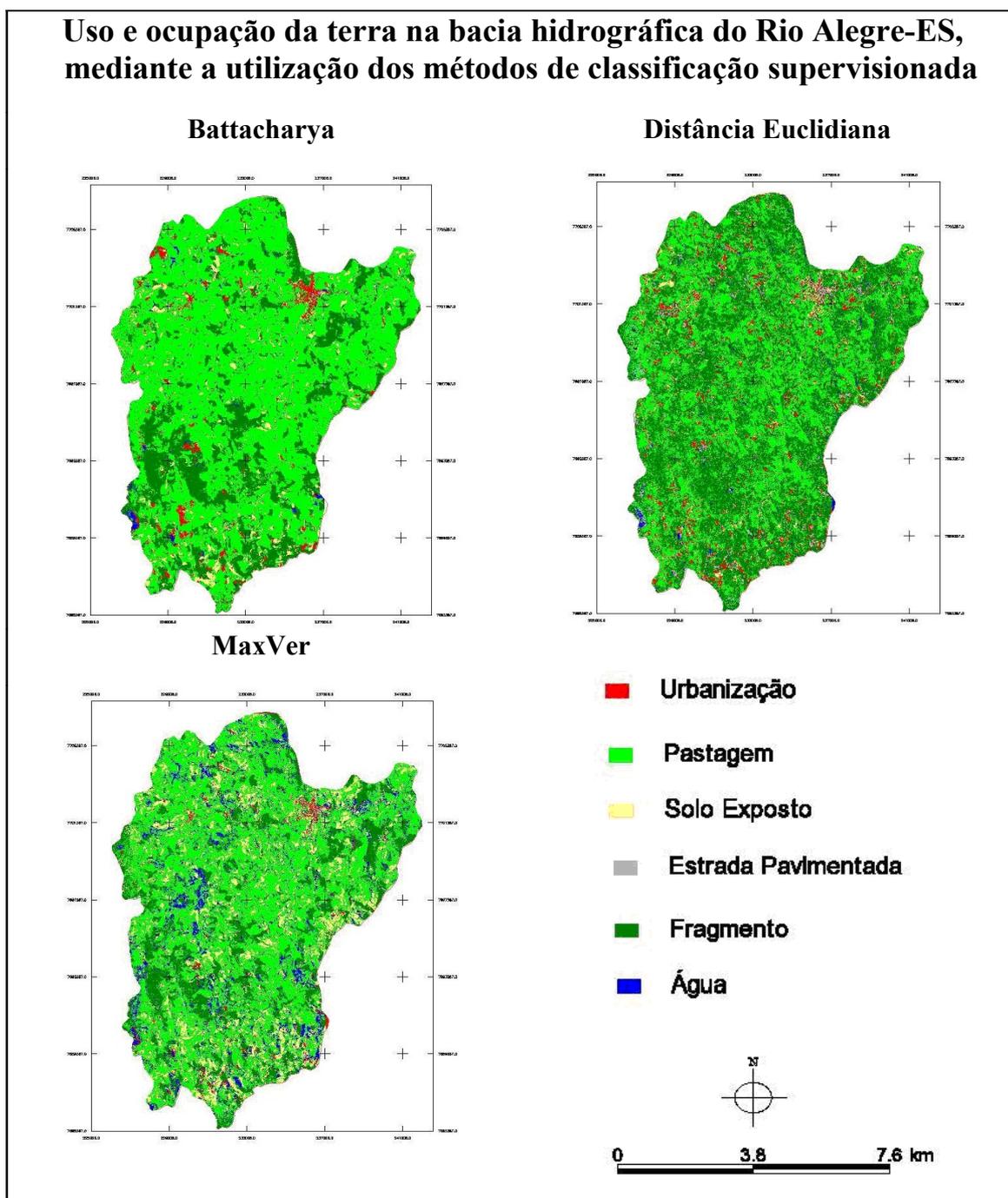


Figura 2- Uso e ocupação da terra na bacia hidrográfica do Rio Alegre-ES, mediante a utilização dos métodos de classificação supervisionada de Battacharya, MaxVer, Distância euclidiana.

O algoritmo de classificação supervisionada de Battacharya foi o que gerou uma melhor imagem temática, ou seja, com melhor discriminação das classes de uso e cobertura do solo presente na área de estudo quando comparados com o MaxVer e Distância Euclidiana.

Tabela 1. Áreas e porcentagens, do uso e ocupação da terra na bacia hidrográfica do rio Alegre-ES, mediante a utilização dos algoritmos Battacharya, MaxVer e distância Euclidiana.

CLASSES	Battacharya		MaxVer		Distância Euclidiana	
	Área (ha)	Porc. (%)	Área (ha)	Porc. (%)	Área (ha)	Porc. (%)
Água	242,8	1,18	1714,0	8,34	295,9	1,44
Estrada Pavimentada	34,2	0,17	89,8	0,44	735,7	3,58
Fragmento	4694,8	22,89	3808,0	18,53	9699,6	47,21
Solo Exposto	1070,9	5,22	3313,2	16,13	568,4	2,77
Pastagem	13832,9	67,44	11137,6	54,21	7939,8	38,64
Urbanização	635,7	3,10	484,2	2,36	1307,4	6,36
<b>TOTAL</b>	<b>20511,3</b>	<b>100</b>	<b>20546,8</b>	<b>100</b>	<b>20546,8</b>	<b>100</b>

Conforme observado na figura 2 e tabela 1 o algoritmo de Battacharya, apresentou maiores valores percentuais em relação ao algoritmo de MaxVer e distância Euclidiana na classe de pastagem (67,44%); Já o algoritmo MaxVer apresentou maiores valores percentuais nas classes de água e solo exposto (8,34% e 16,13%) respectivamente; enquanto o algoritmo de distância Euclidiana apresentou maiores valores nas classes de estrada pavimentada, fragmento e urbanização (3,58%, 47,21% e 6,36%).

Comparando o uso e ocupação da terra pela utilização dos algoritmos de classificação supervisionada verificou-se que no de Battacharya, apresenta ordem decrescente nas classes de pastagens, fragmento, solo exposto, urbanização, água e estrada pavimentada. Pelo algoritmo de máxima verossimilhança ordem decrescente nas classes de pastagens, fragmento, solo exposto, água, urbanização e estrada pavimentada. Enquanto pelo algoritmo mínima distância euclidiana a ordem apresentada é de fragmento, pastagem, urbanização, estrada pavimentada, solo exposto e água.

Apesar de apresentarem valores diferentes o algoritmo Battacharya e MaxVer apresentam a mesma ordem na classificação decrescente no uso e ocupação da terra nas classes de pastagem, fragmento e solo exposto, diferindo nas demais classes. O algoritmo da distância Euclidiana apesar de apresentar o valor total da área igual ao algoritmo de MaxVer difere totalmente na ordem das categorias de uso e ocupação da terra, dos algoritmos de MaxVer e de Battacharya.

Após realizar a classificação supervisionada da imagem, verificou-se sua exatidão utilizando os índices de Desempenho Geral, Kappa e a estatística T. A matriz de erros, bem como os valores dos índices obtidos pelo software Spring 5.1.6 são apresentados na tabela 2.

Tabela 2. Índices de exatidão da classificação supervisionada de Battacharya.

Índices	Battacharya	MaxVer	Distância Euclidiana
Desempenho Geral (DG)	100%	98%	84%
Kappa	100%	96%	74%
Estatística T	100%	97%	81%

O algoritmo da distância Euclidiana estimou valores de área para a classe fragmento florestal, maiores do que para classe pastagem, o que não ocorreu nos outros algoritmos de

classificação supervisionada. Esta discrepância de valores é explicada através dos índices de exatidão, conforme mostrado na tabela 2.

#### 4. Conclusões

A região de estudo é caracterizada por minifúndios, onde se pratica agricultura familiar e a pecuária de subsistência, possuindo desta forma uma cobertura de uso do solo bastante heterogênea, provocando com isso uma miscelânea nos valores de pixel, o que dificulta a classificação da imagem por meio dos algoritmos.

O algoritmo de classificação da Distância Euclidiana estimou valores de área para a classe fragmento florestal, maiores do que para classe pastagem, o que não ocorreu nos outros algoritmos de classificação.

O algoritmo de classificação Battacharya foi o que gerou uma melhor imagem temática, ou seja, com melhor discriminação das classes de uso e cobertura do solo presente na área de estudo. Estes resultados foram comprovados pelos índices de Kappa, Desempenho Global e teste T, sendo eles por sua vez superiores aos demais algoritmos avaliados.

#### Referências Bibliográficas

- Brites, R. S. **Verificação de exatidão em classificação de imagens digitais orbitais: efeito de diferentes estratégias de amostragem e avaliação de índices de exatidão**. 1996. 101 f. Tese (Doutorado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, 1996.
- Congalton, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v. 49, n. 12, p. 1671-1678, 1991.
- Congalton, R. G.; Green, K. **Assessing the accuracy of remotely sensed data: principles and practices**. New York: Lewis Publishers, 1998. 137 p.
- INPE – Instituto Nacional de Pesquisas Espaciais. **Spring**: sistema de processamento de informações georreferenciadas. 2004 (manual de ajuda).
- Lillesand, T. M.; Kiefer, R. W. **Remote sensing and image interpretation**. 2. ed. Chichester: John Wiley & Sons, 1994. 750 p.
- Matsukuma, C. K. **Análise comparativa de algoritmos de classificação não-supervisionada, no mapeamento do uso e cobertura do solo**. 2002. 84 p. Dissertação (Mestrado) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba. 2002.
- Moreira, M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. 2. ed. Viçosa: UFV, 2005. 307 p.
- Santos, J. C. Extração de atributos de forma e seleção de atributos usando algoritmos genéticos para classificação de regiões. 2007. 99 f. Dissertação (Computação aplicada) – INPE, São José dos Campos, 2007. Disponível em: <<http://www.inpe.br/biblioteca/>>. Acesso em: 10 out. 2010.
- Silva, M. L.M.; 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 (SBSR), 13., 2007, Florianópolis. **Anais...** São José dos Campos: INPE, 2007. Artigos, p. 6223-6228. Disponível em: < <http://marte.dpi.inpe.br/col/dpi.inpe.br/sbsr@80/2006/11.15.20.25/doc/6223-6228.pdf>>. Acesso em: 07 out. 2010.