

Revista Brasileira de Cartografia (2016), N° 68/3: 581-594
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto
ISSN: 1808-0936

COMPARAÇÃO E DESEMPENHO DE ALGORITMOS PARAMÉTRICOS NA CLASSIFICAÇÃO SUPERVISIONADA DE ÁREAS NATURALMENTE HETEROGÊNEAS E DINÂMICAS

*Comparison and Performance of Parametric Algorithms in Supervised
Classification of Naturally Heterogeneous and Dynamic Areas*

**¹Eduardo Freitas Moreira, ¹Rafaela Lorena da Silva Santos, ²Washington de
Jesus Sant'anna da Franca-Rocha, ³Danilo Boscolo
& ⁴Camila Magalhães Pigozzo**

¹Universidade Federal da Bahia – UFBA
Programa de Pós-Graduação em Ecologia e Biomonitoramento
Rua Barão de Geremoabo, s/n, Ondina – CEP 40170-115. Salvador, BA, Brasil
{dudubiologico, rafaelalorena}@gmail.com

²Universidade Estadual de Feira de Santana – UEFS
Programa de Pós-Graduação em Modelagem em Ciência da Terra e do Ambiente
Avenida Transnordestina, s/n, Novo Horizonte – CEP 44036-900. Feira de Santana, BA, Brasil
francarocha@gmail.com

³Universidade de São Paulo – USP
Faculdade de Filosofia Ciências e Letras de Ribeirão Preto
Avenida Bandeirantes, n: 3900, Vila Monte Alegre – CEP 14040-901. São Paulo, SP, Brasil
danilo.boscolo@ffclrp.usp.br

⁴Centro Universitário Jorge Amado
Centro Universitário Jorge Amado
Avenida Luís Viana, s/n, Paralela – CEP 41745-130. Salvador, BA, Brasil
camilapigozzo@gmail.com

Recebido em 3 de Dezembro, 2014/ Aceito em 6 de Junho, 2015
Received on December 3, 2014/ Accepted on June 6, 2015

RESUMO

Esse trabalho teve como objetivo geral avaliar comparativamente a eficiência dos quatro algoritmos paramétricos de classificação supervisionada por *pixel* mais utilizados no mapeamento da cobertura do solo, em uma área sob domínio do Cerrado que apresenta uma vegetação heterogênea, complexa e extremamente dinâmica, situada no Parque Municipal de Mucugê (12°59'18"S e 41°20'22"W). Para isso foi utilizado uma imagem de alta resolução espacial (KOMPSAT-2). Os algoritmos escolhidos foram de máxima verossimilhança, paralelepípedo e a distância Euclidiana mínima e de Mahalanobis. Para comparar e verificar qual foi o algoritmo paramétrico que teve o melhor desempenho foram calculadas duas medidas que foram discordância de quantidade e de alocação baseado em uma matriz de erro. Para avaliar se as classes produzidas pelo melhor classificador representavam bem a variação estrutural da vegetação da área de estudo foi realizada uma análise de variância (ANOVA) e o teste *post-hoc* de Tukey. De modo geral, o

algoritmo que apresentou melhor desempenho e que representou melhor a variação da estrutura da vegetação foi o de máxima verossimilhança. Considerando o resultado obtido conclui-se que o algoritmo de máxima verossimilhança é o mais adequado, dentre as possibilidades testadas nesse estudo para realizar mapeamentos de áreas de vegetação nativas naturalmente heterogêneas. Esse estudo pode ter grande utilidade para conservação, monitoramento e gestão de ambientes naturais.

Palavras chaves: Algoritmos de Classificação, Conservação do Cerrado, KOMPSAT-2 e Matriz de Erros.

ABSTRACT

This study aimed to comparatively assess the efficiency of the four parametric algorithms of supervised classification by *pixel* which are most used in land cover mapping, in an extremely and complex Savanna region that has a heterogeneous vegetation, located in the Municipal Park of Mucugê (12°59'18"S and 41°20'22"W). We used a spatial high-resolution image (KOMPSAT-2). The algorithms chosen were maximum likelihood, parallelepiped and minimum Euclidean and Mahalanobis distance. To compare and find which parametric algorithm had the best performance we calculated two measures that were the quantity and allocation of disagreement based on the error matrix. To assess if the classes produced by the best classifier represented well the structural variation of the vegetation of the study area we conducted an analysis of variance (ANOVA) and a post-hoc Tukey test. In general, the algorithm with the best performance and which best represented the variation of vegetation structure was the maximum likelihood. Considering the results obtained, we conclude that the maximum likelihood algorithm is the most suitable among the possibilities tested in this study to map areas of naturally heterogeneous native vegetation. This study can be very useful for conservation, monitoring and management of natural environments.

Keywords: Classification Algorithms, Savanna Conservation, KOMPSAT-2 and Error Matrix.

1. INTRODUÇÃO

As transformações dos ambientes naturais, causadas principalmente pela expansão da ocupação humana, desencadearam uma crise ambiental que vem modificando as características originais da vegetação (BERNARDI *et al.*, 2007). Nesse cenário, as tecnologias de sensoriamento remoto passaram a ocupar uma posição central no desenvolvimento científico, assim como na gestão dos recursos naturais, servindo como uma ferramenta valiosa para a conservação e monitoramento de áreas naturais (RICHARDS & JIA, 2006; JACINTHO, 2003; PRIMACK & EFRAIM, 2001). Diante desse contexto, cada vez mais está surgindo uma demanda por informações geográficas com alta resolução espacial e por dados temporais contínuos, gerando o desafio de lidar com uma grande quantidade de informações de uma forma prática, rápida e acurada (COSTA, 2004).

A interpretação e a classificação das imagens obtidas através de técnicas de sensoriamento remoto podem ser feitas de forma visual e manual ou automatizadas (NOVO, 2010). Contudo, devido à demora e custo desses processos, a forma automatizada tornou-se mais viável para o mapeamento contínuo de

grandes extensões (LANG & BLASCHKE, 2007). Além disso, com a incorporação de um número maior de sensores multiespectrais nos satélites, tornou-se humanamente impossível processar tamanha quantidade de informações sem o auxílio de técnicas analíticas apropriadas (LANG & BLASCHKE, 2007).

Nesse sentido, as técnicas de classificação automática supervisionada por *pixel* podem ser uma alternativa viável para gerar mapas de uso e cobertura do solo (RICHARD & JIA, 2006). Essa técnica consiste na amostragem dos *pixels* das imagens guiadas por seres humanos que delimitam regiões representativas de cada classe presente nas imagens (FLORENZANO, 2011). Esses dados são utilizados para o treinamento de um algoritmo que será utilizado para classificar o restante da imagem (FLORENZANO, 2011).

No entanto, existem vários tipos de algoritmos classificadores, que por sua vez são categorizados em paramétricos e não paramétricos (RICHARDS & JIA, 2006). O primeiro tipo tem caráter probabilístico e está baseado na distribuição normal dos dados. Já os classificadores não paramétricos têm natureza determinística e podem ser utilizados com dados que não se comportam com as características da curva normal (ADREOLA, 2009).

Com o aumento constante da capacidade de armazenamento e processamento dos computadores tornou-se possível a elaboração de algoritmos cada vez mais complexos para lidar com dados cada vez mais detalhados (LU & WENG, 2007). Contudo a capacidade de armazenamento e processamento continua a ser um fator limitante, devido principalmente, ao seu alto custo financeiro (NOVO, 2010). Dessa forma, considerando a sua simplicidade, os classificadores estatísticos paramétricos podem oferecer um meio termo entre o custo e acurácia dos resultados (RICHARDS & JIA, 2006).

Assim, considerando a perspectiva da estatística paramétrica, os algoritmos mais utilizados são os de máxima verossimilhança, o método do paralelepípedo e as distâncias Euclidiana mínimas e de Mahalanobis, (RICHARDS & JIA, 2006). Esses classificadores diferem quanto aos parâmetros estatísticos estimados e a sensibilidade à variabilidade dos dados, podendo apresentar desempenho diferenciado a depender das características da região a ser mapeada ou até mesmo dar origem a mapas bastante distintos, a partir do mesmo conjunto de dados (MENESES & ALMEIDA, 2012).

O algoritmo de máxima verossimilhança utiliza o conjunto de dados oriundo da coleta de *pixels* empregada para o treinamento das classes do mapa e cria uma distribuição normal de probabilidades dos valores de todas as bandas para cada classe (RICHARDS & JIA, 2006). Para classificar os *pixels* da imagem é calculada a probabilidade dos *pixels* pertencerem a cada uma das classes criadas para o mapa. Ao final, os *pixels* serão inclusos nas classes que possuem os respectivos maiores valores de probabilidade. Essa abordagem difere da distância Euclidiana mínima por ser sensível à dispersão das probabilidades da distribuição normal (RICHARDS & JIA, 2006).

O algoritmo de distância Euclidiana mínima é calculado a partir da média dos valores obtidos na coleta de *pixels* para cada assinatura de classe utilizada. Para definir a qual classe um determinado *pixel* pertence é calculada a diferença entre o valor do *pixel* e o valor médio de cada classe. O *pixel* será classificado como pertencendo à determinada classe cuja subtração apresente o menor valor. Essa medida é vantajosa em termos computacionais se comparada ao

algoritmo citado anteriormente, contudo, assume a premissa de que as distribuições de valores de todas as classes apresentam distribuição normal e possuem a mesma curtose (RICHARDS & JIA, 2006).

Já a distância de Mahalanobis, apesar de ser uma medida de distância como a Euclidiana, é ponderada pelo desvio padrão e por isso é sensível à dispersão da distribuição dos valores utilizados para o treinamento. A distância de Mahalanobis é computacionalmente mais custosa do que a distância Euclidiana, mas é menos custosa do que a máxima verossimilhança (RICHARDS & JIA, 2006).

O classificador paralelepípedo segue um princípio diferente dos citados anteriormente. A partir da distribuição dos valores obtidos através da coleta de *pixels* são definidos limiares superiores e inferiores de cada classe. Esses limiares definem os limites do intervalo de valores em uma determinada banda que cada classe pode atingir. Durante a classificação, se um *pixel* é localizado no espaço multidimensional dentro dos limites de uma classe, ele é identificado como sendo dessa classe. Quando o *pixel* está situado em uma região onde há a sobreposição de duas ou mais classes (podendo se encaixar em várias classes), ele é definido como não classificado. Contudo, se um determinado *pixel* apresentar uma assinatura ambígua, por não pertencer a nenhuma classe criada, ele é situado na primeira ou última classe. Essa é a maior desvantagem desse método em relação aos demais, que por padrão podem classificar todos os *pixels* de uma imagem e uma determinada classe pode ter *pixels* que espectralmente não deveriam pertencer àquela classe (RICHARDS & JIA, 2006).

Alguns autores (por ex.: LU & WENG, 2007; HUBERT-MOY *et al.*, 2001) já avaliaram comparativamente os referidos classificadores. Contudo, existem poucos trabalhos comparativos do desempenho dos algoritmos paramétricos na classificação de imagens multiespectrais de satélite de alta resolução espacial em paisagens heterogêneas com fitofisionomias de Savana (Cerrado) (SINTHUMULE & MUNYATI, 2014). O mapeamento de áreas de Savana é complexo devido a sua heterogeneidade natural aliada à dificuldade na distinção de suas diferentes fitofisionomias, pois são muito similares entre si, sendo caracterizadas principalmente pelas

proporções das formas de vida (VELOSO *et al.*, 1991).

Estudos como esse são relevantes devido à demanda pelo monitoramento das áreas de Savana uma vez que são consideradas *hotspots* de conservação por conta da alta biodiversidade e da extrema pressão da expansão da fronteira agrícola à qual estão sujeitas. Mapas detalhados da vegetação, que permitam o monitoramento periódico dessas áreas naturalmente heterogêneas e dinâmicas, são essenciais para o manejo de ambientes naturais em Unidades de Conservação. Nesse sentido, a avaliação das técnicas de mapeamento que permitam a distinção das diferentes fitofisionomias de savana de forma rápida e acurada, é extremamente relevante.

Assim, esse trabalho teve como objetivo geral avaliar comparativamente a eficiência de quatro algoritmos estatísticos de classificação supervisionada por *pixel* no mapeamento da cobertura do solo em uma área com fitofisionomias savânicas a partir de uma imagem multiespectral e de alta resolução espacial. Mais especificamente, avaliamos: (1) se os produtos obtidos pelos quatro classificadores diferem entre si; (2) qual classificador produz o mapa mais acurado; e por fim, (3) se o mapa produzido pelo classificador mais acurado apresenta uma boa representação da variação na estrutura vegetal da região estudada.

2. MATERIAL E MÉTODOS

Esse estudo foi realizado em uma área de 932ha, compreendendo o Parque Municipal de Mucugê e arredores (12°59'18"S e 41°20'22"W), localizada nas imediações do Parque Nacional da Chapada Diamantina, Bahia, Brasil (Fig. 1). O clima local é do tipo Bwb, segundo a classificação de Köppen (1948). Essa região apresenta uma estação chuvosa definida de novembro a abril, com média anual de precipitação em torno de 1147mm, seguida de um período seco, de maio a outubro. A temperatura média anual é de 19,6°C e a altitude média é 981m (AZEVEDO & SILVA, 2000). A classificação da vegetação adotada neste estudo foi baseada em uma adaptação da proposta de Veloso *et al.*, (1991). A vegetação é predominantemente composta por fitofisionomias savânicas (Cerrado) além de solo exposto e água. Sendo assim, tiveram as seguintes classes: (1)

Savana florestada - fitofisionomia de maior porte caracterizada por predominância de micro e nanofanerófitos esclerófilos lenhosas; (2) Savana arborizada rupestre - como a Savana arborizada descrita em Veloso *et al.*, (1991), contudo sobre substrato rochosos, caracterizada por vegetação nanocriptófitica rala e hemicriptófitica; (3) Savana gramíneo-lenhosa - dominada por hemicriptófitos graminóides, geófitos e plantas lenhosas raquíticas espaçadas; (4) Savana gramíneo-lenhosa rupestre - possui uma estrutura similar à Savana gramíneo-lenhosa, mas sobre substrato rochoso; (5) Solo exposto - áreas praticamente sem cobertura vegetal, incluindo também estradas e construções em menor quantidade e (6) Água.

2.1 Tratamento e mapeamento da imagem de satélite

A imagem utilizada para a classificação digital da área de estudo foi a do satélite coreano de alta resolução KOMPSAT-2 (*Korean Multi-Purpose Satellite*) com resolução espacial de 1m no modo pancromático (PAN) e de 4m no modo multiespectral (MS), dentro de uma largura de faixa de 15km. Esse satélite apresenta sensores com os canais PAN (500~900nm), azul (450~520nm), verde (520~600nm), vermelho (630~690nm) e infravermelho próximo (760~900nm) e com uma resolução radiométrica de 10bits/*pixels*, mas sendo fornecidas com 16bits (SIIS, 2015). As imagens foram obtidas em 06 de novembro de 2008 e adquiridas da empresa Hiparc Geotecnologia As imagens foram devidamente ortoretificadas a partir das efemérides, com os parâmetros orbitais do satélite/sensor no momento da aquisição da imagem, a base planimétrica de referência e informações altimétricas de referência. Na ausência de bases cartográficas planialtimétricas de referência de qualidade superior foram adotados os dados do satélite LANDSAT 7 ETM+ 2000, da série NASA/*GeoCover* com resolução espacial de 14,25m (MDA Federal 2004) e imagens do radar SRTM - *Shuttle Radar Topography Mission*, com 30 m de resolução espacial (USGS, 2014a e 2014b), para as informações planimétricas e altimétricas respectivamente, ambas disponíveis na base de dados <http://earthexplorer.usgs.gov/>.

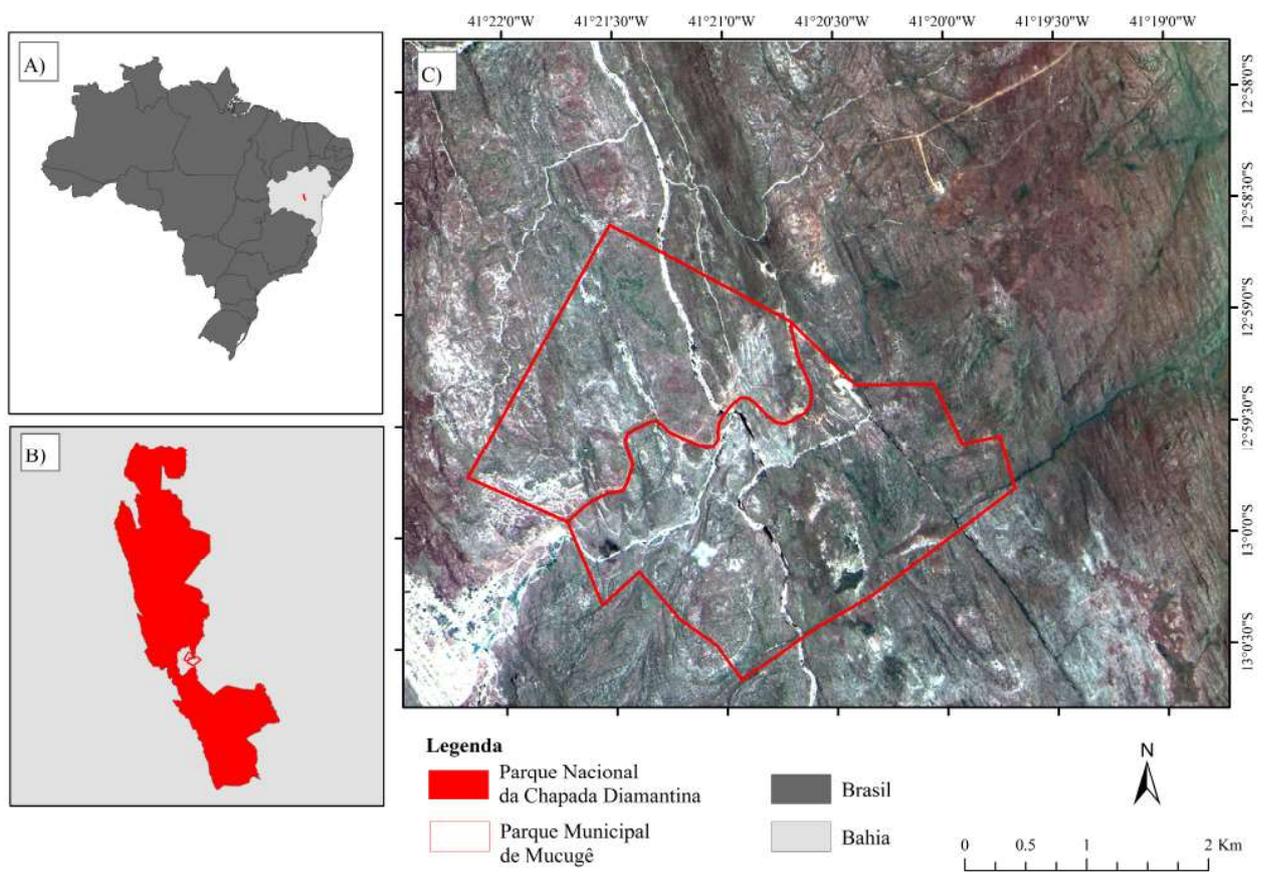


Fig. 1 - Localização da área de estudo; A) Mapa do Brasil, em destaque o estado da Bahia; B) Mapa do Parque Nacional da Chapada Diamantina; C) Imagem KOMPSAT, em destaque o Parque Municipal de Mucugê.

Para a construção dos mapas temáticos de cobertura do solo da região de estudo, utilizamos o método de classificação supervisionada por *pixels* baseado nos quatro algoritmos paramétricos mais utilizados, que são: máxima verossimilhança, distância de Mahalanobis, distância Euclidiana mínima e paralelepípedo, usando o *software* ENVI 4.7 (RICHARDS & JIA, 2006). A coleta de *pixels* na imagem com dados brutos para o treinamento dos algoritmos foi baseada na interpretação visual associada a informações coletadas em campo, que permitiram a identificação dos alvos pertencentes a sete classes utilizadas, que foram: (1) Savana florestada, (2) Savana arborizada rupestre, (3) Savana gramíneo-lenhosa, (4) Savana gramíneo-lenhosa rupestre 1, (5) Savana gramíneo-lenhosa rupestre 2, Solo exposto (6) e Água (7). A Savana gramíneo-lenhosa rupestre foi desdobrada em duas classes distintas que possuem assinaturas espectrais visualmente diferentes. Além das classes Solo exposto e Água.

2.2 Análise dos dados

Com o intuito de avaliar a qualidade da coleta de *pixels* para cada classe realizamos análises exploratórias dos histogramas, média e desvio padrão dos níveis de cinza. Além disso, também avaliamos a separabilidade das classes através do índice de divergência transformada, que permite avaliar quão estatisticamente distintos são dois conjuntos de *pixels* definidos como sendo de duas classes diferentes. Esse índice pode assumir valores entre 0 a 2. Valores acima de 1,9 foram considerados como tendo boa separabilidade (RICHARDS & JIA, 2006). Essas análises foram realizadas no *software* ENVI 4.7. Por fim, os resultados foram ilustrados por gráficos elaborados através do pacote *stats* do *software* estatístico R versão 2.15.0 (R Development Core Team, 2009).

2.2.1 Comparação dos algoritmos paramétricos

Com o intuito de analisar comparativamente os mapas resultantes dos quatro classificadores

realizamos análises exploratórias da proporção das classes de cobertura do solo nos quatro mapas. Além disso, a concordância par-a-par existente entre os mapas oriundos dos quatro classificadores foi analisada através do nível de concordância total.

O nível de concordância total entre os classificadores paramétricos foi calculado a partir das medidas de discordância de quantidade e alocação. Ambas as medidas foram realizadas com base nas matrizes de erro. As medidas de discordância de quantidade e de alocação das discordâncias foram adotadas por serem mais simples, robustas e eficientes para resumir a matriz de erro quantificada, do que o tradicional índice Kappa e suas variações (PONTIUS & MILLONES, 2011). A discordância de quantidade é a medida da diferença na proporção das classes entre os mapas avaliados. Já a discordância de alocação é uma medida da discordância no arranjo espacial dos *pixels* das diferentes classes. A concordância total por sua vez corresponde 100 menos a soma das discordâncias de quantidade e de alocação (PONTIUS & MILLONES, 2011). Os cálculos aqui descritos foram realizados no programa criado pelo professor Gil Pontius, obtido gratuitamente no site: <http://www.clarku.edu/~rpontius> (acessado 28/11/2014). Além disso, foi gerado um gráfico de densidade de pontos a partir das matrizes de erros representando as comparações par-a-par entre os algoritmos paramétricos. Também geramos um gráfico de linhas para representar a variação da proporção das classes referente a cada algoritmo paramétrico usado. Para isso foi utilizado o pacote *stats* do *software* estatístico R versão 2.15.0. Como referência para essa análise, sorteamos coordenadas de forma aleatória, por toda a região de estudo, com o critério de distância mínima de 100m. Os *pixels* que continham as referidas coordenadas foram selecionados como unidades amostrais. Esse procedimento permitiu a seleção 29210 unidades amostrais (*pixels*), homoganeamente distribuídas na região. Cada unidade amostral classificada de acordo com os mapas oriundos dos quatro algoritmos. Esses procedimentos foram realizados com o *software* Hawth's *analysis tools* for ArcGIS (BEYER, 2004), em conjunto com o *software* Arcview 9.3 for Windows.

2.2.2 Avaliação da acurácia dos algoritmos paramétricos

O melhor classificador paramétrico foi selecionado com base na acurácia total, também calculada a partir das medidas de discordância de quantidade e alocação. Ambas as medidas foram realizadas com base nas matrizes de erro, de acordo com o método de Pontius & Millones (2011) descrito na seção anterior. A única diferença é que ao invés dos mapas de cada classificador serem comparados entre si foram comparados com dados de referência correspondentes à realidade local.

Para a análise da acurácia realizamos uma amostragem aleatória estratificada, com 50 unidades amostrais para cada classe de cobertura, totalizando 300 unidades amostrais homoganeamente distribuídas na região, que conferem 298 graus de liberdade à referida análise. Cada unidade amostral foi caracterizada, com base em dados da realidade local obtidos através da observação de 193 coordenadas em campo com o suporte adicional da identificação visual em uma imagem SPOT-5 de 2007, com resolução espacial de 5m e outra imagem KOMPSAT-2 de junho de 2008, com resolução espacial de 1m, ambas com composição R - vermelho, G - infravermelho próximo e B - azul. Os dados de cada coordenada foram comparados com a classificação dos *pixels* que as continham, de acordo com o resultado dos quatro classificadores. Esses procedimentos foram realizados com o *software* Hawth's *analysis tools* for ArcGIS (BEYER, 2004), em conjunto com o *software* Arcview 9.3 for Windows. Os cálculos das discordâncias de quantidade e alocação, bem como da acurácia total foram realizados no programa criado pelo professor Gil Pontius, obtido gratuitamente no site: <http://www.clarku.edu/~rpontius> (acessado 28/11/2014).

2.2.3 Análise da representatividade da variação fitofisionômica pelo mapa mais acurado

Por fim, com o intuito de avaliar se as classes produzidas pelo melhor classificador representam bem a variação estrutural da vegetação realizamos uma análise de variância (ANOVA) usando a estatística F, tendo como fator as classes de cobertura do solo e como

variável resposta o índice de vegetação calculado a partir das imagens de satélite. O índice de vegetação utilizado foi o *two-band enhanced vegetation index* (EVI2) proposto por Jiang *et al.*, (2008) e calculado a partir da imagem do satélite KOMPSAT-2 de novembro de 2008, descrita na seção “2.2 Tratamento e mapeamento da imagem de satélite”. Esse índice está associado com a cobertura foliar, e com a estrutura da vegetação. Sua vantagem sobre os demais índices de vegetação descritos na literatura é que necessita, apenas, das bandas referentes aos comprimentos de onda do vermelho e infravermelho próximo, além do mesmo ponderar o efeito de fatores de confusão, como a refletância do solo e dos aerossóis atmosféricos (JIANG *et al.*, 2008). Para essa análise selecionamos 1000 unidades amostrais (cuja finalidade era de obter um grande número de graus de liberdade para a realização do teste estatístico), para cada uma das seguintes classes: Savana arborizada rupestre, Savana florestada, Savana gramíneo-lenhosa, Savana gramíneo-lenhosa rupestre e uma classe denominada de Solo exposto, totalizando 5000 unidades amostrais, que representam 4998 graus de liberdade para o teste estatístico global. Como o referido conjunto de dados foi testado apenas uma vez, foi adotado o valor de alpha igual a 0,05 como critério de significância, com intuito de estabelecer o equilíbrio entre o erro do tipo I e do tipo II (GOTELLI & ELISON, 2011). Nesse teste a hipótese nula testada é de que não existe diferença entre as médias da distribuição normal dos valores de EVI2 dos pontos correspondentes às sete classes de cobertura do solo. No caso de a hipótese nula ser rejeitada no teste global ($p < 0,05$), foi realizado o teste de *post-hoc* de Tukey para avaliar diferença entre pares de classes do mapa em relação aos valores de EVI2. Todos os testes foram realizados *software* estatístico R versão 2.15.0. A seleção das unidades amostrais foi realizada com a ferramenta *Hawth's analysis tools for ArcGIS* (BEYER, 2004) e o cálculo do EVI2 foi realizado no *software* ArcGIS 9.3.

RESULTADOS E DISCUSSÃO

Ao todo foram coletados 28612 *pixels*, agrupados nas sete classes, Savana florestada (1784 *pixels*), Solo exposto (1675 *pixels*), Savana arborizada rupestre (4225 *pixels*), Savana

gramíneo-lenhosa rupestre 1 (2094 *pixels*), Savana gramíneo-lenhosa rupestre 2 (2229 *pixels*), Savana gramíneo-lenhosa (15210 *pixels*) e Água (1395 *pixels*). De um modo geral, através da análise gráfica dos histogramas de cada classe ao final da coleta de *pixels* foi possível perceber que todas apresentaram distribuições normais dos níveis de cinza nas quatro bandas utilizadas para a classificação, com exceção da classe Água e Solo exposto (Fig. 2).

Inicialmente, a coleta de *pixels* realizada para a classe Savana gramíneo lenhosa rupestre apresentou um padrão bimodal na análise dos histogramas de todas as bandas. Isso ocorreu devido à agregação da cobertura vegetal nas fendas das rochas intercaladas com manchas de rochas nuas, gerando um padrão em mosaico com dois picos de valores mais frequentes no histograma. Para solucionar esse problema a referida classe foi dividida em duas, Savana gramíneo-lenhosa rupestre 1, com cobertura vegetal mais homogênea e Savana gramíneo-lenhosa rupestre 2, com cobertura vegetal mais variável.

Contudo, no processo de pós-classificação, essas classes foram fundidas em uma só denominada Savana gramíneo-lenhosa rupestre. O cuidado na seleção dos *pixels* representativos de cada classe é determinante da qualidade do resultado final da classificação (LU & WENG, 2007). Em se tratando da aplicação de classificadores estatísticos paramétricos é fundamental que os histogramas de todas as classes apresentem a forma simétrica em torno da média, característica da distribuição Gaussiana (QUINN & KEOUGH, 2002; COX, 2006).

A banda que apresentou maiores variações entre as classes de cobertura foi a do infravermelho próximo (Fig. 3). As demais bandas apresentaram comparativamente pequenas variações nas médias das classes, principalmente se compararmos as savanas. A classe de cobertura que apresentou valor de média mais distinto das demais foi o Solo exposto, seguida da Savana florestada. A classe Solo exposto também apresentou os maiores valores de desvio padrão, seguida das classes Água e Savana florestada.

Esses resultados eram esperados considerando os tipos de cobertura da região, uma vez que solo exposto, água e vegetação

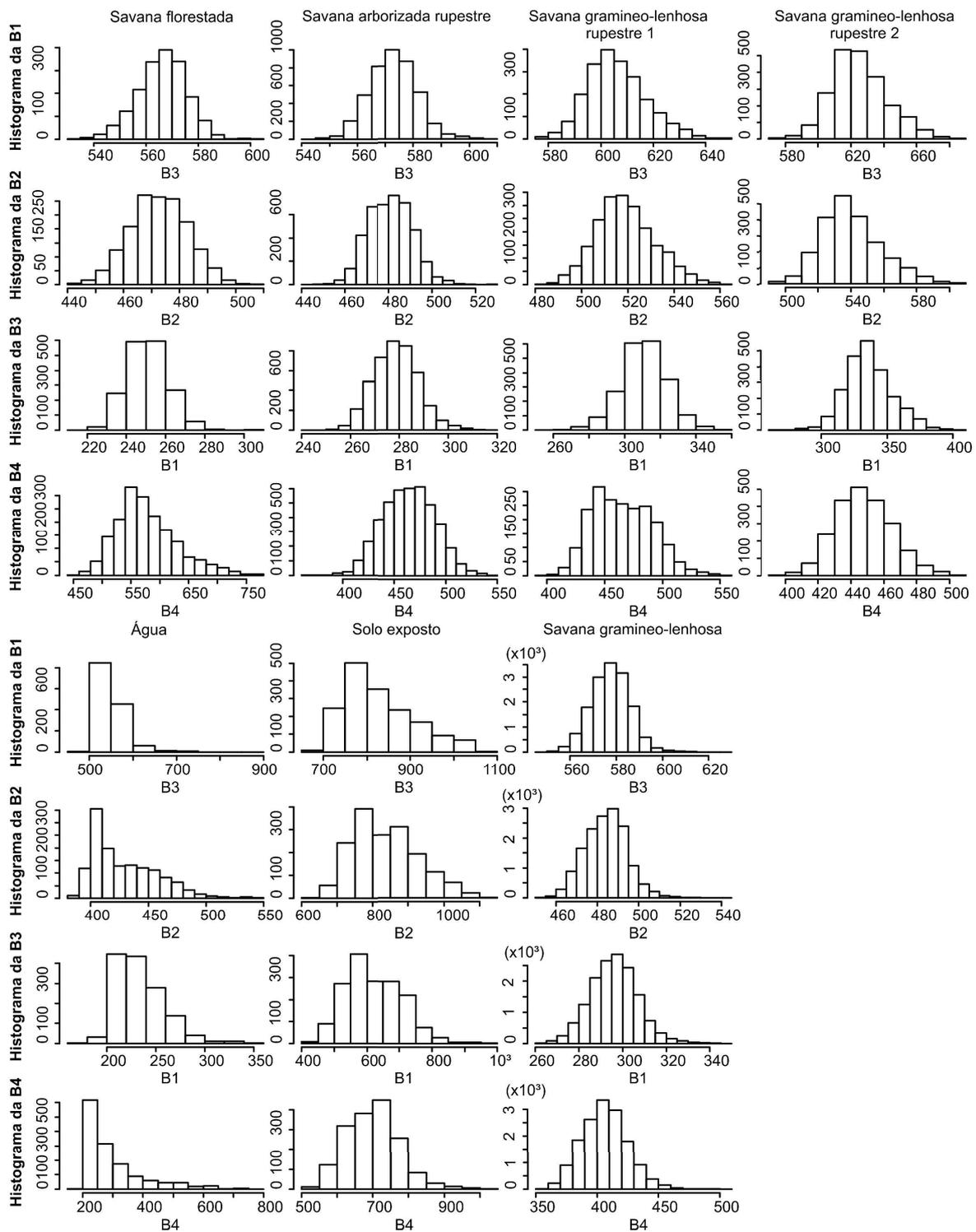


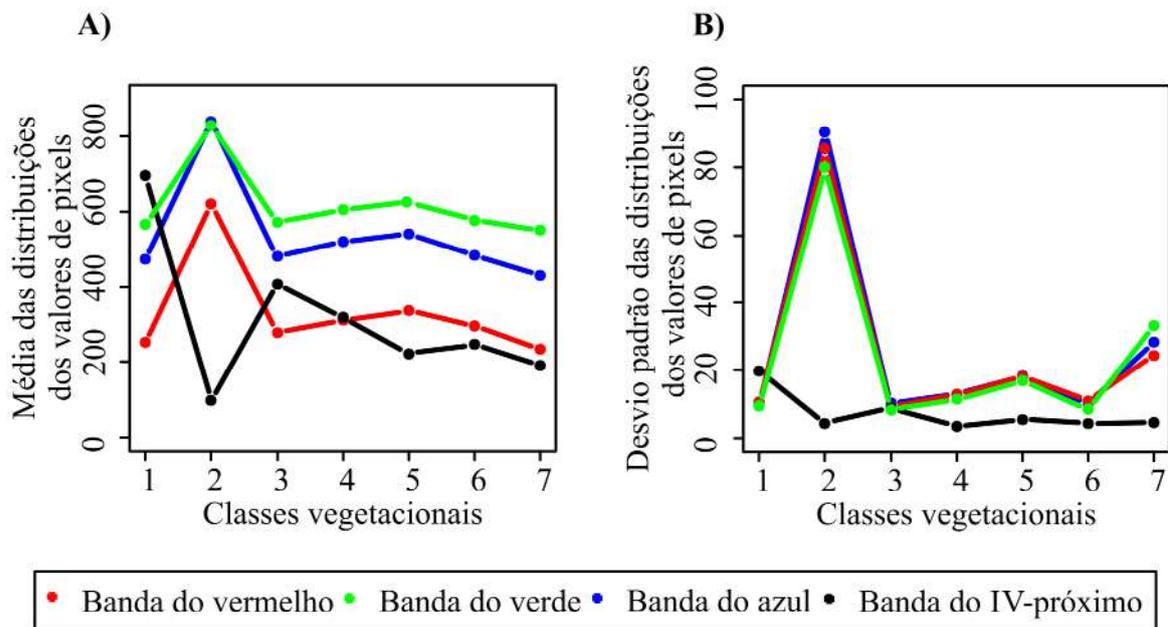
Fig. 2 - Histogramas referentes a cada classe de vegetação nas quatro bandas. No eixo y estão representadas as frequências e no eixo x o nível de cinza; B1 – Banda do vermelho; B2 – Banda do verde; B3 – Banda do azul; B4 – Banda infravermelho próximo.

apresentam diferenças de refletância muito maiores na faixa do infravermelho próximo se comparadas com as diferenças de refletância na faixa do visível (RICHARDS & JIA, 2006). Três das sete classes utilizadas, Savana Florestada, Solo exposto e Água apresentaram boa separabilidade em relação a todas as

demais, com valores do índice de divergência transformada ($d_{ij} > 1,9$). As demais, de um modo geral, apresentaram valores mais baixos em pelo menos uma combinação. A Savana arborizada em relação à Savana gramíneo-lenhosa rupestre 1 apresentou um valor de separabilidade moderado ($d_{ij} = 1,8$). Essa última por sua vez apresentou

uma separabilidade moderada em relação à Savana gramíneo-lenhosa ($d_{ij}=1,8$) e baixíssima em relação à Savana gramíneo-lenhosa rupestre 2 ($d_{ij}=1,2$). Com exceção da classe Água, as diferenças de separabilidade são um reflexo conjunto do porte da vegetação, do tipo de solo e da cobertura foliar presentes em cada classe (PONZONI & SHIMABUKURO, 2007). Como a Savana florestada é dominada por fanerófitas de grande porte (6 a 15 metros) destaca-se facilmente das demais classes por apresentar solo mais profundo e maior cobertura foliar, da

mesma forma que a classe Solo exposto, que se diferencia radicalmente das demais pela ausência quase que absoluta de vegetação (VELOSO *et al.*, 1991). As demais classes são caracterizadas por uma variação gradativa na cobertura do solo e estratificação da vegetação com diferentes proporções de fanerófitas de pequeno porte (<5m), caméfitas, hemicriptófitas e geófitas, e por conta disso tendem a apresentar médias dos níveis de cinza mais próximas entre si (VELOSO *et al.*, 1991; PONZONI & SHIMABUKURO, 2007).



3.1 Comparação dos classificadores paramétricos

Nas quatro classificações feitas (Fig. 4), as classes com maior área na região estudada foram Savana gramíneo-lenhosa rupestre, Savana arborizada rupestre e Savana gramíneo-lenhosa. Contudo, os quatro algoritmos apresentaram uma variação considerável na proporção das classes de cobertura (Fig. 5). Além disso, a concordância variou entre os mapas obtidos pelos diferentes algoritmos, sendo que a maioria das comparações ficou em torno de 70% (Fig. 6). Essas diferenças nos resultados da classificação podem ser atribuídas à diferença de

sensibilidade dos classificadores e à curtose das distribuições de frequência de níveis de cinza. Por exemplo, enquanto a distância Euclidiana desconsidera completamente a dispersão dos valores, a distância de Mahalanobis pondera os valores de distância pelo desvio padrão e o método do paralelepípedo tende a superestimar essa dispersão (RICHARDS & JIA, 2006). Já a máxima verossimilhança representa um retrato mais fiel das distribuições de frequências dos níveis de cinza, pois o cálculo das probabilidades é baseado na própria distribuição, sem privilegiar medidas de tendência central ou de dispersão (QUINN & KEOUGH, 2002; COX, 2006).

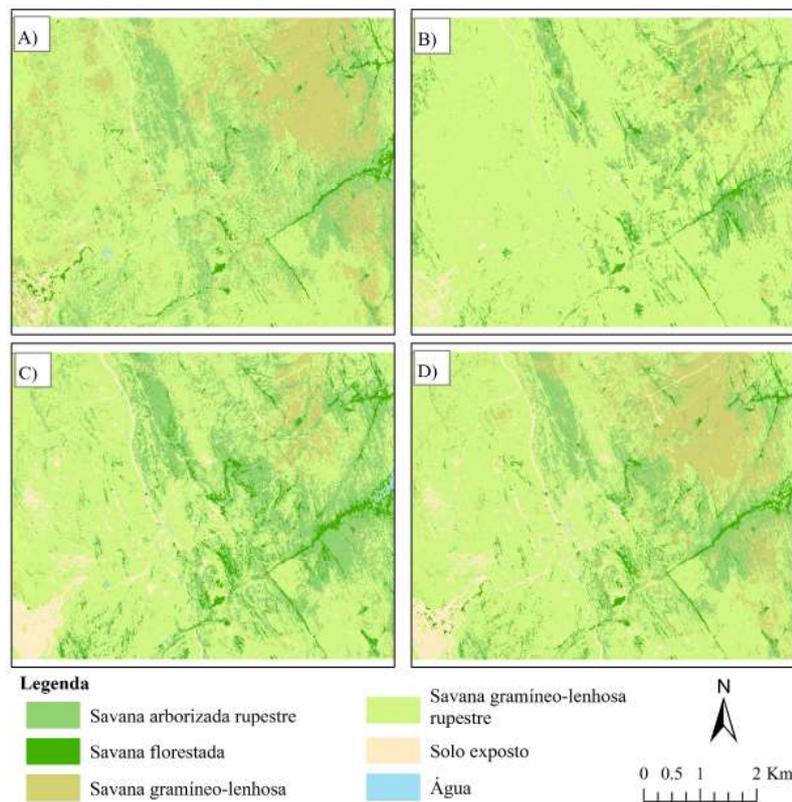


Fig. 4 - Mapas de classes gerados pelos quatro algoritmos paramétricos. A – Distância Mahalanobis; B – Distância Euclidiana mínima; C – Paralelepípedo; D – Máxima verossimilhança.

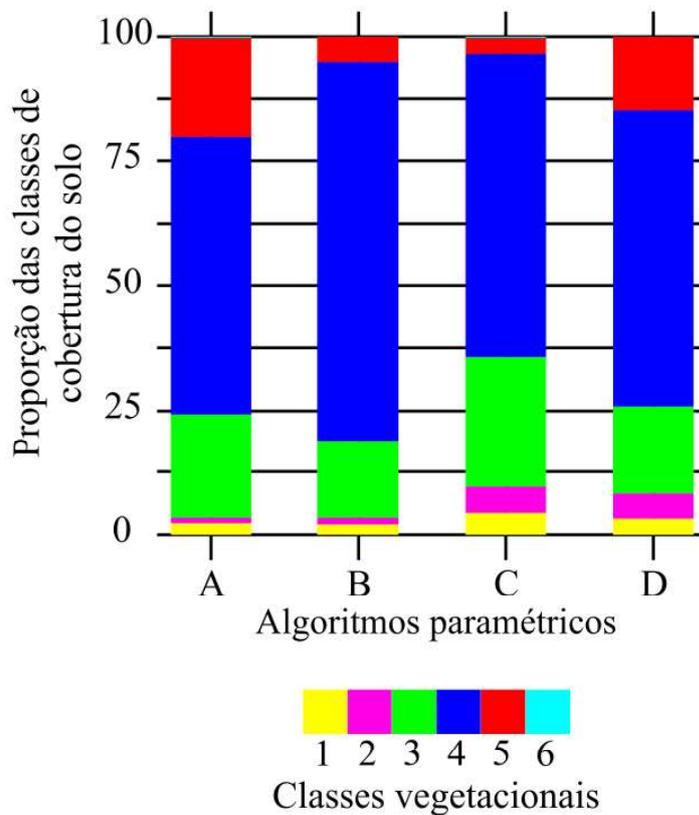


Fig. 5 - Variação da proporção das classes de acordo com o algoritmo paramétrico usado no presente estudo. A – Distância Mahalanobis; B – Distância Euclidiana mínima; C – Paralelepípedo; D – Máxima verossimilhança. Classes vegetacionais: 1 – Savana florestada; 2 – Solo exposto; 3 – Savana arborizada rupestre; 4 – Savana gramíneo-lenhosa rupestre; 5 – Savana gramíneo-lenhosa e 6 – Água.

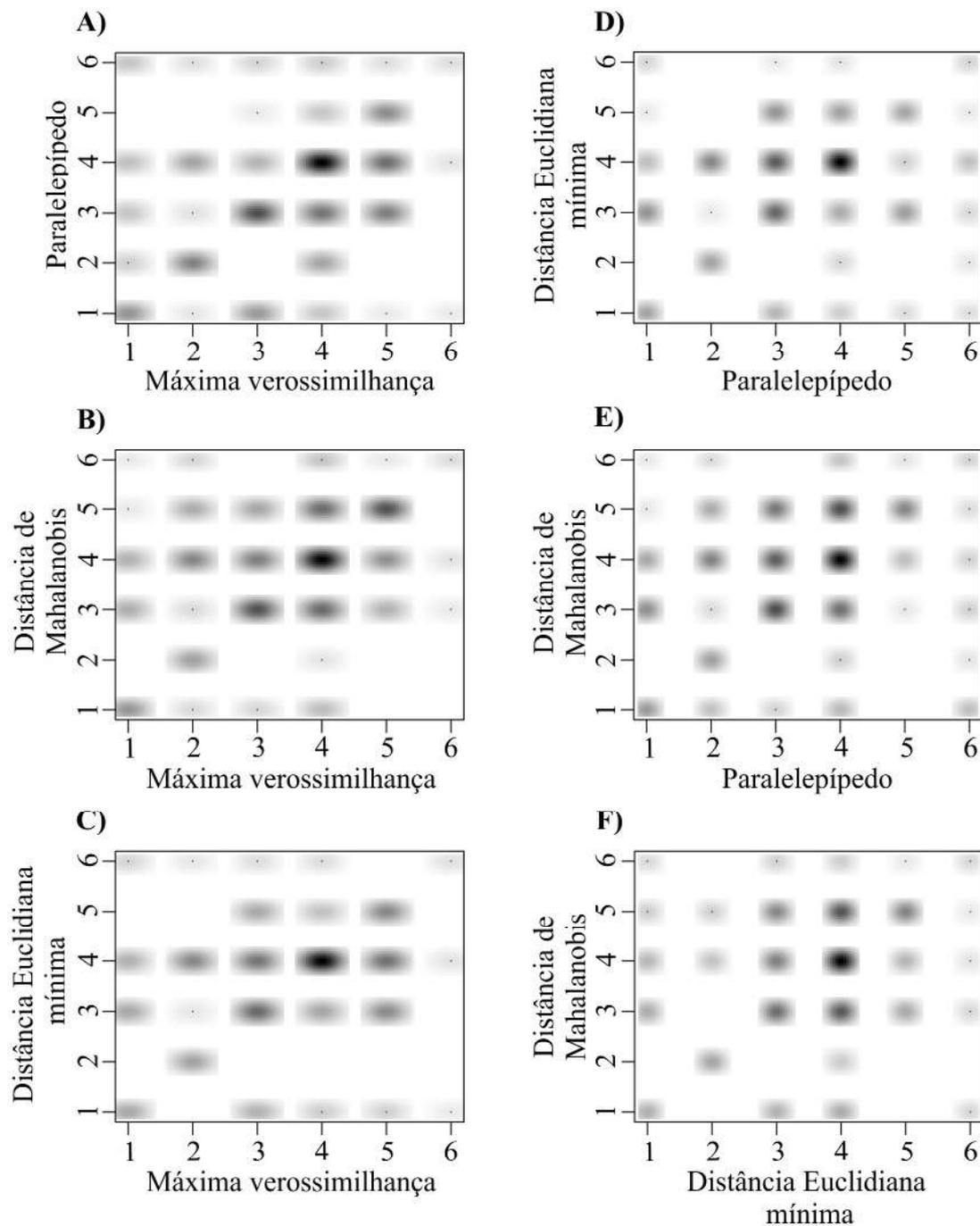


Fig. 6 - Concordância entre os mapas gerados pelos quatro tipos de algoritmos utilizados. As matrizes de erro representam as comparações par-a-par entre os algoritmos, os tons de cinza representam a densidade de pontos, sendo que quanto mais concentrados na diagonal ascendente maior é a concordância entre os mapas; A – 80%; B – 76%; C – 73%; D – 73%; E – 68%; F – 63%.

3.2 Avaliação comparativa da acurácia dos algoritmos paramétricos e da representatividade da variação estrutural da vegetação pela melhor classificação

As diferenças dos resultados dos algoritmos testados foram expressas na acurácia dos mesmos. O melhor classificador foi o de máxima verossimilhança, cuja concordância com a realidade foi classificada com base

na acurácia total (92%). Os demais classificadores apresentaram desempenhos inferiores. O classificador paralelepípedo foi o segundo colocado e apresentou uma concordância substancial com a realidade, apesar do desempenho bastante inferior em relação à acurácia total (75%). As distâncias de Mahalanobis e Euclidiana mínima apresentaram valores de acurácia mais próximos entre si (67% e 71%, respectivamente) (Fig. 7).

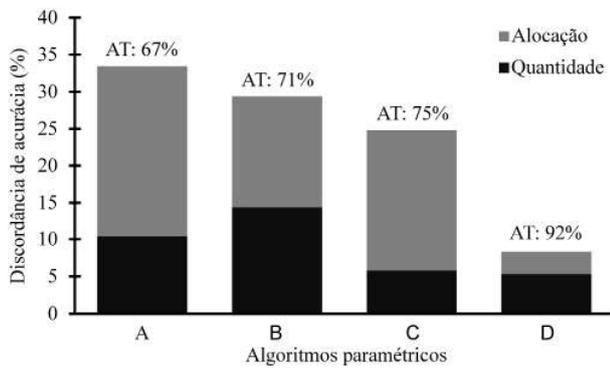


Fig. 7 - Discordância de acurácia (%) e sua acurácia total em % (AT) dos mapas gerados pelos quatro tipos de algoritmos paramétricos utilizados. A – Distância de Mahalanobis; B – Distância Euclidiana mínima; C – Paralelepípedo; D – Máxima verossimilhança.

Apesar de fundamental, a análise da acurácia dos mapas deve ser interpretada com cautela (CONGALTON, 1991). O melhor mapa, com bom desempenho na avaliação da acurácia, não necessariamente é um bom mapa (FOODY, 2002). Por exemplo, considerando que as classes aqui utilizadas refletem uma variação composicional e estrutural da vegetação, é de se esperar que um bom mapa reflita essa variação. Caso contrário, as classes adotadas poderiam ser consideradas muito artificiais, sem nenhum significado ecológico e, portanto, não atingiriam o objetivo de monitoramento da vegetação (CONGALTON, 1991; FOODY, 2002). De fato, a classe com fisionomia mais distinta das demais, a Savana florestada, foi a única em que todos os *pixels* analisados foram classificados corretamente e com baixa confusão pelo classificador de máxima verossimilhança.

As classes que apresentaram maior confusão foram a Savana gramíneo-lenhosa e a Savana gramíneo-lenhosa rupestre, que são muito similares em termos de estrutura e cobertura foliar, diferenciando-se apenas pela composição florística e pelo substrato, Neossolo Quartzarênico e afloramentos de arenito respectivamente (VELOSO *et al.*, 1991; JUNCA *et al.*, 2005). Além disso, mesmo se diferenciando pela estrutura do substrato, ambos os tipos de cobertura compartilham o principal componente do mesmo, que é o quartzo (JUNCA *et al.*, 2005).

Apesar disso, o método de máxima verossimilhança produziu classes que apresentam diferenças estatisticamente significativas nos valores do índice de vegetação EVI2 ($F=10473$, $df=4998$, $R^2=0.68$, $p<0.0001$). Somado a isso, o teste *post-hoc* revelou que essa diferença estatística foi detectada em todas as comparações par-a-par entre as classes de cobertura, considerando um intervalo de confiança de 99.9%. Isso indica que o classificador de máxima verossimilhança foi eficiente em detectar as diferenças na quantidade e estrutura da vegetação entre as classes (Fig. 8). Essas variações estruturais da vegetação podem refletir diferenças composicionais biologicamente relevantes e fundamentais para a conservação. Por exemplo, a espécie *Syngonanthus mucugensis* Giul. (Sempre-Viva), endêmica da região estudada, é ameaçada de extinção pela perda de habitat e extrativismo predatório e é associada às fitofisionomias Savana gramíneo-lenhosa (CERQUEIRA *et al.*, 2008). Portanto, a localização e a quantificação dessa fitofisionomia são fundamentais para a delimitação de áreas de conservação, bem como para a definição de medidas de manejo.

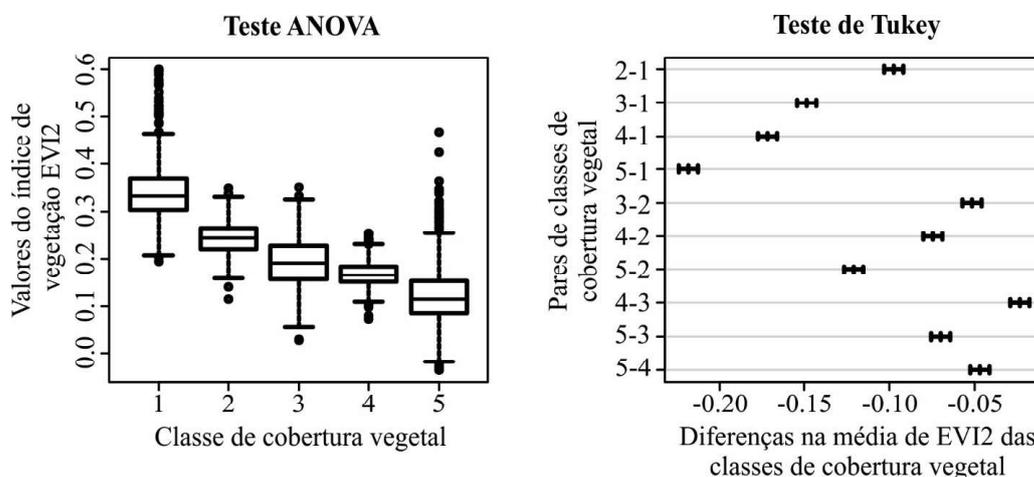


Fig. 8 - Diferença dos valores do índice de vegetação EVI2 entre as classes de cobertura vegetal de acordo com o classificador de máxima verossimilhança.

4. CONCLUSÃO

Com base nos resultados pode-se concluir que classificadores estatísticos paramétricos podem ser utilizados no mapeamento de fitofisionomias de Savana necessários para o monitoramento da vegetação. Os resultados indicam que o algoritmo classificador baseado na máxima verossimilhança é o mais acurado, bem como é eficiente em produzir uma classificação que reflete a variação da estrutura da vegetação, sendo sensível as diferenças entre as fitofisionomias de Savana. Nesse sentido, as técnicas aqui empregadas podem ser aplicadas no monitoramento de unidades de conservação de regiões predominantemente ocupadas por fitofisionomias savânicas, contribuindo assim com a conservação da biodiversidade nesses ecossistemas.

AGRADECIMENTOS

Registra-se agradecimentos ao Conselho Nacional de Pesquisa (CNPq), pelo auxílio financeiro fornecido para a aquisição das imagens de satélites de alta resolução (KOMPSAT-2 e SPOT-5), que permitiu a realização desse trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

ADREOLA, R. **Support Vector Machines na Classificação de Imagens Hiperespectrais**, 130f. Dissertação (Mestrado em Sensoriamento Remoto) – Centro Estadual de Pesquisa em Sensoriamento Remoto e Meteorologia. Universidade Federal do Rio Grande do Sul. 2009.

AZEVEDO, P.V.& SILVA, G. B. Potencial Agroclimático da Região da Chapada Diamantina no Estado da Bahia. **Revista Brasileira de Meteorologia**, v. 15, p. 77-88, 2000.

BEYER, H. L. 2004. **Hawth's Analysis Tools for ArcGIS**. Disponível em: <http://www spatialecology.com/htools>. Acesso: 20 novembro 2014.

CERQUEIRA, C. O.; FUNCH, L. S.; BORBA, E. L. Fenologia de *Syngonanthus mucugensis* Giul. subsp. mucugensis e *S. curralensis* Moldenke (Eriocaulaceae), nos municípios de Mucugê e Morro do Chapéu, Chapada Diamantina, BA, Brasil. **Acta Botanica Brasilica**, v. 22, n. 4, p.

962–969, dez. 2008.

CONGALTON, R. G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. **Remote Sensing of Environment**, v. 46, n. October 1990, p. 35–46, 1991.

COSTA, F. P. M. Use of SAR Satellites for Mapping Zonation of Vegetation Communities in the Amazon Floodplain. **International Journal of Remote Sensing**, v. 25, n. 10, p. 1817-1835, 2004.

COX, D. R. **Principles of Statistical Inference**. Cambridge: Cambridge University Press, 2006. p. 290p.

FLORENZANO, G. T. **Iniciação em Sensoriamento Remoto**. São Paulo, Oficina de Textos, 2011. 128p.

FOODY, G. M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, v. 80, n. 1, p. 185–201, abr. 2002.

GOTELLI, J. N. & ELLISON, M. A. **Princípios em Estatística em Ecologia**. Brasil, Artmed, 2011. 528p.

HUBERT-MOY, L.; DU, L. L.; CHARDIN, A.; PEREZ, P. A Comparison of Parametric Classification Procedures of Remotely Sensed Data Applied on Different Landscape Units. **Remote Sensing of Environment**, p. 174-187, 2001.

JACINTHO, C. R. L. **Geoprocessamento e Sensoriamento Remoto como Ferramentas na Gestão Ambiental de Unidades de Conservação O Caso da Área de Proteção Ambiental (APA) do Capivari-Monos, São Paulo**. 2003. 110f. Dissertação (Mestrado em Recursos Minerais e Hidrogeologia) – Instituto de Geociência. Universidade de São Paulo. 2003.

JIANG, Z.; HUETE, A.R.; DIDA, K. MIURA, T. Development of a Two-Band Enhanced Vegetation Index without a Blue Band. **Remote Sensing of Environment**, v. 112, p. 3833-3845, 2008.

JUNCÁ, F. A.; FUNCH, L.; ROCHA, W. **Biodiversidade e Conservação da Chapada Diamantina**. Brasília: Ministério do Meio Ambiente, 2005. p. 411

- KÖPPEN, W. **Climatologia: Conun Estudio de los Climats de La Tierra**. México, Fondo de Cultura Econômica, 1948. 479p.
- LANG, S. & BLASCHKE, T. **Análise da Paisagem com SIG**. São Paulo, Oficina de Textos, 2007. 424p.
- LU, D. & WENG, Q. A Survey of Image Classification Methods and Techniques for Improving Classification Performance. **International Journal of Remote Sensing**, v. 28, n. 10, p. 823-870, 2007.
- MDA Federal (2004), **Landsat GeoCover ETM+ 2000 Edition Mosaics Tile MES-24-10_LL_2000**, 1.0, USGS, Sioux Falls, South Dakota, 2000.
- MENESES, P. R. & ALMEIDA, de T. **Introdução ao Processamento de Imagens de Sensoriamento Remoto**. Brasília. CNPq, 2012. 266p.
- NOVO, M de. L. M. E. **Sensoriamento Remoto: Princípios e Aplicações**. Blucher, 2010. 387p.
- PONTIUS, Jr. G. R. & MILLONES, M. Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment. **International Journal of Remote Sensing**, v. 32, p. 4407-4429, 2011.
- PONZONI, F. & SHIMABUKURO, Y. **Sensoriamento remoto no estudo da vegetação**. São José dos Campos: Parêntese, 2007. p. 127
- PRIMACK, B. R. & EFRAIM, R. **Biologia da Conservação**. Londrina, Planta, 2001. 328p.
- QUINN, G. P. & KEOUGH, M. J. **Experimental Design and Data Analysis for Biologists**. Cambridge, 2002. 557p.
- RICHARDS, A. J. & JIA, X. **Remote Sensing Digital Image Analysis: An Introduction**. Germany, Springer, 2006. 439p.
- R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. **R Foundation for Statistical Computing**, Vienna, Austria. Disponível em: <www.R-project.org>. Acesso: 15 setembro 2014.
- SIIS. **Technical Introduction of KOMPSAT**. Disponível em: <http://www.si-imaging.com/lfile/Technical_Introduction_of_KOMPSAT_v1.0.pdf>. Acesso em: 8 ago. 2015.
- SINTHUMULE, N.; MUNYATI, C. Quantifying Savanna Woody Cover in the Field and on Historical Imagery: A Methodological Analysis. **South African Journal of Geomatics**, v. 3, n. 2, p. 113–127, 2014.
- USGS (2014)a, **Shuttle Radar Topography Mission**, 1 Arc Second scene SRTM1S14W042V3, Void Filled finished, Global Land Cover Facility, University of Maryland, College Park, Maryland, February 2000.
- USGS (2014)b, **Shuttle Radar Topography Mission**, 1 Arc Second scene SRTM1S13W042V3, Void Filled finished, Global Land Cover Facility, University of Maryland, College Park, Maryland, February 2000.
- VELOSO, P. H.; FILHO, R. R. L. A.; LIMA, A. C. J. **Classificação da Vegetação Brasileira, Adaptada a um Sistema Universal**. Fundação Instituto Brasileiro de Geografia e Estatística IBGE, 1991. 117p.