

Avaliação da delimitação subfisionômica da caatinga a partir da aplicação de análise discriminante em valores de reflectância de imagem do sensor Landsat TM 5

Leilton Santos Damascena¹
Luiz Rogério de Lima Macedo²

^{1,2} Pós-graduações em Modelagem em Ciências da Terra e do Ambiente
Universidade Estadual de Feira de Santana
Av. Transnordestina S/N. Novo Horizonte, Bahia, Brasil
l.s.damascena@gmail.com
rogerio2007mais@gmail.com

Abstract: The objective that guided this study was to evaluate changes in spectral subformations Savannah Steppe (Caatinga) and areas of irrigated agriculture bounded by Veloso (1991) from a technical application of multivariate statistical analysis (discriminant analysis), with the objective of verifying the possibilities of separation of these categories in periods of drought. For its development we used a LANDSAT 5 TM, orbit 216, Item 066, date 20/11/2009, converted to physical values of reflectance to unify the scale of the bands. Data collection was performed by guiding the mapping PROBIO (Conservation and Sustainable Use of Biological Diversity) for choosing the sample areas. The pixels chosen were selected by the operator using the following criteria: i) the more central areas of the class ii) areas not visually degraded, or iii) areas known in the field. The results from the application of discriminant analysis showed a real differentiation between categories. On the other hand, have accused some pixels near boundaries of other categories, probably due to leaf fall pattern of the vegetation types. The bands that best discriminate the categories analyzed in this work were the first band, band 2, band 3 of Landsat 5 TM, covering the entire visible spectrum. Unexpectedly contrasting the importance of using the infrared bands for the classification of vegetation types.

Palavras-chave: Spectral variation, Caatinga, Discriminant Analysis, Variação espectral, Caatinga, Análise Discriminante

1. Introdução:

A delimitação de grupos vegetacionais pelo sistema atual de classificação do Brasil tem nas características fisionômicas o seu principal referencial. Atrélado a este referencial encontramos o uso de espécies marcadoras para cada fisionomia, o que demonstra a influência de sistemas anteriores, que tinham sua base fundamentada sobre a distribuição das espécies. Do ponto de vista do uso do sensoriamento remoto, as características estruturais são captadas pelos sistemas de sensores imageadores dos satélites, de maneira mais objetiva que a análise da distribuição das espécies, que depende diretamente do número de coletas e da distribuição destas.

No intuito de construir crivos espectrais para as categorias, vamos analisar a vegetação da Caatinga, que no sistema proposto por Veloso (1991) é chamada de Savana Estépica. No entanto, a delimitação da categoria Savana Estépica não corresponde aos limites realizados pela província Caatinga ou bioma Caatinga, abarcando áreas disjuntas da região nordestina “estado de Roraima e duas outras áreas também disjuntas chaquenhas, uma no extremo sul do Mato Grosso do Sul e outra na barra do Rio Guará, no Rio Grande do Sul (P. 92, Veloso 1991)”.

O mapeamento proposto por Veloso *et al.* (1991) utiliza-se das características fisionômicas para a delimitação dos grupos vegetacionais. Utilizando tal critério foram delimitados para a região da Savana Estépica (Caatinga) os seguintes grupos: Savana Estépica Florestada (Td), Savana Estépica Arborizada (Ta), Savana Estépica Parque (Tp) e Savana Estépica gramíneo-lenhosa (Tg). A definição destes grupos está apoiada sobre o declínio gradual do porte da vegetação, relacionada a vários fatores ambientais. A unidade Savana Estépica gramíneo-lenhosa, não foi analisada neste trabalho, pois no mapeamento realizado pelo Projeto de Conservação e Utilização Sustentável da Diversidade Biológica-PROBIO (para a Caatinga), a sua área de abrangência é muito reduzida e sua delimitação questionável.

O comportamento espectral resultante da vegetação é constituído por múltiplas interações complexas, com diversos materiais que compõem a vegetação (Huete, 1988). No entanto, de forma geral, a característica mais conspícua dentre a mistura espectral da vegetação é a resposta realizada pela parte foliar. Então, por generalização, o comportamento de reflectância da vegetação é aproximado da assinatura espectral das folhas (Ponzoni e Shimabukuro, 2007).

Contudo, é importante salientar que a vegetação identificada como Caatinga apresenta uma estrutura vegetacional predominante constituída por árvores e arbustos, e, em algumas formações, o estrato herbáceo é relevante. O estrato Arbóreo-arbustivo apresenta um padrão irregular na construção de sua copa, havendo a presença de fuste múltiplo, permitindo que a luminosidade penetre até o solo (Queiroz, 2009). Por isso, é necessário levar em consideração o marcante ciclo sazonal, que molda o comportamento das espécies.

O objetivo que norteou este trabalho foi avaliar, a partir da aplicação uma técnica de análise estatística multivariada (Análise Discriminante), as variações espectrais de cada subformações da Savana Estépica (Td, Ta, Tp) e áreas de agricultura irrigada (Ag) delimitadas por Veloso (1991), com o fim de discutir as possibilidades de separação destas categorias em períodos de estiagem.

2. Metodologia de trabalho

Para o desenvolvimento desta pesquisa foi utilizada uma imagem LANDSAT 5 TM, órbita 216, ponto 066, data 20/11/2009 que recobre parte norte do estado da Bahia e sul de Pernambuco, abrangendo áreas com precipitação de aproximadamente 400 mm/ano. A imagem foi processada para possibilitar a coleta dos dados. O primeiro passo foi o georeferenciamento tendo com base imagens do PROBIO, que foram utilizadas para o desenvolvimento do mapeamento da Caatinga, logo em seguida, a imagem foi convertida em valores físicos de reflectância para unificar a escala das bandas.

A coleta dos dados teve como norteador o mapeamento realizado pelo PROBIO para a escolha das áreas de amostra. Os pixels escolhidos foram selecionados pelo operador através de critérios: i) Áreas mais centrais da categoria, ii) Áreas visualmente não degradadas, ou iii) Áreas reconhecidas em campo. Foram realizadas quatro coletas de nove pixels contíguos em cada categoria, sendo extraído os valores de reflectância nas 6 bandas, para a construção de uma planilha de dados para aplicação da análise de discriminantes.

A utilização de uma única imagem é um limite para a discriminação entre os grupos vegetacionais, mas é preciso levar em conta que a homogeneidade do comportamento da vegetação no período de estiagem é a principal dificuldade para a classificação da cobertura

vegetacional, sendo assim, a aplicação de métodos estatísticos que possam analisar a consistência do sistema em um período de senescência foliar contribui para verificar a adequação dos critérios atuais de delimitação das subformações (Veloso 1991).

A análise de discriminantes permite a discriminação entre categorias pelo estabelecimento de pesos as variáveis estatísticas, para maximizar a variância entre os grupos, no caso os valores de reflectância para cada banda da imagem.

Estas variáveis são sintetizadas na função discriminante linear de Fisher:

$$Z_{jk} = a + W_1X_{1k} + W_2X_{2k} + \dots + W_nX_{nk} \quad (1)$$

Onde,

Z_{jk} = escore Z discriminante da função discriminante j para o objeto k.

a = intercepto

W_1 = peso discriminante para variável independente 1.

X_{1k} = variável independente 1 para o objeto k.

A utilização da análise de discriminantes pode nos oferecer três objetivos básicos:

- i. Identificação das variáveis que melhor discriminam entre dois ou mais grupos mutuamente exclusivos;
- ii. A utilização destas variáveis para criar uma função parcimoniosa entre os grupos;
- iii. A utilização desta função para classificar a priori novos indivíduos nos grupos.

Estas respostas serão fundamentais para responder a consistência da discriminação entre as categorias vegetacionais propostas por Veloso (1991) para a região da Savana Estépica (Caatinga), incluindo áreas de plantio irrigado.

Para a aplicação desta técnica é necessário obedecer algumas critérios: normalidade multivariada das variáveis dependentes; homogeneidade das matrizes de variância e co-variância; ausência de multicolinearidade e linearidade.

A primeira condição busca definir que as variáveis independentes foram amostradas aleatoriamente escolhidas na população e que a distribuição dos seus valores é normal (Tabachnick & Fidel, 2001). A condição de homogeneidade das matrizes explica que a sua desigualdade pode afetar negativamente o processo de classificação (Hair, 2005). E a última condição denota a dependência entre as variáveis e o poder explicativo que cada uma destas pode oferecer para a análise (Hair, 2005).

A classificação dos grupos, ou a determinação de escores de corte também chamado de Z crítico, dependendo do tamanho dos grupos podem ser definidos como escores de grupos de tamanho iguais ou diferentes. Os grupos propostos neste trabalho são iguais, sendo adotado: (Figura 1).

Escore de corte para dois grupos de mesmo tamanho,

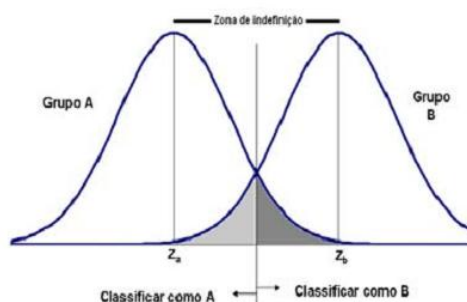
$$ZCE = (ZA + ZB) / 2 \quad (2)$$

Onde,

ZCE = valor do escore de corte crítico para grupos de mesmo tamanho.

ZA = centróide do grupo A.

ZB = centróide do grupo B.



Fonte: Hair (2005, p. 225).

Figura 1. Escore de corte = Z_{CE}

A análise discriminante foi realizada no software SPSS® versão 17.0, utilizado como ferramenta para execução dos testes estatísticos a um nível de significância de 0,05.

3. Resultados e discussão

A análise preliminar das variáveis independentes é realizada através da diferença entre as médias, para verificação da correspondência da contribuição entre as variáveis. As bandas aceitas para fazer parte da análise de discriminantes tem que possuir menos de 95 % de concordância, pois se os dados de uma variável são muito similares à outra variável, a explicação da distribuição dos grupos é pouco afetada por esta última variável.

Na tabela 1 o teste demonstra com 95 % de confiança que as bandas da imagem são dispare, podendo todas elas participar da análise de discriminantes.

Tabela 1. Teste de igualdade entre as médias dos grupos

VARIÁVEIS	Lambda de Wilks	F	gl1	gl2	Sig.
BANDA 1	.201	185.530	3	140	.000
BANDA 2	.229	156.743	3	140	.000
BANDA 3	.231	155.240	3	140	.000
BANDA 4	.282	118.536	3	140	.000
BANDA 5	.265	129.615	3	140	.000
BANDA 7	.231	155.743	3	140	.000

O teste de homogeneidade das matrizes de variância a co-variância (Tabela 2) não foi atendido, para os dados aqui apresentados. A regra do método diz que ao utilizar em sua aplicação amostras de tamanho pequeno e matrizes de tamanhos diferentes a significância

estatística pode ser afetada adversamente. No entanto, o número de amostras aqui colocadas supera, em quantidade, a maioria daquelas encontrados na literatura consultada, além de apresentarem tamanhos iguais. Este foi o único teste que não foi atendido, mas isto não é um impedimento para o uso da estatística.

Tabela 2. Teste de homogeneidade das matrizes de variância e covariância.

M de Box		719.168
F	Aprox.	10.536
	gl1	63
	gl2	45924.921
	Sig.	.000

Seguiu-se a análise dos autovalores (Tabela 3) que apresentam a medida relativa de quão diferente são os grupos na função discriminantes. Encontramos, para nossas categorias três funções, sendo que a primeira concentra mais de 60 % da variância das categorias. Neste mesmo quadro encontramos a correlação canônica, que demonstra o nível de associação entre os escores discriminantes e os grupos. Para o tratamento do escore em porcentagem é importante elevar ao quadrado o escore (Hair, 2005). Na primeira função encontramos mais de 85% de correlação com as variáveis canônicas. As outras funções também apresentam alta correlação, mas a variância explicada pelas funções é pequena.

Tabela 3. Autovalores das funções discriminantes.

FUNÇÃO	Autovalor	% VARIÂNCIA	% ACUMULADA	CORRELAÇÃO CANÔNICA
1	5,800 ^a	60.4	60.4	.924
2	2,525 ^a	26.3	86.7	.846
3	1,278 ^a	13.3	100.0	.749

O próximo teste é o de Lambda de Wilks, que testa a significância das funções determinantes, com base nas diferenças de médias entre os grupos, para este teste buscamos uma diferença de significância de $\alpha = 0.05$. Esta foi encontrada em todas as funções (Tabela 4).

Tabela 4. Teste de Lambda de Wilks.

Contraste entre as funções	Lambda de Wilks	Qui-quadrado	gl	Sig.
1 a la 3	.018	551.983	18	.000
2 a la 3	.125	287.455	10	.000
3	.439	113.606	4	.000

A matriz dos coeficientes estruturais apresentam em cada função as variáveis mais correlacionadas (Tabela 5). Para a função 1 temos as bandas 1, 3 e 2 como as variáveis como maior poder de distinção entre os grupos vegetacionais.

Tabela 5. Matriz de estrutura.

VARIÁVEIS	FUNÇÃO		
	1	2	3
BANDA 1	,731	.522	.383
BANDA 3	,666	.350	.588
BANDA 2	,615	.580	.500
BANDA 4	-.425	,751	.229
BANDA 7	.679	-.009	,718
BANDA 5	.605	-.005	,717

O próximo item (Tabela 6) apresenta as variáveis selecionadas para compor as funções e seus respectivos coeficientes não padronizados.

Tabela 6. Coeficientes da função Discriminantes Canônica.

VARIÁVEIS	FUNÇÃO		
	1	2	3
BANDA 1	,731	.522	.383
BANDA 3	,666	.350	.588
BANDA 2	,615	.580	.500
BANDA 4	-.425	,751	.229
BANDA 7	.679	-.009	,718
BANDA 5	.605	-.005	,717

Portanto, as funções serão descritas:

$$Z_1 = -7.145 + 169.903 \text{ BANDA 1} + 86.004 \text{ BANDA 2} - 92.970 \text{ BANDA 3}$$

$$Z_2 = -13.23 + 12.326 \text{ BANDA 4}$$

$$Z_3 = 1.191 + 32.311 \text{ BANDA 5} - 72.643 \text{ BANDA 7}$$

O resultado da função de centroides dos grupos (Tabela 7) permite o cálculo do Z crítico, através da equação: $ZCE = (ZA + ZB) / 2$ (1). Esta função definirá os valores que limitam os grupos de vegetação.

Para a primeira categoria (Savana Estépica Parque) temos:

$$\text{I. } ZCE = (3.257 + (-1.425))/2 \rightarrow 0.916$$

$$\text{II. } ZCE = (3.257 + 1.113)/2 \rightarrow 2.185$$

$$\text{III. } ZCE = (3.257 + 2.946)/2 \rightarrow 0.155$$

Classificação através dos escores críticos:

$$(1) \quad T_p < 0.916 < T_d$$

$$(2) \quad T_p < 2.185 < T_a$$

$$(3) \quad T_p < 0.155 < A_g$$

$$(4)$$

Tabela 7. Funções de centróides dos grupos.

Categorias de vegetação	Função		
	1	2	3
1,0	3.257	1.222	.796
2,0	-1.425	-1.904	1.202
3,0	1.113	-1.154	-
4,0	-2.946	1.836	1.667

A função de *Fisher* pode ser utilizada para a classificação dos valores de cada banda, conjuntamente com os valores críticos, buscando definir os grupos a partir destas variáveis.

Por fim, foram validados estes resultados com base no processo de validação interna, que consiste em verificar a eficácia da classificação das observações originais e a validação cruzada (Tabela 8).

Tabela 8. Resultado da classificação cruzada. ^{b,c}

categorias de vegetação		Previsão de membros do grupo				Total	
		1,0	2,0	3,0	4,0		
Original	contagem	1,0	35	0	1	0	36
		2,0	0	36	0	0	36
		3,0	2	0	34	0	36
		4,0	0	1	3	32	36
	%	1,0	97.2	.0	2.8	.0	100.0
		2,0	.0	100.0	.0	.0	100.0
		3,0	5.6	.0	94.4	.0	100.0
		4,0	.0	2.8	8.3	88.9	100.0
Validação cruzada ^a	contagem	1,0	34	0	2	0	36
		2,0	0	36	0	0	36
		3,0	4	0	32	0	36
		4,0	0	2	3	31	36
	%	1,0	94.4	.0	5.6	.0	100.0
		2,0	.0	100.0	.0	.0	100.0
		3,0	11.1	.0	88.9	.0	100.0
		4,0	.0	5.6	8.3	86.1	100.0

A classificação correta dos casos dos grupos originais somam 95.1%, já a classificação cruzada chega há 92.4%. O poder de diferenciação entre os grupos foi elevado, não alcançando valores absolutos pelas características homogêneas dos grupos vegetacionais da caatinga, principalmente entre Ta e Tp, que são diferenciadas em algumas áreas.

Com a função de discriminantes utilizada neste trabalho explicamos 60% da variação dos grupos vegetacionais aqui estudados, com base nas três bandas do visível da imagem Landsat 5 TM. Mas, quando observamos as funções discriminantes propostas com as bandas 4, 5 e 7, encontramos conjuntamente a explicação para aproximadamente 30% da variação das categorias.

4. Conclusões

Os resultados encontrados a partir da aplicação da Análise Discriminante apresentou uma real diferenciação entre as categorias. De outro lado, alguns pixels apresentaram limites próximos de outras categorias. Estes fatores são ocasionados pelo padrão de queda foliar entre os tipos vegetacionais, sendo que são confundidos os tipos Tp e Ta, Td e Ag. Esta sobreposição é esperada entre estas categorias pelas características espectrais similares.

As bandas que melhor discriminaram as categorias analisadas neste trabalho foram as banda 1, banda 2, banda 3 do Landsat 5 TM, abarcando todo espectro visível. Inesperadamente contrapondo a importância da utilização das bandas do infravermelho para a classificação dos tipos vegetacionais.

Agradecimentos

Agradeço pós-graduação em Ciências da Terra e do Ambiente pela disponibilização da estrutura física e aos professores da pós.

5. Referências Bibliográficas

- Veloso, H. P.; Antônio, L. R. R. F.; Jorge, A. L. **Classificação da Vegetação Brasileira, adaptada a um sistema universal**. Rio de Janeiro: CDDI, 1991. 124 p.
- Hair, Joseph F.; Anderson, Rolph E.; Tatham, Ronald L.; Black, William C. **Análise multivariada de dados**. Porto Alegre: Bookman, 2005. 688 p.
- Tabachnick, Barbara G.; Fidell, Linda S. **Using multivariate statistics**. 4 ed. Boston: Allyn and Bacon, 2001. 500 p.
- HUETE, A.R. A soil-adjusted vegetation index. **Remote Sensing of Environment**, Elsevier Science Publishing Co., New York, USA. v.25 n. 295-309p.1988.
- PONZONI, F.J., SHIMABUKURO, Y.E. **Sensoriamento remoto no estudo da vegetação**. São José dos Campos: Parêntese. 2007. 127 p.
- QUEIROZ, L.P. de. **Leguminosas da caatinga**. Feira de Santana, Bahia. 2009. 467 p.