

Método de Identificação Espectral (MIE): Novo Classificador Baseado no Método ANOVA e no *Spectral Correlation Mapper* (SCM)

OSMAR ABÍLIO DE CARVALHO JÚNIOR.¹;
ANA PAULA FERREIRA DE CARVALHO^{2 E 3},
PAULO ROBERTO MENESES⁴,
RENATO FONTES GUIMARÃES¹;

¹Departamento de Geografia - Universidade de Brasília (UnB)- Campus Universitário Darcy Ribeiro, Asa Norte, 70910-900, Brasília, DF, Brasil

osmana@tba.com.br

²Departamento de Ecologia - Universidade de Brasília (UnB)- Campus Universitário Darcy Ribeiro, Asa Norte, 70910-900, Brasília, DF, Brasil

³INCRA – SBN Ed. Palácio do desenvolvimento, sala 1205, 70057-900, Brasília, DF, Brasil

⁴Instituto de Geociências - Universidade de Brasília (UnB)- Campus Universitário Darcy Ribeiro, Asa Norte, 70910-900, Brasília, DF, Brasil

Abstract The spectral classifiers generate images that express the probability of the sought material according to a similarity parameter. The value of the similarity parameter that attests the existence of the material varies for each sought spectral curve being accomplished in a manual way by the user. The present work shows the development of a new spectral classifier, denominated of Method of Spectral Identification (MSI), in which supplies estimates, according to confidence levels, of the existence or not of the sought curve. The proposed method is based on two procedures: the algorithm of similarity of the spectral classifier MSI and the statistical method ANOVA. This method generates the image of the similarity parameter so much MIS as three relative binary images the existence of the material according to confidence levels.

Keywords: Spectral Mixture, ANOVA, Spectral Correlation Mapper, Imaging Spectroscopy.

1. Introdução

As técnicas de análise espectral permitem uma boa estimativa para a identificação do material procurado pela similaridade entre a curva de referência e a da imagem. Dentro desse contexto destacam-se os classificadores espectrais *Spectral Angle Mapper*-SAM (Kruse *et al.* 1992, Kruse *et al.* 1993a e b), *Spectral Correlation Mapper* - SCM (Carvalho & Meneses, 2000), *Spectral Feature Fitting* - SFF e o coeficiente de correlação R utilizado pelo Tricorder (Clark & Swayze, 1995).

No entanto, um questionamento existente é a determinação do grau de similaridade que permita validar estatisticamente a existência do elemento procurado. Nos métodos citados os valores que atestam a existência do material é determinada pelo usuário. Questiona-se, portanto, qual o valor do parâmetro de similaridade que confirma a existência do elemento procurado? É

importante destacar que o limite de detecção é variável e dependente do material analisado e da correlação existente entre os materiais em análise. Essa variabilidade dificulta uma delimitação automatizada do material procurado.

Nesse artigo é proposto um Método de Identificação Espectral (MIE) que fornece um novo índice de similaridade e três estimativas segundo níveis de significância das possíveis áreas do material procurado. Esse método fundamenta-se em dois procedimentos estatísticos: ANOVA (Davis, 1973; Steel & Torrie, 1980; Vieira, 1988; Souza, 1998) e coeficiente SCM (Carvalho & Meneses, 2000). Essa informação pode ser utilizada para avaliar o grau de inter-correlação dos materiais em análise, sendo interessante em uma análise exploratória.

2 Método ANOVA

O emprego da ANOVA pode ser sintetizado em três procedimentos

2.1 Formulação da Hipótese

Pela análise de variância, a inclusão de uma curva observada como pertencente ao grupo espectral de uma curva referência é realizada através de um teste de hipóteses. Neste teste são avaliadas duas alternativas:

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

Estabelece-se uma regra de decisão baseada na inclinação **b** da reta de regressão. Se o coeficiente de inclinação **b** é próximo de 0, conclui-se que o espalhamento em Y_r (Estimativa de Y da regressão linear) é baixo, sendo próximo ao Y médio e, conseqüentemente, com reduzido grau de ajuste entre a reta predita e os dados.

Portanto, confirmando-se a hipótese H_0 , o teste instituído descarta a curva observada como sendo da mesma população que a curva padrão. Caso contrário, obtendo-se valores condizentes com a hipótese H_1 tem-se que a curva observada é pertencente ao grupo da curva padrão.

2.2 Nível de Significância

Não é possível estabelecer uma regra que permita rejeitar H_0 , sem que se esteja sujeito a erro. Assim, embora a regra de decisão seja condizente com os dados, não é possível estar certo que a decisão tomada é a correta. Na análise de hipótese pode-se deparar com dois tipos de erros:

- Erro tipo I – consiste em rejeitar H_0 , sendo que H_0 é verdadeira; e
- Erro tipo II – consiste em aceitar H_0 , sendo que H_0 é falso.

A probabilidade da ocorrência do erro tipo I é denominada, no teste de hipótese, como nível de significância (α). Um nível de significância de 10% consiste em dizer que existe uma probabilidade de 10% dos dados possuírem $\beta = 0$ e serem considerados como $\beta \neq 0$. Desta forma, observa-se uma margem de 10 % dos dados que não são oriundos da população do padrão e que erroneamente estão classificadas como dela pertencente. Deve-se atentar que a confirmação da hipótese é na realidade uma negação de que a curva pertença ao padrão.

A colocação do erro dessa forma evita eliminar dados bons pois a área selecionada apresenta-se um pouco maior do que deveria ser de acordo com o nível de significância adotado.

2.3 O Teste F

Para a validação ou não da hipótese é empregado o teste F. Para proceder ao teste realiza-se inicialmente a estimativa de um índice, denominado de F:

$$F = \frac{\text{MQr}}{\text{MQd}} \quad \text{eq. 1}$$

Onde:

MQr – a média da soma dos quadrados da regressão

MQd – a média da soma dos quadrados do desvio

O procedimento do teste consiste em rejeitar H_0 para todo valor de F calculado igual ou maior do que o valor estabelecido segundo uma tabela de distribuição de (Vieira, 1988). O valor de F segundo a tabela de distribuição é função do nível de significância e dos graus de liberdade do numerador e denominador.

O numerador MQr possui apenas um grau de liberdade pois baseia-se somente em duas observações: a curva padrão fornecida e a curva observada da imagem em análise. Portanto, no presente caso o somatório do quadrado da regressão (SQr) é igual à média do quadrado da regressão (MQr).

O grau de liberdade do denominador é expresso como sendo (N-2). A utilização literal do número de bandas para o cálculo dos graus de liberdade proporciona erros para feições com alta inter-correlação. Para as imagens hiperespectrais as bandas são muito redundantes, principalmente, considerando-se apenas a faixa relativa à feição de absorção (≈ 10 bandas). O emprego de uma APC ou MNF demonstra que para uma faixa relativa a uma feição de absorção apenas 3 componentes são responsáveis pela quase a totalidade da variância com grau de liberdade igual a 1. Portanto, no presente método o valor de N, em vez de ser expresso pelo número de bandas, é representado pela sua real dimensionalidade que é obtido pela APC ou pelo MNF.

Para feições espectrais mais complexas e que ocupam maior número de bandas como a feição dupla da caulinita, a real dimensionalidade dos dados aumenta para 4 componentes tornando o grau de liberdade do denominador igual a 2.

Portanto, as condicionantes do método proposto são: graus de liberdade do denominador (1 ou 2), numerador (1) e o nível de significância (2,5%, 5% e 10%). Com esses dados compara-se o resultado de F obtido para cada pixel com os valores da tabela de F. Se o valor for menor recusa-se a hipótese de ser o alvo pretendido, do contrário aceita-se. A **Tabela 1**, conhecida como tabela de análise de variância, consiste em um resumo dos procedimentos estatísticos para a realização dessa análise.

Tabela 2 – ANOVA para uma Regressão Linear Simples (Davis, 1973)				
Origem da Variação	Soma dos quadrados	Graus de liberdade	Média dos Quadrados	F Teste

Regressão Linear	SQr	1	MQr	MQr/MQd
Desvio	SQd	N - 2	MQd	
Total da Variação	SQt	N - 1		

3 Formulação do Método de Identificação Espectral (MIE) Pela Integração da ANOVA com o Coeficiente (SCM)

A análise de variância de uma regressão linear não é sensível à presença da correlação negativa. Portanto, como complemento ao método proposto deve-se utilizar o coeficiente de correlação SCM (Carvalho & Meneses, 2000). Desta forma, pode-se segmentar as áreas com correlação negativa e adicioná-las aos dados obtidos com a ANOVA restringindo as áreas de real interesse. A **Figura 1** demonstra a conjugação dos dois fatores na formulação do método de identificação espectral. Como exemplificação são demonstradas as áreas com a feição da caulinita.

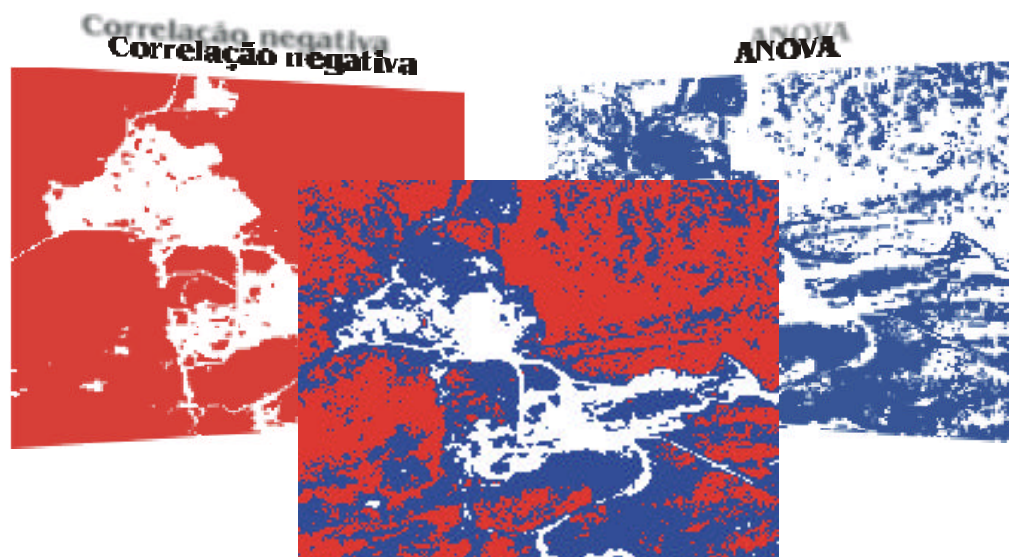


Figura 1. Soma dos dois parâmetros estatísticos restritivos às áreas sem a feição da caulinita em um nível de significância de 2.5%.

4 Análise das Imagens dos Graus de Significância

O presente método gera imagens binárias, relativas à existência ou não da informação pretendida, dentro de uma margem de significância de 2.5%, 5% e 10%. Como exemplificação é demonstrado o emprego desse método para as feições da pimelita/saponita ($2.19\mu\text{m} - 2.36\mu\text{m}$) e da vegetação ($0.54\mu\text{m} - 0.75\mu\text{m}$). Para essas feições utiliza-se o valor de graus de liberdade do denominador igual a 1 devido às suas características espectrais.

A partir das informações obtidas pela análise manual observa-se que a estimativa de 2.5% foi bastante eficiente na discriminação das áreas de interesse para ambas as feições (**Figura 2** e **Figura 3**). Os demais níveis de significância apresentam uma superestimativa da área. No caso do emprego do nível de significância de 10% observa-se que a informação gerada não corresponde à realidade pois muitas das áreas identificadas não apresentam o material procurado. Essa característica é devido à alta inter-correlação da feição procurada com as demais presentes na cena.



Figura 2 – Comparação das áreas definidas para a presença de pimelita pelo método MIE para os níveis de significância 2,5%, 5% e 10%.



Figura 3 – Comparação das áreas definidas para a presença de caulinita pelo método MIE para os níveis de significância 2,5%, 5% e 10%.

5 Análise da Imagem do Fator de Similaridade MIE

Esse coeficiente origina-se do fator F do método ANOVA conjugado com o método SCM, que permite a eliminação das áreas com correlação negativa e baixa. Os valores de MIE variam entre 0 a 1. O valor máximo desse coeficiente corresponde às áreas com nível de significância igual a 2,5%, enquanto que os valores mais baixos são relativos as áreas com SCM abaixo de 40% e de fator F abaixo de 5%. As imagens MIE proporcionam uma intensa expansão das áreas de maior correlação SCM, tornando-as um eficiente e robusto índice de similaridade. A **Figura 4** apresenta as imagens MIE para as feições da pimelita-saponita, vegetação sem nenhum tipo de realce.

Os índices MIE e SCM são descritos por uma função, quando dispostos em um diagrama de dispersão. A **Figura 5** apresenta os diagramas de dispersão SCM x MIE para as feições da pimelita-saponita e vegetação. Destaca-se a intensa dispersão do MIE para os valores de SCM com alta correlação. As curvas geradas estão segmentadas de acordo com os diferentes graus de significância. A área vermelha é relativa a significância de 2,5%, enquanto que a ciano representa a área de acréscimo para 5% e a azul 10%. A área amarela representa valores superiores de 10% que são desconsideradas na análise. As áreas com pimelita e vegetação verde estão todas contidas no pequeno segmento de valor de MIE igual a 1, no ápice da curva. Desta forma, esse índice estabelece uma grande diferença numérica entre os dados pretendidos e os demais, favorecendo sua segmentação.

O Método de Identificação Espectral (MIE) foi implementado pelo autor, em linguagem IDL, dentro do programa ENVI. Esse programa utiliza um espectro de referência e a imagem hiperespectral a ser analisada. O programa gera 4 imagens de saída, sendo que as três primeiras são referentes aos níveis de significância de 2.5%, 5.0% e 10% do método e a última, à imagem do índice de similaridade.

A imagem dos níveis de significância é de caráter binário podendo ser utilizada como uma máscara para delimitar as áreas de interesse para realizar tratamentos de semi - quantificação somente nas localidades onde se tem a feição.

6 Conclusão

O MIE é um novo método de classificação espectral que fornece estimativas segundo níveis de significância das áreas mais prováveis do material procurado. Portanto além de gerar uma imagem relativo ao parâmetro de similaridade o algoritmo realiza um teste de hipótese para definir quais seriam os melhores valores de corte. O método mostrou-se bastante eficaz apresentado excelentes estimativas para área testada. Além disso, o método possui um parâmetro de similaridade que elimina por completo os falsos positivos presentes nos algoritmos do SAM, SFF e Tricorder e ainda maximiza os valores das áreas onde existe o material procurado.

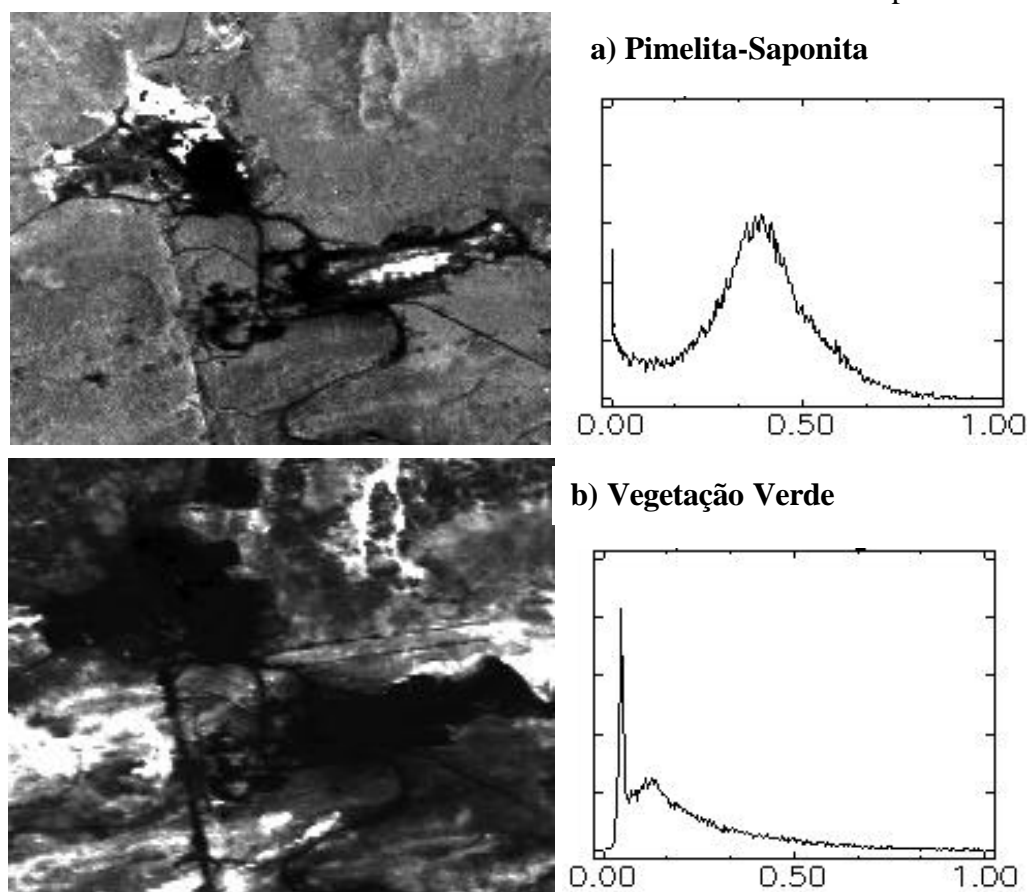
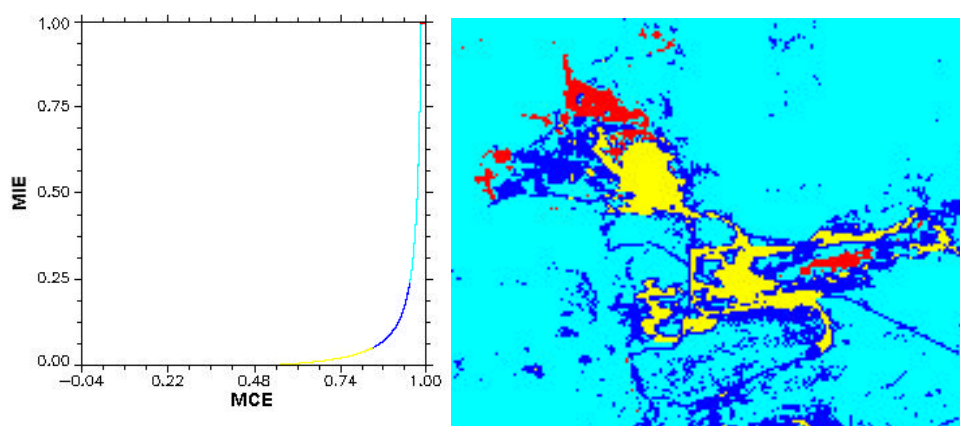


Figura 4 – Imagem MIE para a região da mina de Niquelândia com seu respectivo histograma das feições: a) pimelita e b) vegetação.

a) Pimelita-Saponita



b) Vegetação Verde

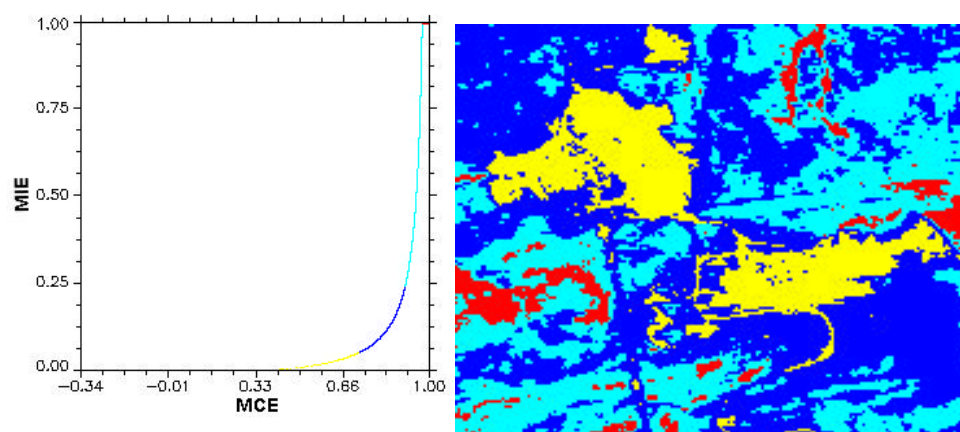


Figura 5 – Diagramas de dispersão entre SCM x MIE para as feições: a) pimenteira e b) vegetação verde. As curvas estão segmentadas de acordo com os graus de significância: 2,5% (vermelho), 5% (ciano) e 10% (azul). As áreas em amarelo são relativas a valores superiores a 10%.

Referências

- Carvalho Jr., O. A. & Menezes, P. R., 2000, Spectral Correlation Mapper (SCM): na Improving Spectral Angle Mapper. Presented in the *Nineth JPL Airborne Earth Science Workshop* (URL: <http://makalu.jpl.nasa.gov/docs/workshops/toc.htm>).
- Clark, R. N. & Swayze, G. A., 1995. Mapping Minerals, Amorphous Materials, Environmental Materials, Vegetation, Water, Ice and Snow, and Other Materials: The USGS Ticorder Algorithm. In: *Summaries of the Fifth JPL Airborne Earth Science Workshop*, JPL Publication 95-1 v.1, p.39-40
- Davis, 1973, *Statistics and analysis in geology* New York, John Willey & Sons, Inc. 550p
- Kruse, F. A.; Lefkoff, A. B.; Boardman, J. W.; Heidebrecht, K. B.; Shapiro, A. T.; Barloon, P. J. & Goetz, A. F. H., 1992. The Spectral Image Processing System (SIPS) - Software for Integrated Analysis of AVIRIS Data. *Summaries of the 4th Annual JPL Airborne Geoscience Workshop*, JPL Pub-92-14, pp 23-25.

- Kruse, F. A.; Lefkoff, B. & Dietz, J. B., 1993, Expert System-Based Mineral Mapping in Northern Death Valley, California/Nevada, Using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), *Remote Sens. Environ.*, Vol. 44, No.2, pp. 309-336.
- Kruse, FA; Lefkoff, A. B; Boardman, J. W.; Heiedbrecht, K. B.; Shapiro, A. T.; Barloon, P. J. & Goetz, A. F. H., 1993, The Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data. *Remote Sens. Environ.* 44:145-163.
- Souza, G. S., 1998, Introdução aos Modelos de Regressão Linear e Não-Linear, Brasília: EMBRAPA – SPI/EMBRAPA-SEA, 505pp.
- Steel, R. G. D.; Torrie, J.K., 1980, Principles and procedures of statistics. 2. Ed. New York: MacGraw-Hill.
- Vieira, S. 1988, Introdução à Biestatística. 5ed. Rio de Janeiro: Campus, 293p.