SPECTRAL IDENTIFICATION METHOD (SIM): A NEW CLASSIFIER BASED ON THE ANOVA AND SPECTRAL CORRELATION MAPPER (SCM) METHODS

OSMAR ABÍLIO DE CARVALHO JÚNIOR.[1]
ANA PAULA FERREIRA DE CARVALHO[2, 3]
PAULO ROBERTO MENESES[4]
RENATO FONTES GUIMARÃES[1]

[1]Departamento de Geografia - Universidade de Brasília (UnB)- Campus Universitário Darcy Ribeiro, Asa Norte, 70910-900, Brasília, DF, Brazil osmana@tba.com.br

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

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

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

Abstract: Spectral classifiers generate images that express the probability of the material sought according to a similarity parameter. The value of the similarity parameter that attests to the existence of the material varies for each spectral curve sought; this is done manually by the user. This work demonstrates the development of a new spectral classifier called the Spectral Identification Method (SIM). It supplies estimates according to confidence levels of the existence or not of the curve sought. The proposed method is based on two procedures: the algorithm of similarity of the SIM spectral classifier and the ANOVA statistical method. This method generates an image of the similarity parameter as much from the SIM as from three relative binary images of the existence of the material according to confidence levels.

Keywords: Spectral mixture, ANOVA, Spectral Correlation Mapper and imaging spectroscopy.

1 Introduction

Spectral classifiers allow good mapping of the materials based on spectral signatures: the *Spectral Angle Mapper-SAM* (Kruse *et al*. 1992, Kruse *et al*. 1993a and b), the *Spectral Correlation Mapper* - SCM (Carvalho & Meneses, 2000), the *Spectral Feature Fitting* - SFF and the correlation coefficient R used by Tricorder (Clark & Swayze, 1995).

However, an issue is the determination of the similarity degree that allows the statistical validation of the existence of the sought element. In the mentioned methods, the user determines the values that attest to the existence of the material. Is it asked, therefore, which is the value of the similarity parameter that confirms the existence of the sought element? It is important to highlight that the detection limit is variable and dependent on the analyzed material and the existent correlation among the materials in analysis. That variability hinders an automated delimitation of the sought material.

In this paper, the Spectral Identification Method (SIM) is proposed to establish a new similarity index and three estimates according to the levels of significance of materials. The method is based on two statistical procedures: ANOVA (Davis, 1973; Steel & Torrie, 1980; Vieira, 1988; Souza, 1998) and the SCM coefficient (Carvalho & Meneses, 2000). This information can be used to evaluate the degree of correlation among the materials in analysis.

2 The ANOVA Method

The variance analysis uses a hypothesis test in order to determine if the spectral image belongs to the spectral reference group. In this test two alternatives are appraised based on the slope coefficient of linear regression:

$$Ho: \beta = 0$$
$$H1: \beta \neq 0$$

If the slope coefficient is close to 0, it is possible to conclude that scatter in Yr (Estimate of Y about linear regression) is low, close to medium Y and with low fit regression. The determination of Ho hypothesis establishes that the spectral image doesn't belong to the same population as the pattern curve. The hypothesis analysis can have two types of errors:

- Error type I - consists of rejecting Ho, and Ho is true; and
- Error type II - consists of accepting Ho, and Ho is false.

The occurrence probability of error type I is denominated, in the hypothesis test, as level of significance (α). A level of significance of 10% means that there exists a probability that 10% of the data has $\beta = 0$ and they be considered as $\beta \neq 0$.

The F test is used to validate (or not) the hypothesis. The F index calculation is:

$$\mathbf{F} = \frac{\mathbf{MSr}}{\mathbf{MSd}} \qquad \text{eq. 1}$$

Where:

MSr – Sums of squares mean regression

MSd – Sum of squares mean deviation

The test rejects the hypothesis (Ho) for the whole F value equal to or greater than the value contained within an F table dependent on the degree of freedom and level of significance of the numerator and denominator (Vieira, 1988). The characteristics of the F test for a simple linear regression are shown in table 1.

The degree of freedom of the denominator is expressed by (N-2), where N is a band number. However, hyperspectral images have a redundancy of information. With the use of APC or MNF, an absorption feature can be represented by three components. Therefore, in the present method, the N value, instead of being expressed by the number of bands, is represented by the number of APC or MNF components with information. For more spectral features that occupy major numbers of bands, as the case of the double feature of kaolinite, four components can be considered. Thus, the conditions of the proposed method are the degrees of freedom of the denominator (1 or 2), numerator (1) and the level of significance (2.5%, 5% and 10%).

Table 1 – ANOVA for Simple Linear Regression (Davis, 1973)

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F Test
Linear Regression	SSr	1	MSr	MSr/MSd
Deviation	SSd	n –2	MSd	
Total Variation	SSt	n –1		

3 Formulation of the Spectral Identification Method (SIM) for the Integration of ANOVA with the Coefficient (SCM)

The variance analysis of a linear regression is not sensitive to the presence of a negative correlation. Therefore, as complement to the proposed method, the correlation coefficient SCM should be used (Carvalho & Meneses, 2000). The areas with negative correlation are detected and integrated in the data that were obtained with ANOVA (**Figure 1**).

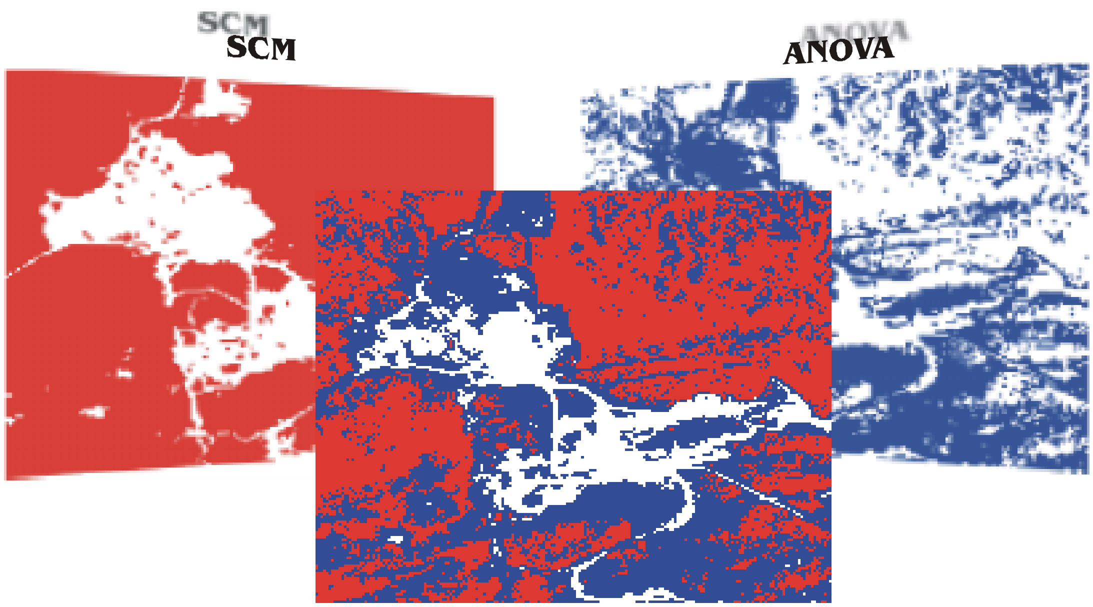


Figure 1. SIM method from SCM and ANOVA union.

4 Degrees of Significance Images

The present method generates binary images, relative to the existence (or not) of the information, inside a level of significance of 2.5%, 5% or 10%. This method was used for the features of pimelite/saponite (2.19µm - 2.36µm) and of vegetation (0.54µm - 0.75µm). These features used the value of degrees of freedom in the denominator equal to one.

Compared to the manual analysis, it is observed that the estimate of 2.5% was quite efficient to identify the areas for both features (Figure 2 and Figure 3). The other significance levels present an overestimated area. That characteristic is due to the high spectral correlation of the scene.

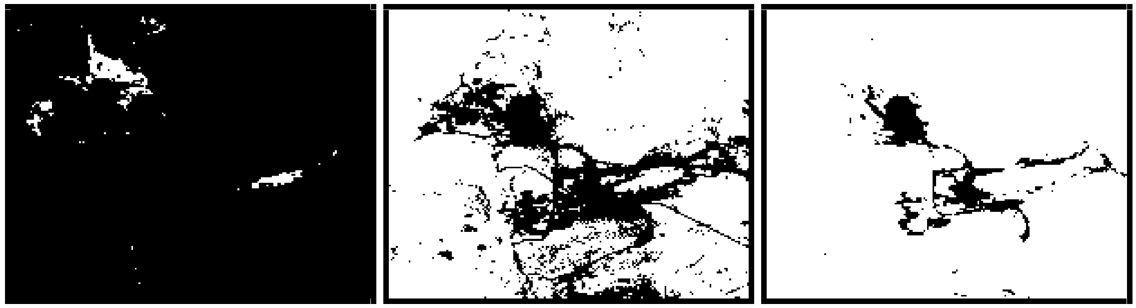


Figure 2. Comparison of the areas defined for pimelite presence, using the SIM method for the significance levels 2.5%, 5% and 10%.

Figure 3. Comparison of the areas defined for kaolinite presence, using the SIM method for the significance levels 2.5%, 5% and 10%.

5 Analysis of the Factor of Similarity SIM Image

The coefficient conjugate of the ANOVA and SCM methods varies between 0 and 1. The maximum value of that coefficient corresponds to the areas with levels equal to 2.5%, while the lower values are related to areas with SCM below 40% and F factor below 5%. The SIM images promote an intense expansion of the areas of greater SCM correlation, giving them an efficient and robust similarity index. Figure 4 presents the SIM images for the features of the pimelite-saponite and vegetation.

The SIM and SCM indexes are described by a function. Figure 5 presents the SCM x SIM scatter plot for the pimelite-saponite and vegetation features. The SIM promotes an expansion for SCM values with high correlation.

The generated curves are segmented in agreement with the differing degree of significance. The red area is relative to the significance of 2.5%, cyan 5%, blue 10% and yellow higher than 10%. The areas with pimelite and green vegetation are contained in the SIM value equal to one. Thus, this index establishes a great numeric difference between the intended data and the others. The author, in IDL language, inside of the ENVI program implemented the SIM.

6 Conclusion

SIM is a new method of spectral classification that provides estimates according to levels of significance of the most probable areas of the sought material, and an image related to the similarity parameter. The method demonstrated excellent estimates for study area.

The method has a similarity parameter that completely eliminates the false positives present in the algorithms of SAM, SFF and Tricorder and still maximizes the values of the areas where the sought material exists.

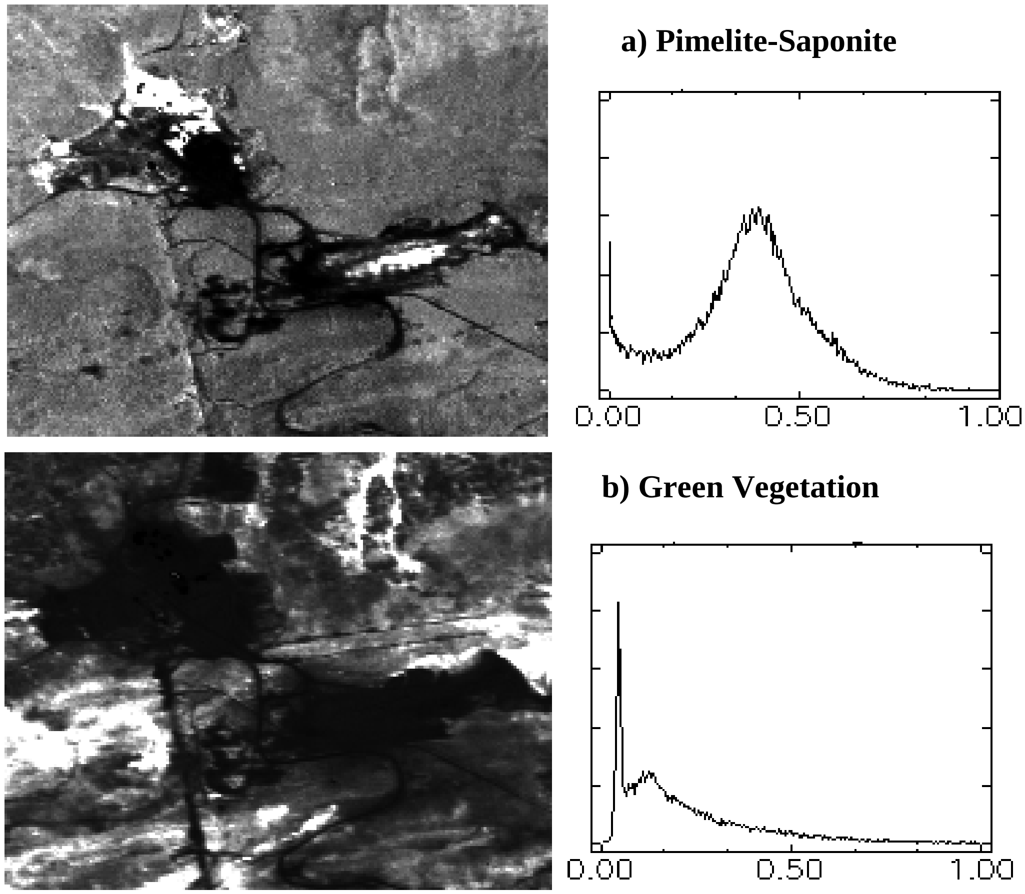


Figure 4. –SIM image for the area of the *Niquelândia* mine with its respective histogram of the features: a) pimelite, and b) vegetation.

References

Carvalho Jr., O. A. & Meneses, P. R., 2000, Spectral Correlation Mapper (SCM): An Improving Spectral Angle Mapper. In: *Ninth JPL Airborne Earth Science Workshop.* JPL Publication 00-18, pp. 65-74.

Clark, R. N. & Swayze, G. A., 1995. Mapping Minerals, Amorphous Materials, Environmental Materials, Vegetation, Water, Ice and Snow, and Other Materials: The USGS Tricorder 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. 550 pp.

Kruse, F. A.; Lefkoff, A. B.; Boardman, J. W.; Heiedbrecht, 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 Publication 92-14, pp. 23-25.

Kruse, F. A.; Lefkoff, B. & Dietz, J. B., 1993a, 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.

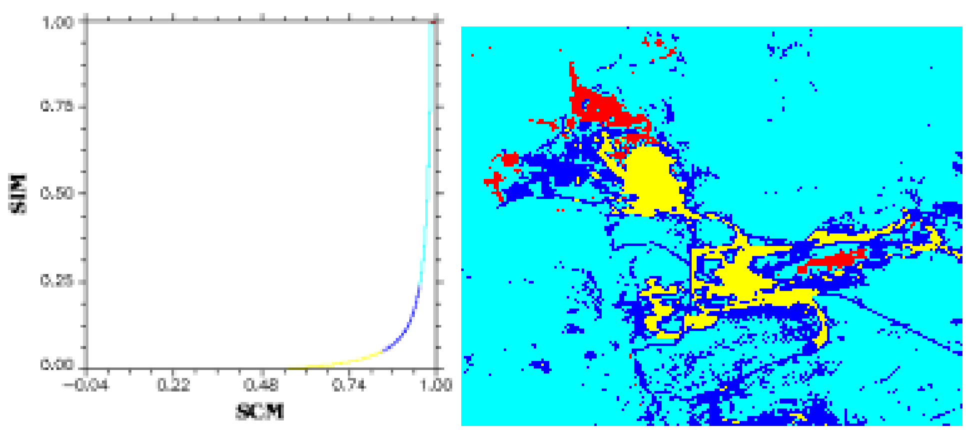
Kruse, FA; Lefkoff, A. B; Boardman, J. W.; Heiedbrecht, K. B.; Shapiro, A. T.; Barloon, P. J. & Goetz, A. F. H., 1993b, The Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data. *Remote Sens. Environ.* 44:145-163.

Souza, G. S., 1998, Introdução aos Modelos de Regressão Linear e Não-Linear, Brasília: EMBRAPA – SPI/EMBRAPA-SEA, 505 pp.

Steel, R. G. D.; Torrie, J. K., 1980, Principles and procedures of statistics. 2nd ed. New York: MacGraw-Hill.

Vieira, S., 1988, Indrodução à Bioestatística. 5th ed. Rio de Janeiro: Campus, 293 pp.

a) Pimelite-Saponite



b) Green Vegetation

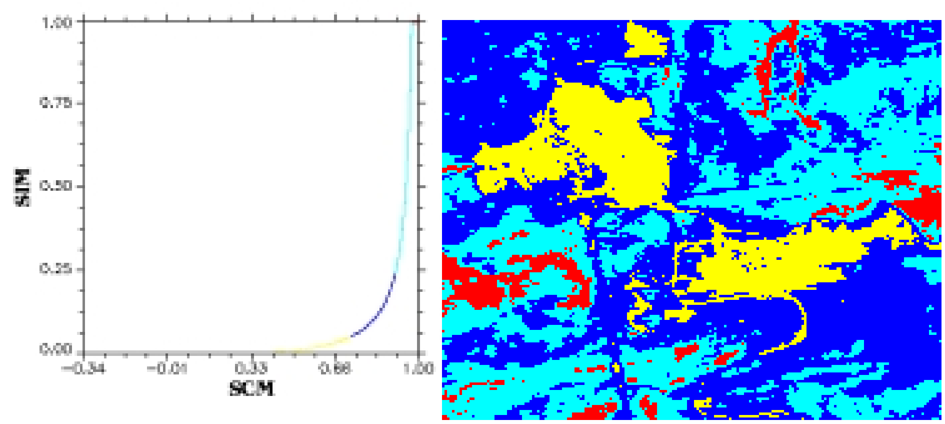


Figure 5 –Scatter plot among SCM x SIM for the features: a) pimelite and b) green vegetation. The curves are segmented in agreement with the degrees of significance: 2.5% (red), 5% (cyan) and 10%(blue). The areas in yellow have a relative value higher than 10%.