An Approach to Hyperspectal Image Classification Based on Linear Regression

Osmar Abílio de Carvalho Jr,[1, 2] Ana Paula Ferreira de Carvalho,[3] Paulo Roberto Meneses[4]

[1]Departamenteo de Geografia da Universidade de Brasília - Campus Universitário Darcy Ribeiro, Asa Norte, 70910-900, Brasília, DF, Brasil osmana@tba.com.br

[2]COTER – Exército Brasileiro - SMU QGEx Bl H térreo

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

[4]IG/UnB-Instituto de Geociências da Universidade de Brasília - Campus Universitário Darcy Ribeiro, Asa Norte, 70910-900, Brasília, DF, Brasil

1 INTRODUCTION

Many of the methods used in hyperspectral data classification are based on the linear regression between the Reference Endmember (RE) and the Image Spectrum (IS) as is the case in the Tricorder and Spectral Feature Fitting (SFF) methods. This linear regression can be done in two ways:

a) RE to IS regression (RE⇒IS)

$$\mathbf{RE = a + bIS} \qquad \text{Equation 1}$$

b) IS to RE regression (IS ⇒RE)

$$\mathbf{IS = a + bRE} \qquad \text{Equation 2}$$

The degree of adjustment between those spectra is obtained using the correlation coefficient (R) or by using the total Root-Mean-Square error (RMS). Both ways of regression present the same values of correlation coefficient and different values of RMS.

2 RE⇒ IS REGRESSION CHARACTERISTICS

The combined use of RMS and R is widely practiced in statistical analysis and presents some advantages in spectral analysis. Images of RMS and R are obtained using RE ⇒ IS regression, where the dependent variable is always a constant datum, described by a function.

For example, the correlation coefficient and RMS were calculated between a reference spectrum of kaolinite and a hyperspectral image from the AVIRIS sensor for the Niquelandia area, in Brazil. The Spectral Correlation Mapper (SCM) index based on Pearson' s correlation was also used (Carvalho *et al.*, 2000).

Scatterplots of bands $RMS_{\mathbf{RE \Rightarrow IS}}$ x R and $RMS_{\mathbf{RE \Rightarrow IS}}$ x SCM describe a parabolic function. The parabola's vertex presents a null value to the correlation coefficient and the highest value of $RMS_{\mathbf{RE \Rightarrow IS}}$ (Fig. 1). A spectra series along the generated curve allows for the accompaniment of sign degradation according to $RMS_{\mathbf{RE \Rightarrow IS}}$ and SCM values (Fig. 2). The absorption feature of kaolinite is more evident at point 1, which presents the highest value of correlation coefficient and a minor $RMS_{\mathbf{RE \Rightarrow IS}}$ value. Sign degradation increases towards the parabola's vertex (point 5). The points 6 and 7 represent the spectra of negative correlation.

The definition of the main points along the curve allows for image classification. Figure 3 shows the image split into three areas: a) areas of negative correlation (white), b) areas with low correlation and high values of $RMS_{RE \Rightarrow IS}$ (gray) and c) areas with correlation above 80% and low values of $RMS_{RE \Rightarrow IS}$ (black).

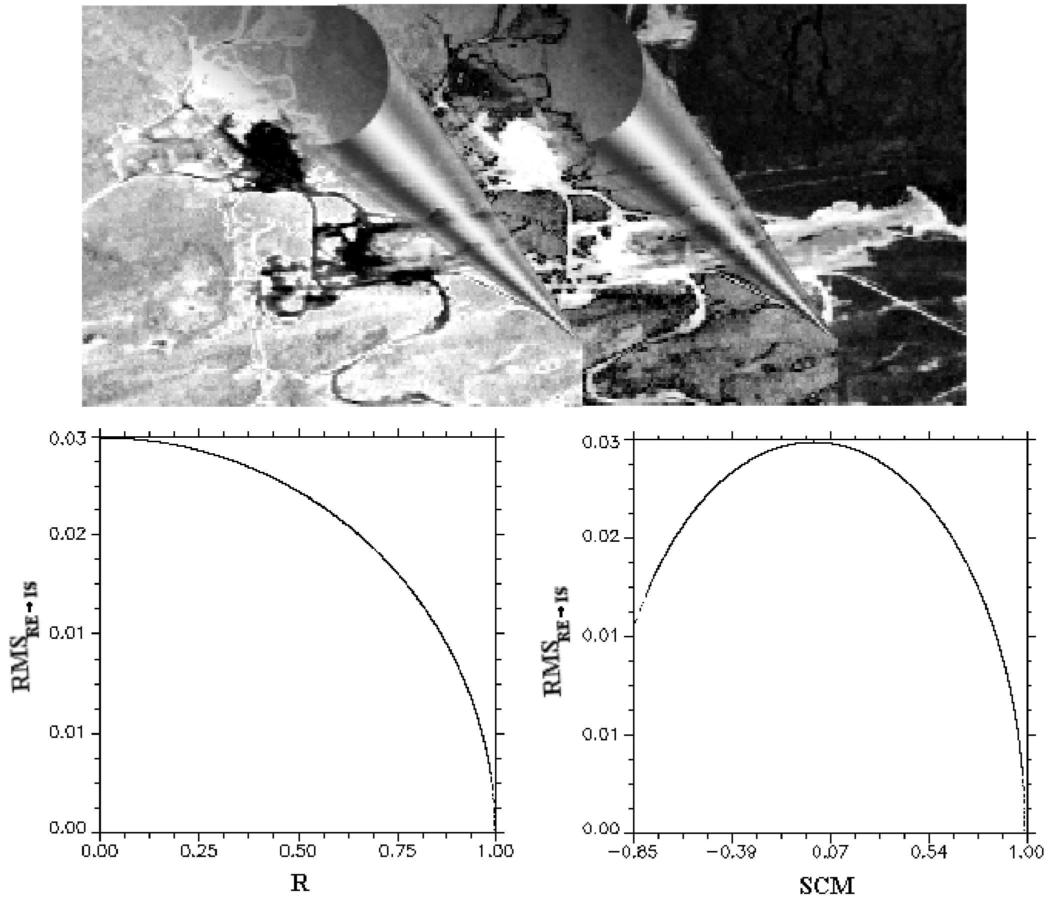


Figure 1 – Scatterplots a) $RMS_{RE \Rightarrow IS}$ x R and b) $RMS_{RE \Rightarrow IS}$ x SCM for the kaolinite feature.

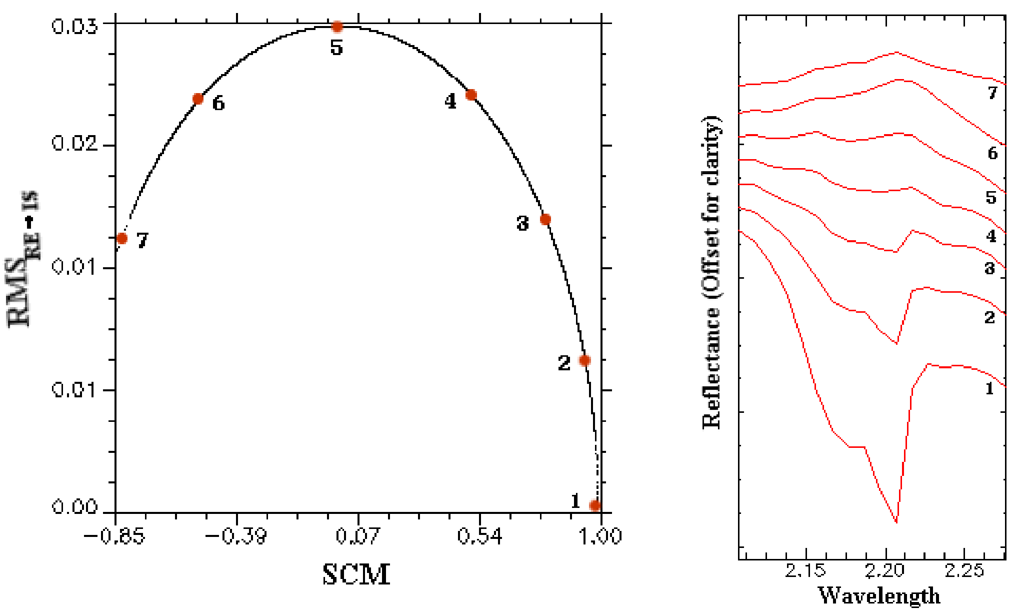


Figure 2 – Behavior of kaolinite's features in Scatterplot of images $RMS_{RE \Rightarrow IS}$ x SCM.

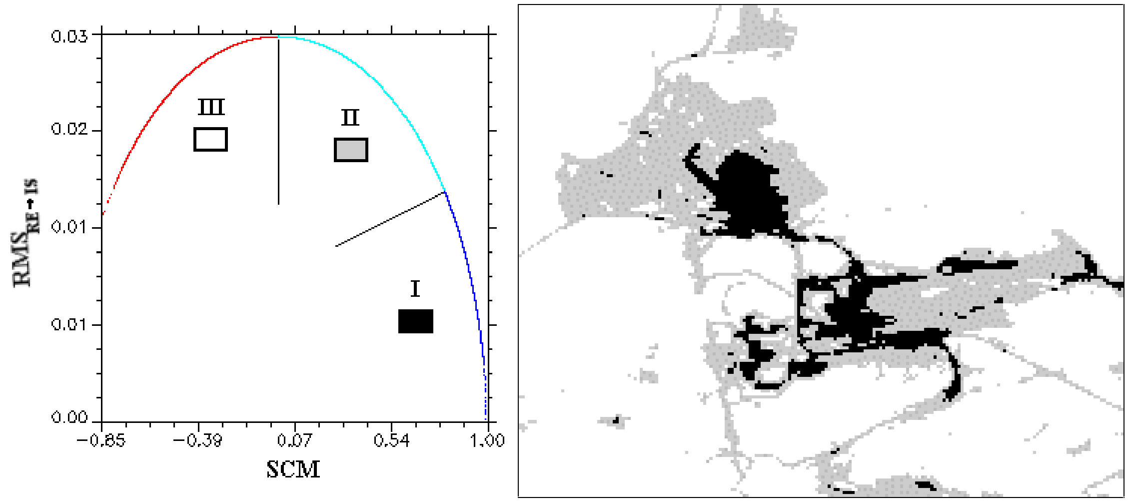
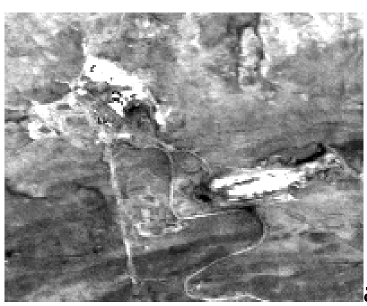


Figure 3 – Classification image through the scatterplot of SCM x RMS$_{\mathbf{RE} \Rightarrow \mathbf{IS}}$.

3 IS ⇒ RE REGRESSION CHARACTERISTICS

The IS ⇒ RE regression does not establish a function between RMS$_{\mathbf{IS \Rightarrow RE}}$ and the R images. Those images show an intense visual difference. The RMS$_{\mathbf{IS \Rightarrow RE}}$ image for the kaolinite feature has a histogram where the highest frequency of data is located in the lowest values; the opposite of the RMS$_{\mathbf{RE \Rightarrow IS}}$ image where the peak of frequency is located with highest values (Fig. 4).


a) RMS $_{\mathbf{IS \Rightarrow RE}}$

b) RMS$_{\mathbf{RE \Rightarrow IS}}$.

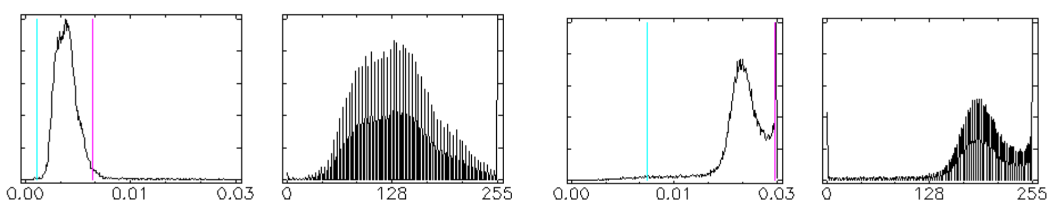


Figure 4 – Images of RMS $_{\mathbf{IS \Rightarrow RE}}$ and RMS$_{\mathbf{RE \Rightarrow IS}}$ for the kaolinite feature.

The RMS $_{\mathbf{RE \Rightarrow IS}}$ enables kaolinite's areas to be better enhanced than the RMS$_{\mathbf{IS \Rightarrow RE}}$, found by spectra extraction. The scatterplot presents differentiated behavior between the two RMS types (Fig. 5). Notice the reverse of RMS $_{\mathbf{IS \Rightarrow RE}}$ in relation to RMS$_{\mathbf{RE \Rightarrow IS}}$ values.

The scatterplot of images RMS $_{IS \Rightarrow RE}$ and SCM presents a strong variation of SCM for the same values of RMS $_{IS \Rightarrow RE}$ (Fig. 6). In the spectral curves of Fig. 6, the same results of RMS $_{IS \Rightarrow RE}$ present different spectral curves according to the SCM values.

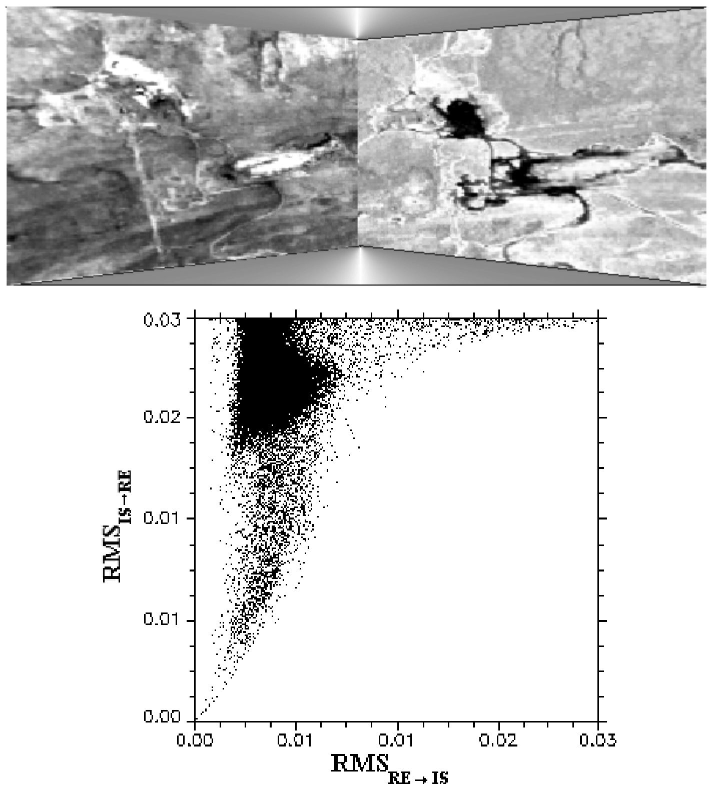


Figure 5 – Scatterplots between the data of RMS $_{IS \Rightarrow RE}$ and RMS$_{RE \Rightarrow IS}$ for the kaolinite feature.

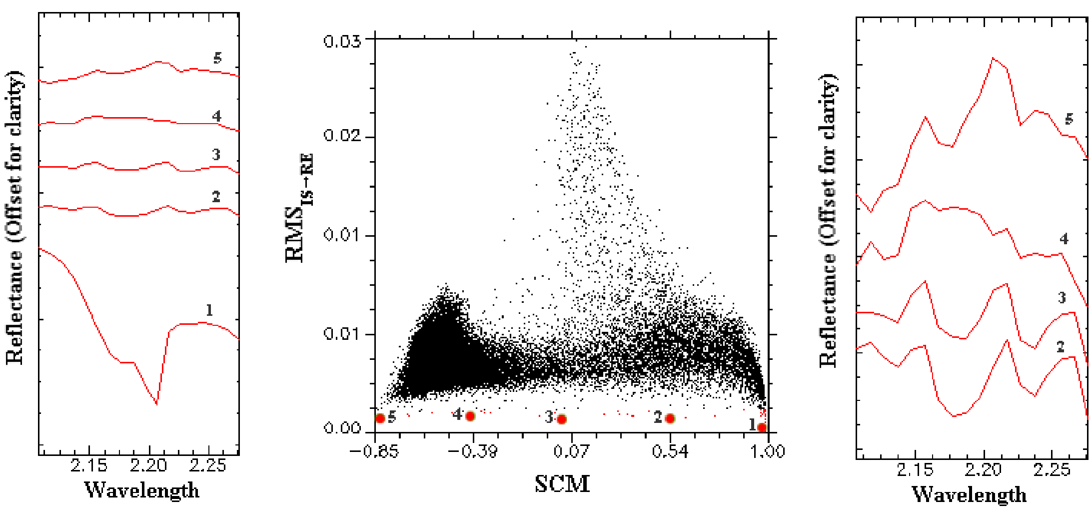


Figure 6 – Behavior analysis of kaolinite features through scatterplot of images of RMS$_{RE \Rightarrow IS}$ x SCM and a spectra's series. The left side series are without the spectrum of point 1, allowing for the distinction of other spectra.

4 RE⇒IS REGRESSION CHARACTERISTICS WITH CONTINUUM-REMOVED DATA

The continuum-removed data process is an important procedure to spectral feature detection. The RE⇒ IS and IS ⇒ RE regressions used on the continuum-removed spectrum also present equal correlation coefficient values as well as different RMS values.

The Tricorder program calculates the correlation coefficient-**R** of a continuum-removed spectrum. The scatterplot of $RMS_{\mathbf{RE \Rightarrow IS}}$ and **R** describes a parabolic function, as it was described to spectra without the continuum removed. (Figs. 7 and 8).

The utilization of SCM, instead of R, allows the detection of areas with negative correlation (Fig. 9). However, it occurs a decrease of negative correlation values with the use of continuum-removed data.

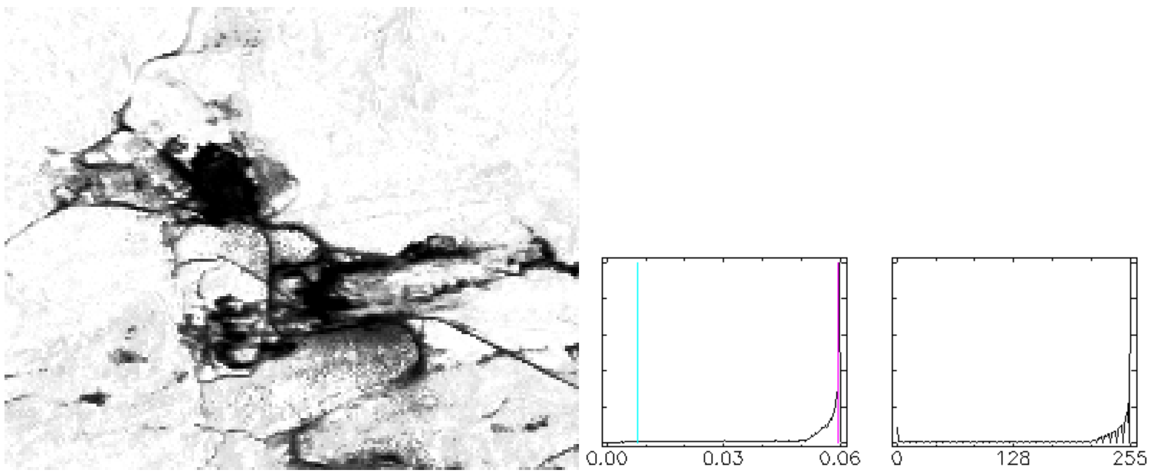


Figure 7 – Images of $RMS_{\mathbf{RE \Rightarrow IS}}$ with its respective histogram.

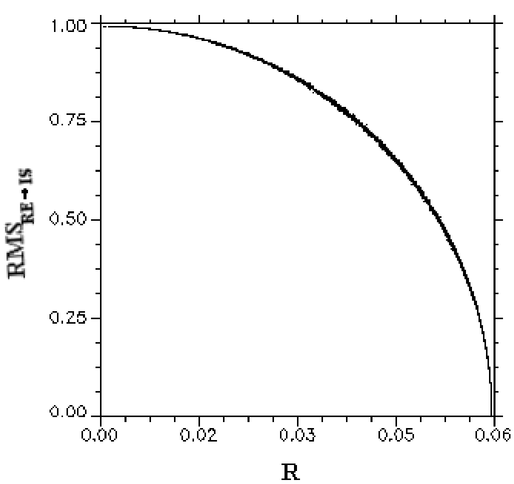


Figure 8. Scatterplot of images $RMS_{\mathbf{RE \Rightarrow IS}}$ and R.

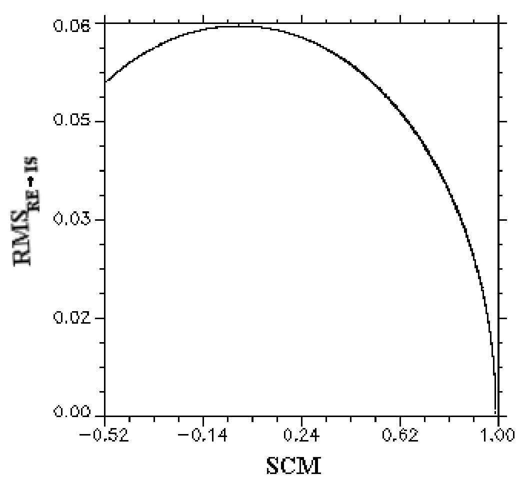


Figure 9. Scatterplot of images $RMS_{\mathbf{RE \Rightarrow IS}}$ and SCM.

Analogous to spectra analysis without the removal of the continuum, the scatterplot of SCM and $RMS_{RE \Rightarrow IS}$ presents a curve that permits the following of signal degradation. This procedure enables the determination of the point where the interested feature is no longer observed. Figure 10 presents a spectra sequence along the parabolic curve, where the kaolinite feature identification decreases since 1 to 7. The signal begins to lose its features after point 3. In spite of enhancements of the feature of interest, the continuum removal can either enhance noise features from mixtures or interferences. To facilitate the analysis, another spectra sequence can be used without the continuum for the same points selected in the curve.

Demarcation of principal modification locations of spectrum behavior within the curve from established points permits the classification of the image. Figure 11 presents a classification of the kaolinite image according to three classes. The area with the mineral appears dark blue and contains high values of SCM and low values of $RMS_{RE \Rightarrow IS}$.

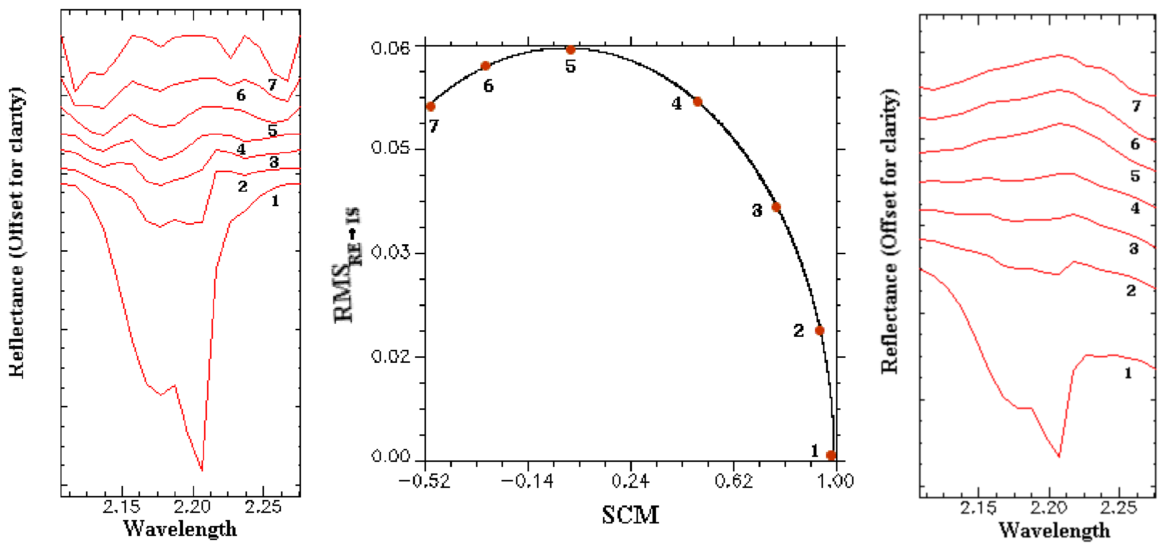


Figure 10 – Scatterplot of images $RMS_{RE \Rightarrow IS}$ and **SCM** with removal of continuum on the left side and without removal of continuum on the right side.

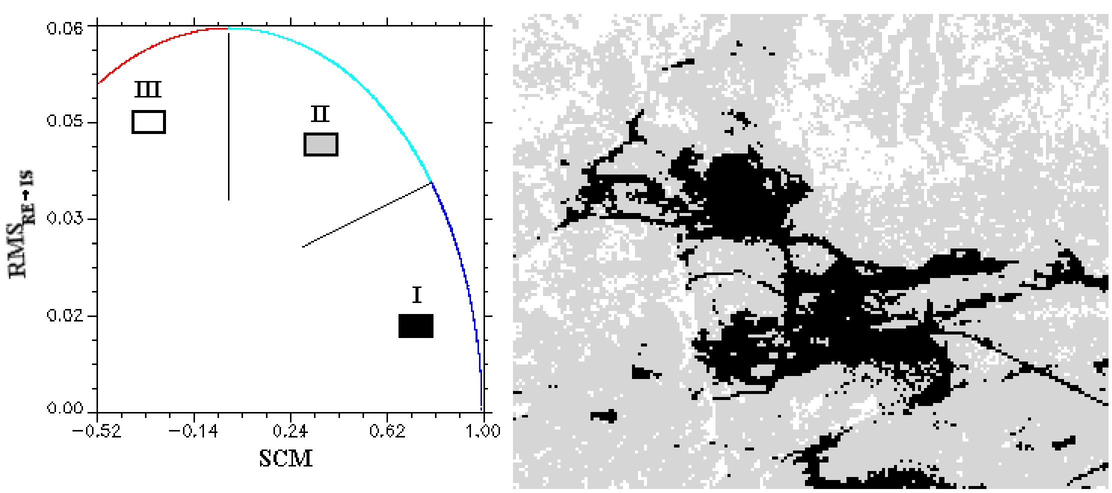


Figure 11 – Segmentation of the image from a function analysis established between $RMS_{RE \Rightarrow IS}$ and **SCM.**

5 IS ⇒ RE REGRESSION CHARACTERISTICS WITH CONTINUUM REMOVAL

The scatterplot of $RMS_{IS \Rightarrow RE}$ and R exhibits differentiated behavior between the two parameters (fig.12). As seen in the image without continuum removal, the kaolinite's feature is more coherent with R than $RMS_{IS \Rightarrow RE}$. The example in Fig. 13 serves to illustrate this point. According to the $RMS_{\mathbf{IS} \Rightarrow \mathbf{RE}}$ analysis; point 5 is considered equal to point 1, which are actually different considering their spectral curves. This reveals the $RMS_{\mathbf{IS} \Rightarrow \mathbf{RE}}$ limitation.

The Tricorder program uses the coefficient of correlation (R^2) while SFF uses $RMS_{IS \Rightarrow RE}$. The scatterplot of these two show an inverse visual similarity while their histograms show the distribution of distinct data (Fig. 14). Hence, the SFF, in spite of the use of a similar Tricorder formulation, presents different and inferior results.

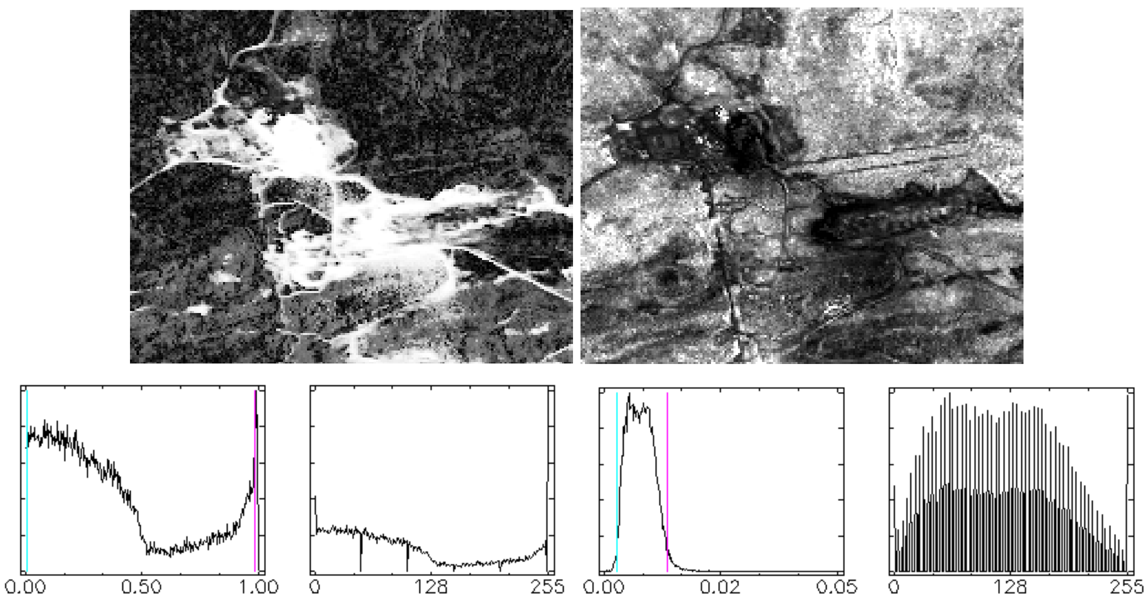


Figure 12 – Image of correlation - **R** and $RMS_{IS \Rightarrow RE}$ for the kaolinite feature with removal of continuum and their respective histograms.

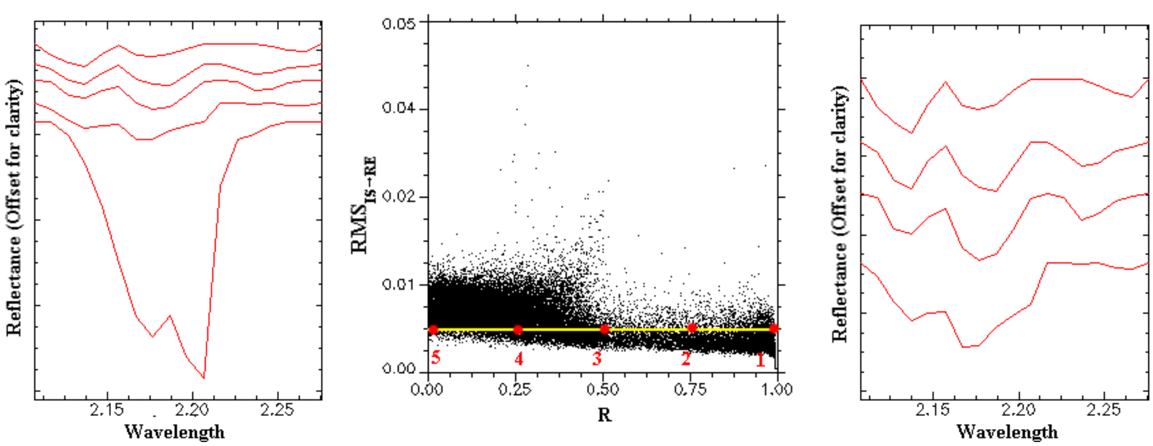


Figure 13 – a) Scatterplot of images R and $RMS_{\mathbf{IS} \Rightarrow \mathbf{RE}}$ for the kaolinite feature and a spectrum series. The series on the left are without the spectrum of point 1, allowing the distinction of other spectra.

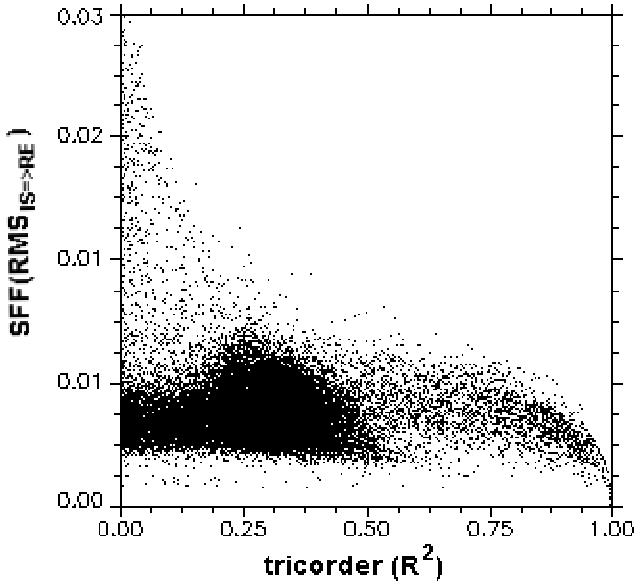


Figure 14 – Scatterplot of SFF x Tricorder images.

6 CONCLUSION

The combined use of $RMS_{\mathbf{RE \Rightarrow IS}}$ and R or SCM allows for an accurate spectral analysis. The employment of the SCM instead of the R has the advantage of the detection of negative correlation. When the continuum-removed data is used, negative correlation decreases and results became closer to SCM and R.

The employment of $RMS_{\mathbf{IS \Rightarrow RE}}$ presents a different and inferior performance than with $RMS_{\mathbf{RE \Rightarrow IS}}$. Due to the limitations of $RMS_{\mathbf{IS \Rightarrow RE}}$; the SFF program presents different results from those obtained by Tricorder.

In order to compare different methods, an applicative onto the ENVI program was implemented, which permits the R, SCM, $RMS_{\mathbf{IS \Rightarrow RE}}$ and $RMS_{\mathbf{RE \Rightarrow IS}}$ calculations to be performed.

Acknowledgements

The Brazilian Army has supported this work. We also thank Dr. Mírian Trindade Garret for her review and suggestions.

References

Carvalho Jr, O. A.; Carvalho, A. P. F.; Mensese, P. R., 2000, Spectral Correlation Mapper (SCM): An Improvement on the Spectral Angle Mapper (SAM). Summaries of the Ninth Annual JPL Airborne Earth Science Workshop, JPL Pub 00-18, Jet Propulsion Laboratory, Pasadena, CA.

Clark, R. N; Gallagher, A. J.; Swayze, G. A., 1990. *Material Absorption Band Depth Mapping of Imaging Spectrometer Data Using a Complete Band Shape Least-Squares Fit With Library Reference Spectra* Summaries of the 2nd Annual JPL Airborne Geoscience Workshop, JPL Pub 90-54, AVIRIS Workshop. Jet Propulsion Laboratory, Pasadena, CA, pp. 176-186.

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* Summaries of the Fifth JPL Airborne Earth Science Workshop, JPL Publication 95-1, v.1, pp. 39-40.

ENVI ®, 1997, Tutorials Better Solutions Consulting Limited Liability Company Lafayette, Colorado, USA, 370 pp.