CLUEGO, AN INFORMATIONAL HYPERSPECTRAL CLASSIFIER. L. Pompilio[1], G. Pedrazzi[2], M. Pepe[3], L. Marinangeli[1], [1]DISPUTer, D'Annunzio University, Via Dei Vestini, 31, Chieti I-66013, pompilio@irsps.unich.it, [2]Department of Neuroscience, University of Parma, Via Volturno, 39, Parma I-43125, [3]CNR-IREA, Via Bassini, 15, Milano I-20133.

Introduction: Hyperspectral scanners are widely used for applications based on remote measurements on Earth and planetary surfaces, as well. The high spectral resolution of hyperspectral sensors allows for the application of very sophisticated processing techniques, in order to retrieve exhaustive information from highly detailed spectra.

Nevertheless, the inherent dimensionality of hyperspectral imagery is responsible for important limitations in the application of both supervised and unsupervised classifiers already developed for multispectral systems [1, 2, and references therein]. As a consequence, the application of feature selection and reduction techniques prior to data processing is recognized as a fundamental step for improving the classification results.

Despite the numerous approaches to feature reduction, so far developed [3, 4, 5, 6], analytical methods capable of preserving as much information as possible are still required. The present research is focused on the development and testing of CLUEGO, an hyperspectral classifier aimed at preserving the informational content of hypercubes.

The main goals of the present research are: a) to provide an alternative method for reducing the inherent dimensionality of hypercubes and simultaneously preserving the informational content of data; b) to use the reduced dataset in order to develop a classification method based on spectral analysis; c) to demonstrate the benefits of the present method versus the cluster analysis of the original dataset. In order to achieve these purposes, we used a number of simulated hypercubes and evaluated CLUEGO performances versus one of the most common clustering algorithms (e.g., the k-means) applied to both the original and reduced data cubes. The results have been statistically evaluated.

Theory: The key features of the CLUEGO hyperspectral classifier are: 1) reducing the dimensionality of the dataset; and 2) clustering the results.

Feature reduction. In order to preserve the informational content inherent in the dataset, we accomplish the feature reduction processing through parametrization of the hypercube in the spectral direction. We apply the EGO (Exponential Gaussian Optimization) decomposition [7, 8] to model the spectral response of each pixel. As a result, we transform the original dataset in a reduced space including 7 dimensions per spectral feature, those dimensions being the EGO center position, width, depth, asymmetry and saturation; the continuum slope and intercept. We run the code over array structures, thus to accomplish the spectral modeling of both simulated and real hypercubes on a pixel by pixel basis. The code returns the reduced cube, residual and error matrices, the latter including the standard error of the estimate and the adjusted coefficient of determination.

Cluster analysis. By applying the CLUES algorithm [9] to the reduced dataset, we automatically estimate the optimal number of clusters and the partition of data points with the only input of a convergence criterion. The subsequent cluster analysis by using the PAM or k-means methods with the number of clusters provided by the CLUES algorithm, allow us to retrieve the classification raster image.

Results: In order to accomplish the performances of the CLUEGO classifier, we tested the code on a number of synthetic hypercubes (fig. 1) and compared the results with the application of the much more common k-means algorithm to the original hypercubes, in terms of correctness of class identification and computational time. The test application is addressed to model spectral features occurring near 1.0 μm, which are diagnostic of Fe-bearing minerals. We used known laboratory spectra of minerals (fig. 2) and a small array of 60x40 pixels. At each simulation, we decreased the number of channels in the same spectral range and the SNR.

CLUEGO classifier demonstrates to perform very well at each specific test. It was always able to identify the 5 classes and match their distribution within the test hypercubes. It appears not to be influenced by Gaussian noise nor by broader channels, as expected. The k-means algorithm performed well at each test, but with higher computational time.

Conclusive remarks: The k-means algorithm is a well established technique of image segmentation, where the role of the user is limited to the choice of the number of classes to be retrieved. Therefore, though the correct number of classes, the good performance of the k-means classifier can be assessed as a matter of computational time. CLUEGO classifier always performed well even when no information on the number of classes to identify was provided, and the time required for computation was always lower. In addition, the feature reduction step included in the CLUEGO algorithm preserves all the informational content of spectra because the classification involves the model parameters. This properties also allows for a semi-complete reversibility to the modeled feature.

After having tested the algorithm against a number of well established techniques of image segmentation, we will dedicate a separate task to applications to real hyperspectral data, acquired on Earth and other planets, as well. The ultimate goal is to directly link the classification method to the informational content included in the spectral data.

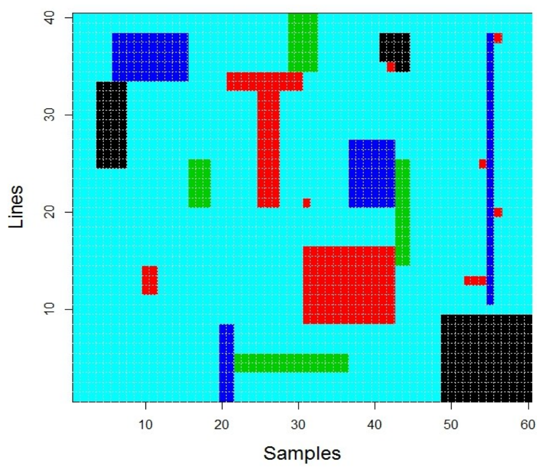

Figure 1. Single layer of the test hypercube used for the computation. Basically, the image is a 60x40 pixel matrix; the spectral dimension is variable, according to the different levels of data degradation, starting from 100 spectral channels as full resolution. Colored areas reflect the spatial distribution of the spectra used for populating the array. For the legend, see Fig. 2

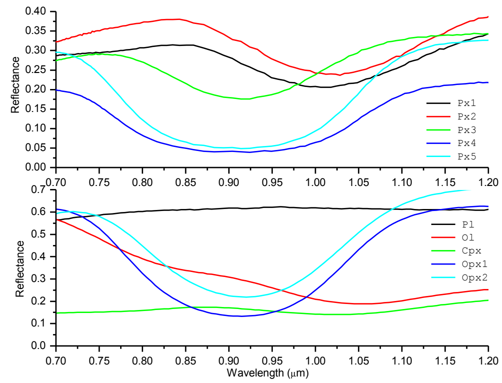

Figure 2. Plot of the reflectance spectra forming the dataset. Top plot: one of the simulations included 5 spectra of piroxenes, distributed as in Fig. 1. Bottom plot: the other simulations included spectra of 5 different minerals (Pl, plagioclase; Ol, olivine; Cpx, clinopyroxene; Opx1 and Opx2, orthopyroxenes), distributed as in Fig. 1. Then we progressively degraded the dataset starting from this latter hypercube.

References: [1] Richards J. A. and Jia X. (2006) *Remote Sensing Digital Image Analysis. An Introduction,* Springer-Verlag, Berlin. [2] Plaza A. et al. (2009) *Remote Sensing of Environment, 113*, s110-s122. [3] Landgrebe D. A. (2003) *Signal theory methods in multispectral remote sensing*, J. Wiley & Sons. [4] Van Der Meer F. (2004) *International Journal of Applied Earth Observation and Geoinformation, 5*, 55-68. [5] Hsu P. (2007) *ISPRS Journal of Photogrammetry and remote Sensing, 62*, 78-92. [6] Pelkey S. M. et al. (2007) *JGR, 112*, doi:10.1029/2006JE002831. [7] Pompilio L. et al. (2009) *Icarus, 201*, 781-794. [8] Pompilio L. et al. (2010) *Icarus, 208*, 811-823. [9] Wang X. et al. (2007) *Computational Statistics and Data Analysis, 52*, 286-298.