Geophysical Research Abstracts,
Vol. 10, EGU2008-A-04899, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-04899
EGU General Assembly 2008

Clustering of laboratory spectral data: evaluating applications to interpretation of planetary spectra

G. A. Marzo (1), T. L. Roush (1), R. C. Hogan (2)

(1) NASA Ames Research Center, MS 245-3, Moffett Field, CA 94035-1000, (2) Bay Area Environmental Research Institute c/o NASA Ames Research Center MS 245-3, Moffett Field, CA 94035-1000 (giuseppe.marzo@le.infn.it / Fax: +1 650-604-6779)

Planetary space experiments are currently collecting a huge amount of information which needs to be explored in order to understand its scientific content. Marzo et al. [1] developed and evaluated an unsupervised statistical clustering scheme able to reduce a spectral data set to a few clusters allowing for more focused, and rapid, evaluation of their scientific meaning. The technique adopts a partitioning clustering algorithm based upon the Calinski and Harabasz [2] criterion for identifying the natural number of clusters. Here we extend the original criterion to account for the measurement uncertainty. We also investigate limitations to the technique and, eventually, associate a scientific meaning to each of the clusters.

We apply the clustering technique to ASTER and RELAB reflectance spectral libraries of minerals, focusing our analyses from visible to near-infrared and on three different particle size ranges. Spectral libraries are documented data sets providing the ability to assign each spectrum a label reflecting its physical and chemical properties. We assess the ability of the original and extended criteria to identify natural clusters of the library spectra and estimate associated uncertainties of the results. Several independent clustering analyses are performed using each criterion. The number of natural clusters obtained for each analysis is highly reproducible. We evaluate the scientific meaning of the derived clusters, based upon the associated compositional labels contained within each cluster, following the approach described by Roush and Hogan [3].

Once the clusters meanings are defined, we test our classification scheme. The test uses the approach of [3]: initially define clusters using a randomly selected subset of the spectral library; define labels for each cluster; assign new spectra not included in the initial training to a cluster; and evaluate the accuracy of assigning the unknown spectra to the correct cluster.

Future efforts will focus on applying this approach to MRO/CRISM spectroscopic data, but the technique is appropriate for a wide range of applications including analysis of different terrestrial and planetary data sets and enabling pattern recognition capabilities on next generation planetary rovers.

[1] Marzo,G.A.; Roush,T.L.; Blanco,A.; Fonti,S.; Orofino,V., 2006, Cluster analysis of planetary remote sensing spectral data, JGR, 111, E3, E03002

[2] Calinski, T., Harabasz, J., 1974, A dendrite method for cluster analysis. Communications in Statistics - Theory and Methods, A, 3, 1-27

[3] Roush, T.L., Hogan, R., 2006, Automated Classification of Visible and Near-Infrared Spectra Using Self-Organizing Maps, Proceedings IEEE 2004 Aerospace Conference, Big Sky, Montana, March 2007, paper #1456.