AUTONOMOUS MINERAL DETECTORS FOR VISIBLE/NEAR-INFRARED SPECTROMETERS AT MARS. Martha S. Gilmore[1], Rebecca Castaño[2], Benjamin Bornstein[2] and James Greenwood[1], [1]Dept. of Earth and Environmental Sciences, Wesleyan University, Middletown, CT 06459 mgilmore@wesleyan.edu, [2]Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109 {firstname.lastname}@jpl.nasa.gov.

Introduction: Spectroscopy is the primary tool by which we can remotely assess the chemical composition and mineralogy of a planetary surface. Currently, the most advanced available technology for the observation of planetary surfaces is imaging spectrometers, which combine high spatial resolution (10s m) imaging with hundreds of spectral bands. Hyperspectral systems collect huge volumes of multidimensional data. Analysis of these data is time consuming, costly and often requires an expert to robustly identify target compositions by comparison to spectral libraries. For planetary missions, where global datasets are collected, analysis costs restrict rapid classification to only a subset of an entire mission dataset, reducing mission science return. Data downlink restrictions from planetary missions also highlight the need for robust mineral detection algorithms. For example, both the Observatoire pour la Minéralogie, l'Eau, les Glaces, et l'Activité (OMEGA) aboard Mars Express and the Compact Reconnaissance Imaging Spectrometer (CRISM) aboard MRO will map only ~5% of the Mars surface at full spatial and spectral resolution. While some targets are preselected for full resolution study (e.g., the MER landing sites), other high priority targets on Mars will be selected in response to observations made by the instruments in a multispectral survey mode. The challenge is to create mineral detection algorithms that can be utilized to analyze any and *all* image cubes for a selected system to help ensure that priority targets are not overlooked in these datasets.

The primary goal of this study is to design automated algorithms that will rapidly classify hyperspectral data and identify geologically important minerals. We focus on the identification of specific important mineral compositions within the data rather than trying to automate the reliable unmixing of the complex materials on a planet's surface. For Mars, high priority targets include minerals associated with the presence of water. Such algorithms can help ensure that priority targets are not overlooked in these datasets. These detectors will be tunable to a range of hyperspectral systems, providing a mechanism to search all data for targets of interest. We focus on the specific application of building detectors to identify minerals in visible/near-infrared VIS/NIR (350-2500nm) hyperspectral data collected at Mars by OMEGA (300 m to 2-5 km/pixel, 352-400 bands over 350-5100 nm) with eventual application to the CRISM (18 to 200 m/pixel, 500 or 60 bands over 400-4050 nm). The algorithms are also tested on data from JPL's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS; 17 m/pixel, 224 bands over 400-2500 nm [1]).

Classification Methodologies: Each mineral has a specific crystal structure that governs the absorption of energy at particular wavelengths. The position and depths of absorption features in a spectrum can be used to remotely identify the constituent minerals and determine their abundance by comparison with library spectra of known materials. While individual high-resolution spectra can be analyzed manually to characterize absorption bands and identify the minerals present, with hyperspectral image data, the large numbers of individual spectra make this individual analysis of all spectra cost-prohibitive.

Recognizing this problem, a number of approaches to the automated classification of image data have been developed. For multispectral (10s of bands) image data, a common approach is to use statistical methods to exploit the large number of samples available to separate mixed pixels from pure (end-member) pixels. The drawback of this approach is that it is scene-specific. Results vary based on the mineral mixtures present, observing conditions (e.g., atmospheric effects), and whether there are representative examples of pure pixels. As a consequence the results may not generalize well to other image cubes. Another approach to multispectral analysis has typically been to use ratios of bands to identify characteristic behaviors from which the minerals can be ascertained. Although ratioing effectively reduces the fidelity of hyperspectral data, band ratioing and band depth identification represents the state-of-the-practice for hyperspectral analysis of planetary data [2-6]. A number of other methods have been developed to facilitate the retrieval of mineralogical information from imaging spectrometer data including an expert system/decision tree [7], Spectral Angle Mapper [8,9], and binary encoding [10]. Several of these methods are regularly used in the interpretation of spectra from terrestrial imaging spectrometers such as AVIRIS. These methods can be reliable but require considerable domain expertise to develop and often also require considerable domain expertise to use. Our techniques utilize supervised classification methods that leverage recent advances in machine learning. Our approach is to look for specific minerals in a spectrum (as opposed to deconvolving a spectrum) reducing computational cost. The CRISM team has recognized the need for triage of hyperspec-

tral data and has defined a simple set of parameters to help classify spectra according to expected mineral types [2]. Our classification methods enhance this approach, enabling rapid development of new detectors rather than requiring extensive manual labor. In addition, the detectors trained by our system will have different sensitivities, complementing the currently available detection methods.

Detector Construction: Support Vector Machines (SVMs*)* belong to a category of machine learning techniques that produce empirically derived classification algorithms by explicitly attempting to maximize the margin, the boundary that separates one type of classified data from another. Maximizing the margin minimizes overfitting of the training data, which can lead to poor algorithm performance on new input data sets. Thus, maximizing the margin leads to better generalization of the classifier. Supervised classifiers such as SVMs require training data to learn the difference between positive and negative examples. In this work, we utilize different approaches to obtaining training data: 1) using a *generative model* to create a large set of training data from library spectra, and 2) using training data from known targets in the image itself.

The Generative Model. A drawback of SVM algorithms is that large numbers of training data are often required for the algorithm to converge (learn). This makes the cost of collecting a sufficient number of field or laboratory spectra for training prohibitive. We addressed this problem in our previous work by developing a generative model for spectra that allows us to obtain spectral data with many of the characteristics of field data at almost no cost. The model mixes spectra drawn from both JPL ASTER [11] and USGS [12] and CRISM spectral libraries based on mineral compositions (rocks) we specify. Compositions consist of one or more minerals where each mineral has a percent present range (to model modal mineralogy) and is identified to be *essential*, *non-essential*, or *accidental* (rare) to the composition. Many of the compositions were carefully constructed to represent known and predicted martian mineralogy and petrology as derived from the theoretical literature and study of martian meteorites. Finally, the model can optionally add Gaussian noise in order to simulate measurement noise present in spectra collected outside of a laboratory. With this method, thousands of example spectra can be created and used as examples to train the detector.

Two algorithms trained using the generative model have been designed to detect calcite ($CaCO_3$) and jarosite ($(K,Na,H_3O)Fe_3(SO_4)_2(OH)_6$) and performed well (~80-90%) on the VisNIR point spectra of rocks and minerals collected both in the laboratory and the field [13-16].

Within-image training. If the composition of pixels in a scene is known, a subset of pixels can be selected for training and another for testing within an image. This has the advantage of effectively normalizing the pixels for instrument error, atmospheric effects and for Mars, the effects of dust if comparable amounts of dust are expected on both training and unclassified pixels.

For AVIRIS data, we rely on an image of Cuprite, Nevada, which is well studied and serves as a training site for other classification algorithms [7,17]. For Mars, training data for this method can be derived from regions on Mars with known mineralogy. This currently includes hematite and sulfates identified at Opportunity site by TES/THEMIS, OMEGA [5] and MER, olivine identified at Nili Fossae [18], pyroxenes identified by ISM [19,20], TES/THEMIS [21] and OMEGA [6,19], and sulfates identified by the OMEGA in other regions of the planet [3,22].

Atmosphere and dust correction. Radiometrically and geometrically corrected OMEGA data are atmospherically corrected using the method outlined in [23] which assumes that the surface and atmospheric contributions to the spectrum are multiplicative, and that the atmospheric contribution follows a power law variation with altitude. In this way, an atmospheric spectrum correlates with altitude and the consequent depth of the CO_2 band at 2.75 microns can be derived from the spectra at the summit and the base of Olympus Mons [2,3,6].

The effect of the ubiquitous martian dust is to subdue primary absorptions in each spectrum, making the absorptions more difficult to recognize by humans and algorithms alike. We have addressed this in the OMEGA data by training the detectors using the spectra from the hyperspectral data themselves. Another solution, which we have not yet implemented, is to modify the detectors created using the generative model, by training them on library spectra that have been artificially degraded to simulate the effects of dust.

Results and Discussion: *Proof-of-Concept: Detection of calcite in AVIRIS data*. As a proof-of-concept, we built a classifier to detect the mineral calcite based on a SVM using the generative model. The detector was trained and tested on continuum-removed spectra resampled to match the wavelengths in the AVIRIS scene. This detector has been applied to the analysis of an atmospherically corrected AVIRIS scene of Cuprite, Nevada (Figure 1) to demonstrate the applicability of these detectors to the analysis of hyperspectral data. The detector successfully identified the majority of calcite occurrences in the scene (green areas in Figure1a). Note that this analysis is not performing an unmixing, as some spectral analysis methods attempt to do. Rather, we are identifying the pres-

ence or absence of specific minerals. Further, we are not attempting to estimate the abundance of the mineral present (although the strength of a mineral's absorption features correlates with its detection). Since each mineral detector is focused on only a single mineral (e.g., calcite) or mineral subclass (e.g., carbonate), entire hyperspectral data cubes can be rapidly analyzed. This makes our detectors ideal for use in exploratory data analysis, where rapid feedback enhances discovery potential.

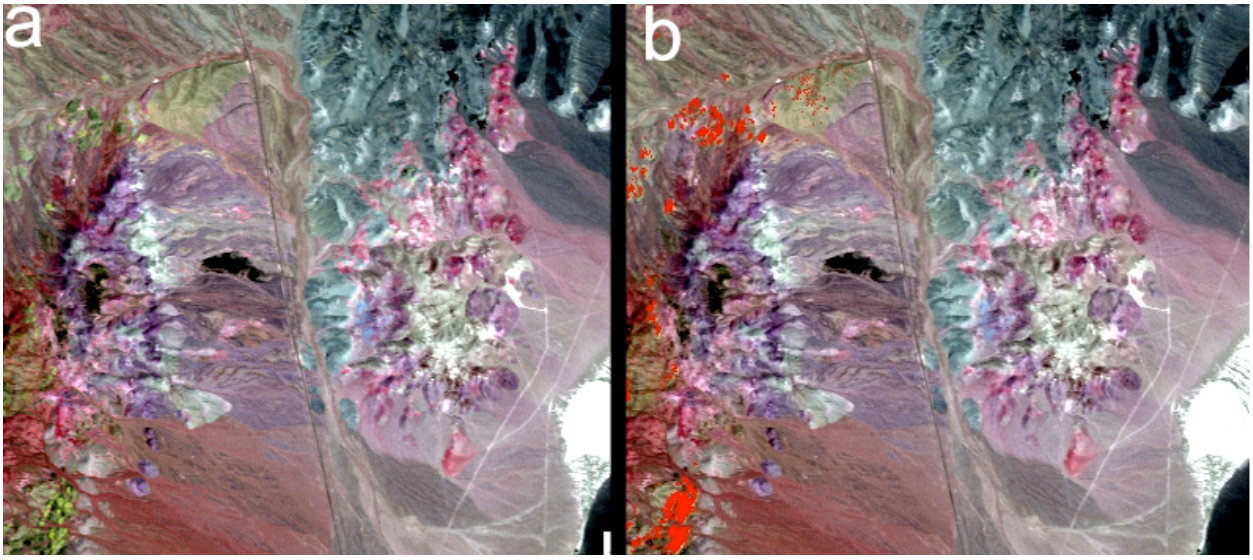

Figure 1. a) Portion of 1995 AVIRIS scene for Cuprite, Nevada. Bands 183 (2101 nm), 193 (2201 nm), 207 (2340 nm) as RGB. A highway runs north-south (north is towards the top) through the center of the image for scale. In this rendition, calcite appears as green hues. b) Same as in (a). Areas in red are identified as calcite by our SVM detector.

Detection of Gypsum in OMEGA data. We have applied the within-image training method to a single OMEGA image cube to detect the mineral gypsum ($CaSO_4 \bullet 2H_2O$; Figure 2). The detector was trained on 100s of pixels identified as containing gypsum by [3]. To reduce complexity and computation time, the detector input was limited to the 1200-2600 nm wavelength region that contains the primary gypsum absorptions. The detector succeeded in finding all of the pixels identified as gypsum outside of the training set. There were no false negatives and false positives comprised 0.8% of the total number of pixels. That the false positives are clustered in regions near the true positives (as opposed to being scattered throughout the scene, e.g.) and that the average spectra of the classified pixels show the training pixels, true positives and true positives are all of similar shape (Figure 3), may indicate the detector is sensitive to occurrences of gypsum outside of the original labeled set.

Conclusion: Hyperspectral data are powerful tools for planetary and terrestrial geology, resource identification and targeting of future landed assets. The algorithms proposed here will reduce mission analysis costs by providing a mechanism to search *all* data for minerals of interest, thus maximizing the chances of finding minerals that are of geologically importance. The state of the art in spectral identification is still a very labor-intensive endeavor. Our supervised machine learning techniques, coupled with our generative models, substantially reduce the effort required to find high priority mineral targets and mineral mixtures in hyperspectral images and point spectra. Supervised machine learning classifiers enable minerals of interest to be targeted directly and since classifiers, by design, create an internal representation to summarize their training data (spectra), they can require less computing power than the other methods outlined in this section, which makes them suitable for onboard classification and downlink prioritization. For instance, we have previously demonstrated an autonomous SVM classifier onboard the Earth Observing One (EO-1) spacecraft [24] for terrain cover classification.

The algorithms designed here are specifically for the VIS-NIR portion of the spectrum, as the lab and field instrumentation is affordable and in place. However, the methods developed here are applicable to, and set the basis for, mineral detectors in any wavelength range. These algorithms may be modified for a variety of instruments including thermal, Raman and Mössbauer spectroscopy. Automated classifiers such as those developed here may assist in the prioritization of targets that are been studied by an array of instruments, where desirable features in more than one data set would give the target a higher priority.

References: [1] Vane, G. et al. (1993) *Rem.Sens. Env. 44,* 127. [2] Pelkey, S. M., et al. (2005) *LPSC 36,* #1458. [3] Gendrin, A. et al. (2005) *Science 307*, 1587. [4] Gomez C. et al. (2006) *LPSC 37*, #1405. [5] Arvidson R. E. et al. (2005) *Science 307*, 1591. [6] Mustard J. F. et al. (2005) *Science 307,*1594. [7] Clark R. N. et al. (2003) *JGR 108*, (E12). [8] Boardman J. W. et al. (1995) *Fifth JPL Airborne Earth Science Workshop, JPL Publ. 95-1*, 23. [9] Kruse F. A. et al. (1993) *Rem. Sens. Environ., 44*, 145. [10] Mazer A. S. et al. (1988) *Rem. Sens. Environ., 24*, 201. [11] Hooke S. (2000) The JPL ASTER Spectral Library v1.1, http://speclib.jpl.nasa.gov. [12] Clark R. N. et al. (1993) U. S. Geological Survey Open File Report, v. 93-592, http://speclab.cr.usgs.gov. [13] Gilmore et al. (2000) *JGR 105*, 29223. [14] Gilmore et al. (2004) *Icarus 172*, 641. [15] Bornstein et al. (2005) *Proc. IEEE Aerospace Conf.,* #1527. [16] Bornstein et al. (2006) *Proc. IEEE Aerospace Conf.,* #1513. [17] Swayze G. A. et al. (1998) *Seventh JPL Airborne Earth Science Workshop,* JPL Publication 97-21, 383. [18] Hamilton V. E. and Christensen P. R. (2005) *Geology 33*, 433. [19] Birbring J.-P. (2005) *Science 307*, 1576. [20] Mustard J. F. and Sunshine J. M. (1995) *Science 267*, 1623. [21] Christensen P. R. et al. (1998) *Science 279*, 1692; (2000) *JGR 105,* 9609. [22] Langevin Y. et al. (2005) *Science 307*, 1584. [23] Bibring J.-P. et al. (1989) *Nature 341*, 591. [24] Castaño R. et al. (2006) *Proc. Of Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Figure 2. Results of gypsum detection on OMEGA cube 353_2 (lani Chaos). The yellow pixels correspond to the training areas, green are true positives and red are false positives.

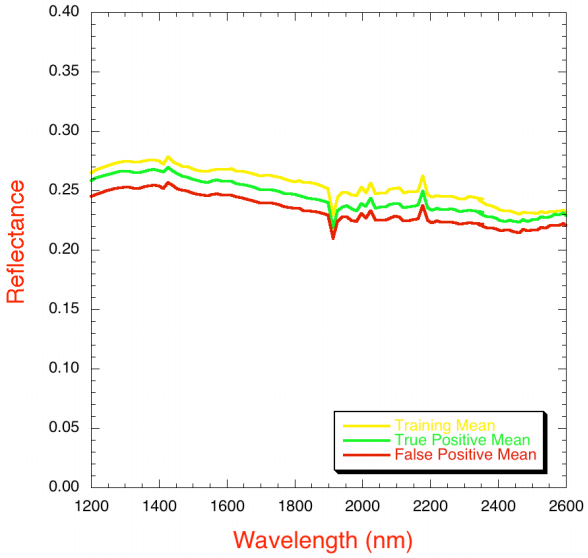

Figure 3. Average spectra of classified pixels in Figure 2.

Acknowledgements: We thank Jack Mustard and Shannon Pelkey for providing the atmospherically-corrected OMEGA image cubes as well as valuable conversations. Discussions with Scott Murchie are also appreciated. The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Specific funding for this work was provided by the Interplanetary Network Technology Program.