AUTOMATICALLY IDENTIFYING ROCK COATINGS IN LABORATORY LIBS DATA USING MACHINE LEARNING ALGORITHMS. D. A. Oyen[1] and N. L. Lanza[1], [1]Los Alamos National Laboratory, Los Alamos, NM 87545, doyen@lanl.gov.

Introduction: As ChemCam data continues to be collected from the martian surface by the Curiosity rover, automated methods for analyzing both the depth and breadth of information are needed. An important aspect of the Curiosity rover mission is the search for surface compositional features on rocks such as coatings and weathering rinds; these features provide important information about past aqueous environments and water-rock and rock-atmosphere interactions [1]. here we apply the machine learning algorithm of Gaussian graphical model structure learning to identify a rock varnish signature. In this study, we apply the algorithm to terrestrial manganese-rich rock varnish coatings on basalt rocks analyzed in the laboratory with the ChemCam engineering model to validate the use of the same algorithm to Mars samples to identify targets that indicate a rock varnish.

The analysis of compositional depth trends is not straightforward due to the high dimensionality of laser-induced breakdown spectroscopy (LIBS) spectral data (~6000 channels) [2]. Typically, to determine the presence of weathering rinds or surface coatings, the sequence of LIBS spectra obtained by 30-150 laser pulses in a single location are analyzed for systematic increases or decreases in specific elements. Instead, we use a machine learning algorithm to give a visual summary of the depth trends in rocks based on the full LIBS spectrum. By starting with the more general summary of depth trends, a broader array of patterns can be discovered before narrowing in on a specific signature. In this paper, we show that this automated method identifies compositional depth trends associated with varnish and weathering rinds on laboratory samples. With this validation, we can better understand the depth trends observed on martian rocks using ChemCam and our machine learning tool.

Method: In machine learning, probabilistic graphical models [3], and specifically, Gaussian graphical models (GGM) [4], provide a visual representation of conditional dependency relationships among a set of objects contained within a data set. Graphical models have been used extensively in fields such as biology to identify gene interaction networks and neuroscience to infer functional pathways of neural activity. Previously, we showed that we can apply this GGM algorithm to LIBS spectra to analyze the change in chemistry among sequential LIBS shots, and that these changes are indicative of depth trends consistent with geological features of interest [5]. Here we validate the GGM method on known targets.

To analyze the depth trend of a rock target at a location, we estimate *partial correlations* among spectra using the GGM algorithm. A partial correlation between shot X and shot Y is the residual correlation after accounting for all other shots. Thus, a partial correlation is an estimate of a direct dependency. If the partial correlation between X and Y is 0 then X and Y are conditionally independent. A GGM is estimated from a data matrix $\boldsymbol{X}$, where each column X_j is a shot j with spectral values X_{ij} for i in $\{1, \ldots, n\}$ wavelengths. The sample covariance matrix, $\boldsymbol{\Sigma}$, is calculated from $\boldsymbol{X}$, then the best sparse approximation, $\boldsymbol{\Theta}$, to the partial correlation matrix for a given sparsity constraint, λ, is estimated. The number of non-zero partial correlations is controlled by the value of λ, which can be any non-negative real number. If $\lambda = 0$, then all values in $\boldsymbol{\Theta}$ are non-zero. As λ increases, the more zeros that $\boldsymbol{\Theta}$ contains.

The resulting GGM is displayed using a spring layout that places strongly correlated nodes near each other as if the correlation weights are springs pulling nodes together in space. If there are no systematic trends, then the non-zero partial correlations will appear on seemingly random pairs of shots, and the displayed GGM will look like an amorphous *blob* (or *hairball* in graph theory terminology). More visually

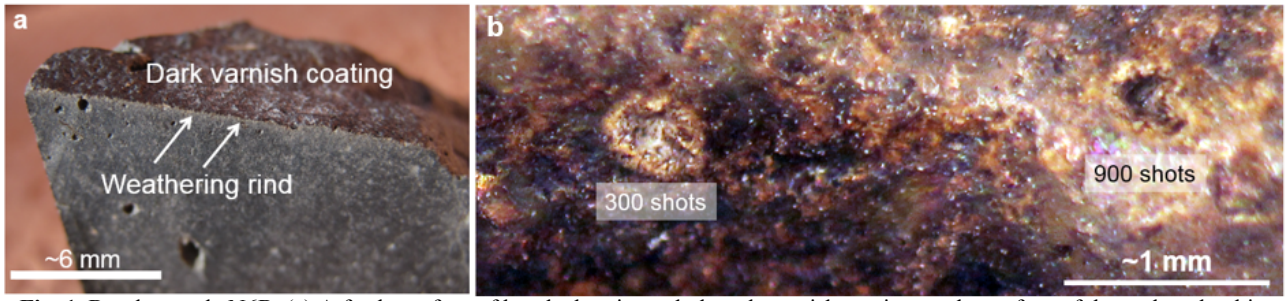

Fig. 1. Basalt sample N6B. (a) A fresh cut face of basalt showing a dark rock varnish coating on the surface of the rock and a thin bright weathering rind below the coating; the dark grey interior rock is also visible. (b) Two LIBS ablation pits in the N6B rock varnish (exterior); note that the grey interior rock is visible within the bottom of the pits. After [1].

interesting patterns emerge when there are interesting depth trends, such as a chain for systematic decrease/increase in elements, or clusters for sudden change in chemistry (such as a layer).

Data: LIBS depth profiles were performed on samples of naturally weathered basalts from the Black Point Lava Flow in Arizona, U.S.A. [1, 6] (see Fig. 1). These samples have surface coatings of rock varnish ~10 μm thick and weathering rinds ~500 μm thick below the coatings. The varnish contains abundant Mn while the basalt below contains low Mn, and previous work has shown that varnish can be discerned in LIBS data with other techniques [1, 6]. Samples were analyzed with the ChemCam engineering model under simulated martian atmospheric conditions (~6 mbar CO_2) at a standoff distance of 1.6 m. Each LIBS shot ablates on average ~0.3-0.5 μm into the rock surface, depending on rock type [1, 2]. Following standard ChemCam data treatments, data is preprocessed after the methods of [2]. We remove data for wavelengths above 850 nm, set all negative values to zero, and normalize the values for each of the three component spectrometers separately by total emission to mitigate fluctuations in laser intensity due to matrix effects.

Results: For each location, the data are made up of spectra from a sequence of 300 or 900 shots in a single location. Each shot is represented as a *node* (or *vertex*) in the graphical model. Partial correlations among nodes are calculated from the spectral data, as described in the Method section. *Links* (or *edges*) in the graphical model represent the estimated direct dependencies (or partial correlations). If two nodes are not connected by a link in the graph, then the two nodes are conditionally independent of each other given the other nodes. This means that they may be correlated, but the correlation is indirect and depends on one or more other nodes in the graph.

Fig. 2 shows the GGM graphs learned from 300 LIBS shots on the exterior (left) and the interior (right) of basalt sample N6B (Fig. 1, also see [1, 6] for additional details about this sample). The GGM from the exterior of the rock displays a distinct chain-like structure in the first 50-100 LIBS shots, indicating that the chemistry is changing systematically along these shots. The structure of the GGM from the interior sample (Fig. 2 right) is less distinct, indicating that the geochemistry is more homogenous among all LIBS shots.

Discussion: The GGMs estimated from the varnish-coated surface samples show a distinct chain-like structure that connects links between consecutive LIBS shots. This chain is most prominent on the first 50 shots, which is consistent with the laser penetrating the coating. The GGMs estimated from the interior samples show a more random assignment of links between non-consecutive shots. The chain-like structures are consistent with the rock targets having surface features with systematic change in composition (i.e., a transition between the varnish and the basalt), and not showing correlations with the interior basaltic composition until after shot 50, as was reported in [1, 6].

Further work developing machine learning algorithms to aid in discovering signatures in LIBS data, particularly for use with ChemCam, are needed; such as for identifying the specific elements that are changing in a depth trend. Machine learning tools are good at quickly identifying general trends or patterns in large data sets; and through customization machine learning algorithms can be used to discover signatures of geological features of interest. ChemCam continues to generate a wealth of data with which we hope to answer complex science questions. The development of machine learning tools tailored to this data set can aid in answering these questions and exploring interesting patterns in the data.

References: [1] Lanza et al. (2015). *Icarus*. [2] Wiens R. et al. (2012) *Space Sci. Rev.,* 170. [3] Koller and Friedman. (2009). *Probabilistic Graphical Models*. [4] Zhao T. et al (2012) *J. Machine Learning Research, 13(1):*1059–1062. [5] Oyen and Lanza. (2015). *LPSC* abstract 2940 [6] Lanza et al. (2012). *Appl. Optics* 51(7), B74-B82.

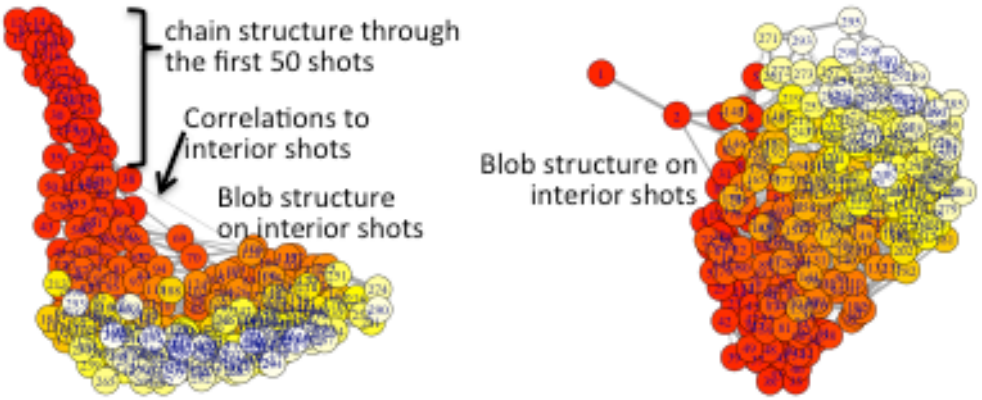

Fig. 2. (left) GGM estimated from 300 LIBS shots at Location 1 of exterior of sample N6B. **(right)** GGM estimated from 300 LIBS shots at Location 1 of interior of sample N6B. In both plots, nodes are colored from dark red for Shot number 1 to pale yellow for shot number 300.