

DISCOVERING CHEMICAL STRUCTURE IN CHEMCAM TARGETS USING GAUSSIAN GRAPHICAL MODELS: COMPOSITIONAL TRENDS WITH DEPTH. D. A. Oyen¹ and N. L. Lanza¹, ¹Los Alamos National Laboratory, Los Alamos, NM 87545, doyen@lanl.gov.

Introduction: As ChemCam data continues to be collected from the martian surface during Mars Science Laboratory rover operations, automated methods for analyzing both the depth and breadth of information are needed. The depth profiling capability of ChemCam provides information about the presence of chemical gradients in the near surface of rock and soil targets; thus ChemCam can directly observe surface features such as coatings, weathering rinds, and thin stratigraphic layers should they be present [1]. Given the high number of ChemCam observations to date (~200,000 spectra) and the high dimensionality of laser-induced breakdown spectroscopy (LIBS) spectral data (~6000 channels) [2], it is not straightforward to evaluate depth trends in composition. Here we use an automated data mining method to identify the relationships between LIBS spectra collected by Curiosity's ChemCam in Gale crater. Other automated methods typically treat each observation as independent from other observations for simplicity. However, such a simplifying assumption discards valuable information about the structure of rocks.

Method: Gaussian graphical models (GGM) [3] provide a visual representation of conditional dependency relationships among a set of objects contained within a data set. The sequence of shots measured by ChemCam on a target are not independent samples of rock, rather they are dependent samples. A correlation analysis would likely reveal correlations among all pairs of shots, yet to discover the structure of dependencies in the data, we need to analyze the *partial cor-*

relations. 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. When calculating partial correlations from sample data, the value is nearly guaranteed not to be exactly 0, even if there is no true direct dependence between variables, because of sample noise. Therefore, to estimate a graphical model, a sparsity constraint limits the number of non-zero partial correlations.

A GGM is estimated from a data matrix \mathbf{X} , where each column X_j is a shot j with spectral values X_{ij} for i in $\{1, \dots, n\}$ wavelengths. The sample covariance matrix, Σ , is calculated from \mathbf{X} , then the best sparse approximation, Θ , to the partial correlation matrix for a given sparsity constraint, λ , is estimated, as shown in Table 1. 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 Θ are non-zero. As λ increases, the more zeros that Θ contains.

Table 1. Estimation of a Gaussian Graphical Model

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij} \quad \text{and} \quad \|\Theta\|_1 = \sum_{ij} |\Theta_{ij}|$$

$$\Sigma_{jk} = \frac{1}{n} \sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$$

$$\hat{\Theta} = \arg \max_{\Theta \succeq 0} [\log \det \Theta - \text{tr}(\Sigma^{-1}\Theta) - \lambda \|\Theta\|_1]$$

Data: During normal operations, ChemCam typi-

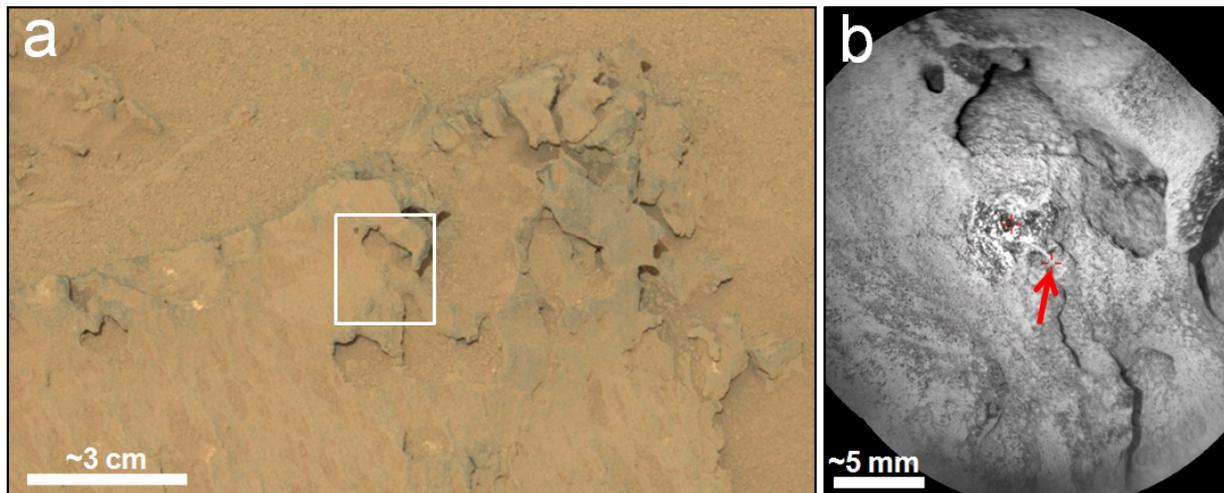


Figure 1. ChemCam target Bell Island, sampled on sol 113 (ccam02113). (a) Mastcam image of rock overview; the white box represents the ChemCam sampling region (mcam00694). (b) ChemCam Remote Microimager (RMI) mosaic of Bell Island. The red arrow indicates sampling location 1, which is described in detail in this abstract and shown in Fig. 2.

cally obtains 30 shots on a single sampling location, and often multiple sampling locations per target. Each shot ablates $\sim 0.3\text{-}0.5\ \mu\text{m}$ of material [1, 2]. We analyzed all sampling locations on rock targets in the first 587 sols of the mission. We removed data for wavelengths above 850 nm, set all negative values to zero, and normalized the values for each of ChemCam's three component spectrometers separately by total emission.

Results: For each location, the data includes spectra from a sequence of several shots in a single location, typically 30 shots. To analyze the structure of the rock target at this location, we estimate a GGM from the spectral data. Each shot is represented as a *node* (or *vertex*) in the graphical model. 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.

The likelihood that the link represents a true direct dependency, is quantified by estimating several GGMs from the same set of data with 40 different values of λ . Links that appear across more values of λ are weighted more heavily, a method known as Bayesian model averaging [4].

Figure 2 shows the graph estimated for sampling location 1 on the rock target Bell Island on Sol 113 (Fig. 1b). It shows that the first several shots are directly dependent on the prior shot and the following shot in the sequence, creating a "chain" from Shot 1 through Shot 10. This means that, for example, if Shots 4 and 6 are correlated, all of that correlation can be explained by the correlation between Shots 4 and 5 combined with the correlation between Shots 5 and 6. Shots 4 and 6 are conditionally independent given Shot 5. Shots 10 through 30 do not demonstrate the same type of chain structure, indicating that the changes in spectra from shot to shot are less distinct.

Discussion: The GGM estimated from a sequence of shots on a single location indicate that there are shot-to-shot dependencies that follow expected patterns. Without any prior information about the shot sequence, the GGM is able to reconstruct the shot sequence on the first 10 or so shots, which is consistent with observations that these first shots may differ from the underlying rock. In this case, the surface of Bell Island is extremely dusty, making dust a likely contributor to the spectra of the initial shots [e.g., 5]. Addi-

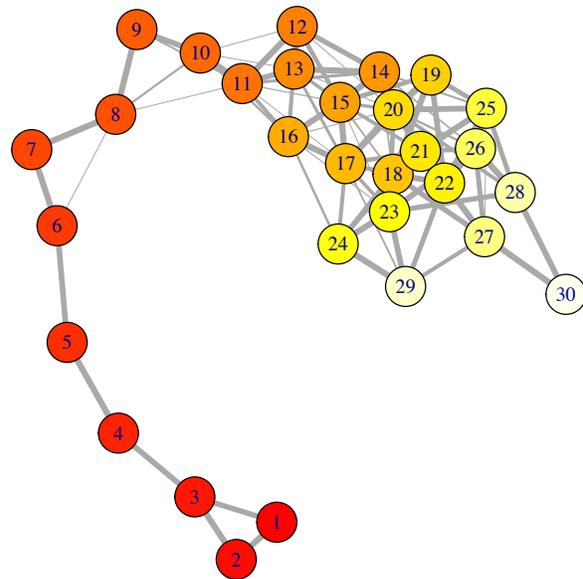


Figure 2. Graphical model of the shot-to-shot relationships in location 1 on Bell Island (30 shots, Sol 113, ccam02113, shown in Fig. 1). Nodes of graph represent shots, numbered 1 through 30. Nodes are colored by shot number, starting with dark red at shot 1 and getting lighter to light yellow at shot 30 to aid visualization.

tionally, the first sampling location of Bell Island appears to be a thin, sulfate-rich vein [6]. Our results suggest that in its 30 shot analysis at this location, ChemCam sampled through surface dust into the sulfate vein below. Examination of the spectral data confirm that the elements enriched in dust (e.g., Mg [5]) decrease with depth (shot number) while S and Ca increase. Further examination of the spectra will help to elucidate the transition between dust and vein compositions.

Further investigations are planned to generate maps of multiple locations from the same rock target, as well as relationships among rocks. Other patterns of interest may emerge from our automated data mining algorithm that will make it easier to interpret large collections of data without losing information about the dynamic nature of the observations.

References: [1] Lanza et al. (2014). *Icarus*, in press. [2] Wiens R. et al. (2012) *Space Sci. Rev.*, 170. [3] Zhao T. et al (2012) *J. Machine Learning Research*, 13(1):1059–1062. [4] Wasserman L. (2000) *J. Math. Psych.*, 44(1):92–107. [5] Lasue et al. (2014). *LPSC XLV*, abstract #1224. [6] Nachon et al. (2014). *J. Geophys. Res. Planets* 119, 1991-2016.