UNUSUAL CHEMCAM TARGETS DISCOVERED AUTOMATICALLY IN CURIOSITY'S FIRST NINETY SOLS IN GALE CRATER, MARS. K. L. Wagstaff[1], N. L. Lanza[2], and R. C. Wiens[2], [1]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109 USA, kiri.wagstaff@jpl.nasa.gov, [2]Los Alamos National Laboratory, nlanza@lanl.gov.

Introduction: As mission data archives grow, it becomes progressively more difficult to spot individual observations that possess unusual features that merit further study. We have developed an automated method for prioritizing observations within large data sets. The DEMUD (Discovery through Eigenbasis Modeling of Uninteresting Data) algorithm [1] uses principal components modeling and reconstruction error to prioritize data. DEMUD goes beyond standard methods for anomaly or novelty detection to also offer *explanations* for its selections. We applied DEMUD to data collected by the ChemCam instrument during the Curiosity rover's first 90 sols on Mars and identified several targets with unusual chemical compositions. Ultimately, DEMUD can save significant human review time by quickly focusing attention on the most unusual observations. Human judgment can then determine whether these anomalies are due to instrument/data artifacts or constitute a novel scientific discovery.

Method: DEMUD [1] is an iterative method designed to facilitate scientific discovery in large data sets. It progressively selects the next most interesting (i.e., surprising with respect to what has already been seen) observation. Starting with an arbitrarily chosen observation, it performs a truncated Singular Value Decomposition (SVD) to obtain a low-dimensional eigenbasis representation of that observation. For a data set $\mathbf{X}$, we obtain the model $\mathbf{U}$ such that $\mathbf{X} = \mathbf{U \Sigma V^T}$. The rest of the data is projected into this model, then reconstructed in the original feature speace to obtain a novelty score, which is the reconstruction error: $\mathbf{R(x)} = ||\mathbf{x} - (\mathbf{UU^T}(\mathbf{x} - \boldsymbol{\mu}) + \boldsymbol{\mu})||_2$, where $\boldsymbol{\mu}$ is the data mean. The item $\mathbf{x}$ with the largest $\mathbf{R(x)}$ is selected and removed from $\mathbf{X}$, and the model $\mathbf{U}$ incorporates $\mathbf{x}$ using an incremental SVD update. Thus, each new observation becomes part of DEMUD's growing knowledge and biases it away from selecting similar observations.

Explanations. DEMUD differs from all existing anomaly detection methods by providing explanations. Each feature of the selected item $\mathbf{x}$ has a reconstruction error that contributes to $\mathbf{R(x)}$. Features with large errors are those that best explain why $\mathbf{x}$ is anomalous.

Data: ChemCam is a Laser-Induced Breakdown Spectroscopy (LIBS) instrument onboard the Curiosity Mars rover [2]. During the rover's first 90 sols on Mars, ChemCam collected 11,750 spectra from martian targets. We removed observations 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. We then applied a median filter with a window size of 7 to remove shot noise.

Results: We ran DEMUD using K=1829 principal components (captures 90% of data variance) to model the data with $\mathbf{U}$. We initialized the model with the single most unusual spectrum in the data set (as chosen by a regular SVD), which was shot 12 on Kilian from sol 72 (DEMUD's selection 0). DEMUD's next 10 selections are shown in Table 1. Each selection is accompanied by an automatically generated explanation in which DEMUD identified spectral features with the largest reconstruction error and looked up the corresponding element associated with each peak. Explanations apply to individual shots, not the entire target. Here, "elevated" and "reduced" are judgments about the observed values compared to the reconstructed values that DEMUD expected to see, given the preceding observations. Thus, Epworth shot 8 has elevated Ca and reduced Na with respect to the first two samples (Kilian shot 12 and Rocknest3_3 shot 2), while

Table 1. First ten ChemCam targets selected by DEMUD from 11,750 spectra collected during MSL's first 90 sols on Mars.

Sel	Target	Sol	Shot	Explanation (automatically generated by DEMUD)
1	Rocknest3_3	88	2	Reduced Ca, Na, and O
2	Epworth	72	8	Elevated Ca; reduced Na
3	Kam	43	18	Elevated C, Si, and Al
4	Rocknest3_3	88	22	Reduced Ca and O
5	Kenyon	82	25	Elevated O; reduced Ca and Na
6	Murky	22	19	Elevated Mg
7	Rocknest3_3	88	28	Elevated Ca
8	Stark	15	48	Elevated C, Si, and O; reduced Al
9	Kilian	72	10	Reduced Ca
10	Thor_Lake	22	34	Elevated Ba; reduced Ca and K

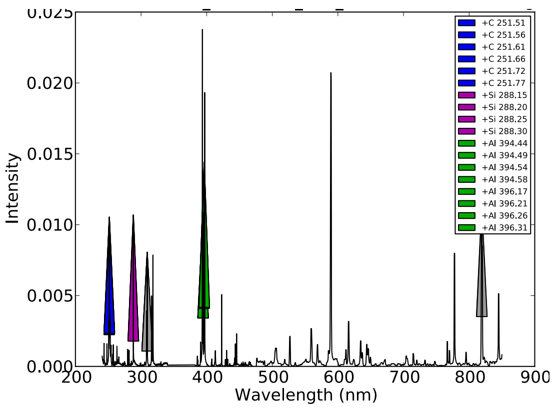

Figure 1. DEMUD selection 3 (Kam shot 18)

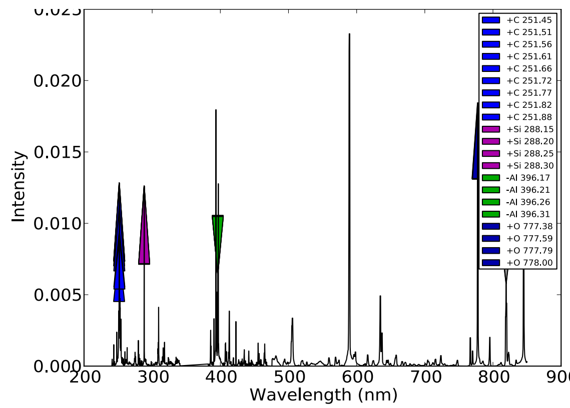

Figure 2. DEMUD selection 8 (Stark shot 48)

Kam shot 18 has elevated C, Si, and Al with respect to the first three samples. DEMUD's explanations cover only what is unexpected and therefore of interest. If all samples possess a peak at the same wavelength, it will not be included in the explanations.

The explanations can also be shown visually, as in Figure 1. The explanations for why Kam shot 18 was selected appear as colored triangles that point from the expected value to the observed value. The C (blue, possibly atmospheric), Si (purple), and Al (green) peaks are all higher than expected given the three preceding spectra. The Na peak at 588 nm is not highlighted because it was already observed in Kilian shot 12 and therefore is not unusual. Gray triangles are unexpected deviations that do not map to known elements in DEMUD's reference library. They may be artifacts or omissions in the library.

Figure 2 shows Stark shot 48, which also shows elevated C and Si because these peaks are higher than those observed in Figure 1 (and other preceding selections). DEMUD also highlighted the fact that Stark's Al peak at 398 nm is smaller than expected.

Discussion: DEMUD's are consistent with summary observations for this data set, e.g., Epworth has elevated Ca [3,4] and Stark has elevated Si [3]. However, DEMUD digs deeper to identify individual shots that exhibit unusual spectral features, rather than summarizing a target with a single characterization. For example, Rocknest3_3 appears three times in the first ten selections with different explanations. Rocknest as a whole has elevated Ca [4], but DEMUD highlights between-shot variation. In addition, DEMUD emphasizes that Rocknest3_3 appears to be a relatively heterogeneous target in the locations sampled on sol 88. Interestingly, although Rocknest3 contains higher Fe than many other Gale crater targets [5], this was not highlighted by DEMUD; this is likely due to the fact that other elements show more variation than Fe in the shots analyzed here (from the first 90 sols only).

DEMUD selects indiidual observations that differ in some way from previous selections. Like a human, it learns progressively; its selections become more refined as more items are selected and then incoporated into its model. It highlights individual features (for ChemCam, individual elemental emission peaks) that provide an initial explanation for why each item may be of interest. It is then up to human review to determine whether these features represent scientifically meaningful deviations or instrument artifacts. For example, DEMUD identified a peak in data from Thor_Lake at 453 nm and linked it to the closest peak in its library, which is Ba at 455.4 nm. Subsequent work on trace elements found that this particular peak is likely due to interfering Ti rather than Ba [6]. However, DEMUD correctly flagged this sample for review, since differentiating between interfering peaks requires careful further study by a subject expert.

DEMUD can serve as a useful operational tool to quickly direct attention to observations of interest. It complements other analysis methods that extract overall trends in the data set by instead highlighting unusual individual features and items. In an operational setting, DEMUD could model all data from all preceding sols and perform a first-cut prioritization of the spectra collected from the current sol accompanied by explanations for how the items were ranked.

References: [1] Wagstaff K. L. et al. (2013) *AAAI Conf. on Artificial Intelligence*. [2] Wiens R. et al. (2012) *Space Sci. Rev.,* 170. [3] Gasnault O. et al. (2013) *LPS XLIV*, Abstract #1994. [4] Clegg S. M. et al. (2013) *LPS XLIV,* Abstract #2097. [5] Blaney, D. et al. (submitted) *JGR-Planets MSL special issue*. [6] Ollila, A.M. et al. (accepted) *JGR-Planets MSL special issue*.

Acknowledgments: This work was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Government sponsorship acknowledged.