COMPARISON OF LASSO AND ELASTIC NET REGRESSION FOR MAJOR ELEMENT ANALYSIS OF ROCKS USING LASER-INDUCED BREAKDOWN SPECTROSCOPY (LIBS). M. V. Ozanne[1], M. D. Dyar[1], M. L. Carmosino[2], E. A. Breves[1] , S. M. Clegg[3], and R. C. Wiens[3], [1]Dept. of Astronomy, Mount Holyoke College, 50 College Street, South Hadley, MA 01075, ozann20m@mtholyoke.edu; [2]Computer Science Dept., University of Massachusetts, 140 Governor's Dr., Amherst, MA 01003, [3]Los Alamos National Laboratory, P.O. Box 1663, MS J565, Los Alamos, NM 87545.

Introduction: Laser-induced breakdown spectroscopy (LIBS) will be used by the Mars Science Laboratory rover, *Curiosity*, to analyze soil and rock. Quantitative elemental analysis from LIBS spectra presents challenges to statistical analyses due to noise and the fact that each element has many peaks with highly-correlated intensities. Several different regression methods eliminate the collinear nature of the covariates by shrinking some model coefficients and reducing others to zero. Two of these are compared here: the lasso (least absolute shrinkage and selection operator) [1] and the elastic net [2].

Samples and Experimental Methods: LIBS was used to analyze a suite of 100 igneous rocks at Los Alamos National Laboratory as described in [3]. Samples were analyzed from a standoff distance of 9 m; they were placed in a chamber filled with 7 Torr CO_2 to simulate the Mars surface pressure. A 1064-nm Nd:YAG laser operating at 17 mJ/pulse was used to ablate the samples. Three Ocean Optics HR2000 spectrometers with UV (223-326 nm), VIS (328-471 nm), and VNIR (495-927 nm) wavelength regions were used to collect the optical emission from the sample plasma. Major element oxide concentrations were measured by XRF at the Univ. of Massachusetts, Amherst, in the lab of Dr. Michael Rhodes [4].

Statistical Methods: The lasso is a penalized shrunken regression method that selects specific channels for each element that explain the most variance in its concentration. It creates a sparse model by shrinking some coefficients and setting others equal to zero. Coefficients are constrained by values of the t and α penalties, where t controls the level of sparsity of the model and α controls the balance between averaging correlated features and the number of non-zero coefficients. The model finds a smaller subset of predictor variables that have the strongest effects on the response variable. This reduces a large model to a sparse, more interpretable model [1]. In these models, y_i is the elemental abundance in sample i and x_{ij} is the intensity at wavelength j of sample i. The lasso shrinks the model coefficients (β_j) according to the following penalty.

$$\hat{\beta}^{lasso} = argmin_{\beta} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2,$$
$$subject\ to \sum_{j=1}^{p} |B_j| \le t.$$

In the lasso penalty, $t \ge 0$.

Elastic net regression is a generalization of the lasso. It retains the sparse properties of lasso regression and the stability of a saturated model when the number of observations (wavelength channels) per sample (p) is much greater than the numberof samples (N) (the $p>>N$ case). It can also select groups of correlated variables. "It is like a stretchable fishing net that retains 'all the big fish'" [3]. The elastic net shrinks the model coefficients as shown below, with $t \ge 0$, $0 \le \alpha \le 1$. The lasso is a special case of the elastic net with $\alpha = 0$.

$$\hat{\beta}^{elastic.net} = argmin_{\beta} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2,$$
$$subject\ to \sum_{j=1}^{p} \alpha|B_j| + (1-\alpha)\beta_j^2 \le t.$$

The second term of the elastic net penalty promotes averaging of highly correlated features. Thus, unlike the lasso, which indiscriminately selects a feature from a cluster of highly correlated features to represent in the model, the elastic net retains information about all of the features in the cluster by averaging them.

The lasso and the elastic net penalties are compared using their mean-squared error (MSE) values for the elemental oxide compositional predictions [5]. The MSE values are compared for the ten major elemental oxides: SiO_2, Al_2O_3, MnO, MgO, CaO, Na_2O, K_2O, TiO_2, P_2O_5, and Fe_2O_3.

Statistical Analysis: Wavelength calibration was performed for each of the three spectral regions to standardize the wavelength scale. Laser shots were averaged and smoothed; the baseline was modeled and subtracted. Both lasso and elastic net models were built using the R statistical software package glmnet [6]. The elastic net penalty used the literature standard value of $\alpha = 0.5$ to calculate the model [2]. The number of components for both models was chosen using a heuristic that selects the component number with a MSE of prediction within one SE of the minimum (Fig. 1). As such, the elastic net is only partially tuned: t has been optimized but not α. Further work with cross-validation procedures will optimize prediction results by manipulating α values.

Discussion: Root mean squared errors of prediction (RMSEP) for both lasso and elastic net have a maximum value of <3.0 in units of wt.% oxide for SiO_2 (Table 1) and are smaller for all other elements.

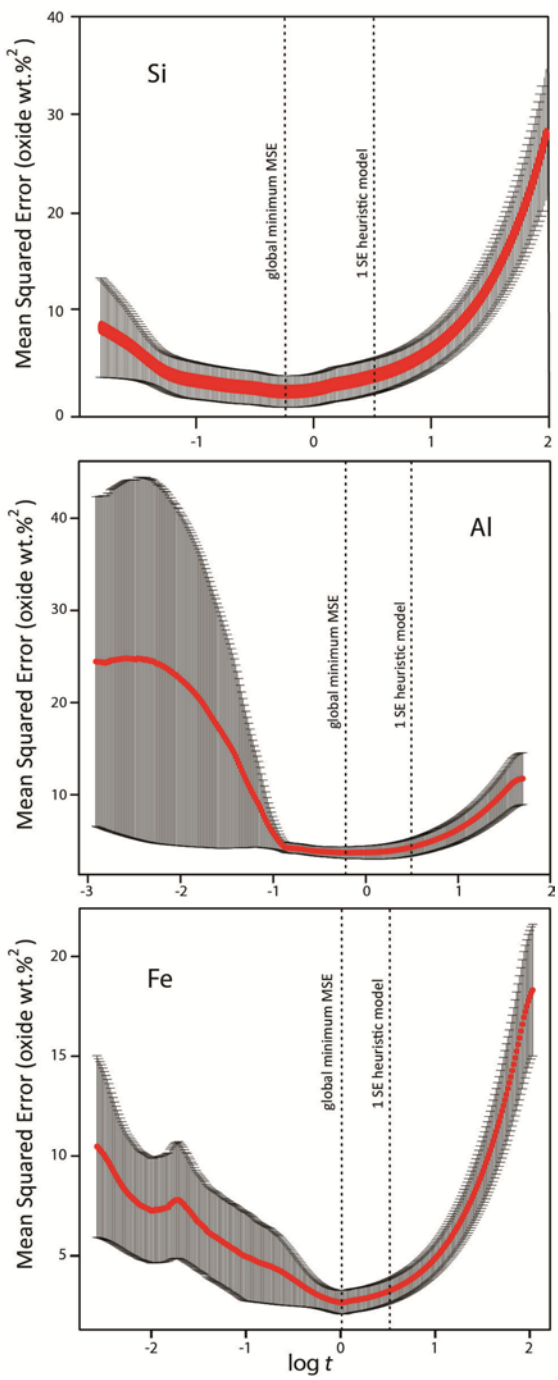

Figure 1. Mean square error (in units of wt% oxides squared) plotted against the log of the t value for Si, Al, and Fe. The global minimum MSE on each plot indicates the value of t for which the smallest prediction error was obtained. The 1 SE heuristic model is the value of t used for predictions because this value of t strikes a balance between accuracy and model simplicity.

For seven of the ten major elemental oxide compositions, the elastic net model produces a smaller RMSEP value. The values for both models, however, are largely comparable. The elastic net model has more non-zero coefficients than the lasso for each of ten major elemental oxides. This is due to the averaging nature of the elastic net penalty. This should provide a more stable model than the lasso model while retaining the sparse property inherent in the lasso.

Without tuning of α, the lasso and the elastic net produce roughly comparable results. They are useful for feature selection in different situations because of the number of variables that they select for their models. In cases where the amount of data is fairly substantial, the elastic net model is preferable for feature selection because it provides more information and a more stable model. In instances where the amount of data is lacking, the lasso allows for feature selection without overwhelming the data with too many model variables. Thus, both methods are useful.

Table 1. Prediction Errors for Tuned Models

	Lasso			Elastic Net		
	RMSEP	MSE	SE	RMSEP	MSE	SE
SiO_2	**2.61**	6.82	1.87	2.69	7.25	1.78
Al_2O_3	2.41	5.81	1.26	**2.18**	4.76	0.88
TiO_2	**0.40**	0.16	0.03	0.41	0.17	0.04
Fe_2O_3	1.78	3.15	0.63	**1.77**	3.15	0.77
MgO	1.91	3.65	0.52	**1.76**	3.10	0.51
MnO	0.03	0.00	0.00	**0.03**	0.00	0.00
CaO	**1.03**	1.06	0.17	1.04	1.08	0.17
K_2O	0.52	0.27	0.04	**0.47**	0.22	0.06
Na_2O	0.74	0.54	0.10	**0.70**	0.49	0.07
P_2O_5	0.16	0.03	0.01	**0.14**	0.02	0.00

RMSEP = root mean squared error of prediction, in the same units as the original measurements (wt.% oxide). MSE is the mean squared error, and SE is the standard error of the MSE.

Implementation of these methods not only yields robust elemental compositions, but also provides channel selections for each element that can be directly related to known atomic emissions, reflecting underlying physical processes. The sparser sets of correlation coefficients make them well-suited for understanding LIBS spectra (with other external variables in addition to spectral features) that require lower dimensional data with easily-interpreted features. Further work is in progress to compare these results to PLS and other shrunken regression methods for this same data set.

Acknowledgments: This research was supported by NASA grants NNG06GH35G and NNX09AL21G from the Mars Fundamental Research Program.

References: [1] Tibshirani R. (1996) *Journal of the Royal Satistical Society. Series B*, 58, 267-288. [2] Zou, H. and Hastie, T. (2005) *Journal of the Royal Statistical Society. Series B*, 67, 301-320. [3] Tucker J.M. et al. (2010) *Chem. Geol.*, 277, 137-148. [4] Rhodes J. M. and Vollinger M. J. (2004) *Geochem. Geophys.Geosyst., 5*, Q03G13. [5] Hastie et al. (2009) *Elements of Statistical Learning*, 661-664. [6] Friedman J. et al. (2010) *glmnet*, CRAN-R project.