Spectrochimica Acta Part B 107 (2015) 1–10

Contents lists available at ScienceDirect

Spectrochimica Acta Part B

journal homepage: www.elsevier.com/locate/sab

A study of machine learning regression methods for major elemental analysis of rocks using laser-induced breakdown spectroscopy

Thomas F. Boucher [a,*], Marie V. Ozanne [b], Marco L. Carmosino [a], M. Darby Dyar [b], Sridhar Mahadevan [a], Elly A. Breves [b], Katherine H. Lepore [b], Samuel M. Clegg [c]

[a] School of Computer Science, University of Massachusetts Amherst, 140 Governor's Drive, Amherst, MA 01003, United States.
[b] Department of Astronomy, Mount Holyoke College, South Hadley, MA 01075, United States
[c] Los Alamos National Laboratory, P.O. Box 1663, MS J565, Los Alamos, NM 87545, United States

ARTICLE INFO

Article history:
Received 6 May 2014
Accepted 3 February 2015
Available online 12 February 2015

Keywords:
Laser-induced breakdown spectroscopy (LIBS)
Partial least squares (PLS)
Support vector regression (SVR)
Lasso
Principal component regression (PCR)

ABSTRACT

The ChemCam instrument on the Mars *Curiosity* rover is generating thousands of LIBS spectra and bringing interest in this technique to public attention. The key to interpreting Mars or any other types of LIBS data are calibrations that relate laboratory standards to unknowns examined in other settings and enable predictions of chemical composition. Here, LIBS spectral data are analyzed using linear regression methods including partial least squares (PLS-1 and PLS-2), principal component regression (PCR), least absolute shrinkage and selection operator (lasso), elastic net, and linear support vector regression (SVR-Lin). These were compared against results from nonlinear regression methods including kernel principal component regression (K-PCR), polynomial kernel support vector regression (SVR-Py) and *k*-nearest neighbor (*k*NN) regression to discern the most effective models for interpreting chemical abundances from LIBS spectra of geological samples. The results were evaluated for 100 samples analyzed with 50 laser pulses at each of five locations averaged together. Wilcoxon signed-rank tests were employed to evaluate the statistical significance of differences among the nine models using their predicted residual sum of squares (PRESS) to make comparisons. For MgO, SiO_2, Fe_2O_3, CaO, and MnO, the sparse models outperform all the others except for linear SVR, while for Na_2O, K_2O, TiO_2, and P_2O_5, the sparse methods produce inferior results, likely because their emission lines in this energy range have lower transition probabilities. The strong performance of the sparse methods in this study suggests that use of dimensionality-reduction techniques as a preprocessing step may improve the performance of the linear models. Nonlinear methods tend to overfit the data and predict less accurately, while the linear methods proved to be more generalizable with better predictive performance. These results are attributed to the high dimensionality of the data (6144 channels) relative to the small number of samples studied. The best-performing models were SVR-Lin for SiO_2, MgO, Fe_2O_3, and Na_2O, lasso for Al_2O_3, elastic net for MnO, and PLS-1 for CaO, TiO_2, and K_2O. Although these differences in model performance between methods were identified, most of the models produce comparable results when $p \leq 0.05$ and all techniques except kNN produced statistically-indistinguishable results. It is likely that a combination of models could be used together to yield a lower total error of prediction, depending on the requirements of the user.

1. Introduction

A laser-induced breakdown spectrometer (LIBS), along with a remote microscopic imager, comprises ChemCam [1,2], a payload instrument on the Mars Science Laboratory (MSL) rover *Curiosity*. This LIBS instrument records emission spectra in the ultraviolet (UV), violet (VIO), and visible to near-infrared (VNIR) ranges. The laser can be focused on a small location size of roughly <0.5 mm from a standoff distance of up to 7 m. ChemCam is being used to determine chemical compositions of dust, rocks, and minerals on the Martian surface. To aid in such quantitative analyses, a broad training set of LIBS spectra of geological standards with known compositions is being developed for calibration [3]. The goal of ChemCam is to produce robust, accurate chemical analyses of minerals, rocks, and soils on the Martian surface.

However, producing quantitative chemical analyses from LIBS data is a challenging task due to the wide variety of chemical compositions found on Mars. Ionization states from the many different elements found in geological materials may interact in the LIBS plasma, causing variations in line intensities that defeat univariate analysis techniques using single-peak calibrations of intensity vs. concentration. Multivariate analysis techniques are thus needed to account for the covariate interactions that occur within the LIBS plasma. They are designed to provide stable models when the data suffer from multicollinearity, and are better suited to LIBS data analysis.

* Corresponding author.
E-mail address: boucher@cs.umass.edu (T.F. Boucher).

http://dx.doi.org/10.1016/j.sab.2015.02.003
0584-8547/

Thus, this paper explores a variety of current machine learning algorithms for regression problems and compares their performance on a suite of 100 spectra from igneous and meta-igneous rocks. LIBS spectral data are analyzed here using linear methods including partial least squares (PLS-1 and PLS-2), principal component regression (PCR), least absolute shrinkage and selection operator (lasso), elastic net, and linear support vector regression (SVR-Lin). These were compared against results from nonlinear methods including kernel principal component regression (K-PCR), polynomial kernel support vector regression (SVR-Py) and *k*-nearest neighbor (*k*NN) regression to discern the most effective models for interpreting elemental concentration from LIBS spectra of geological samples. Ten-fold cross-validation was used to train the parameters and tune the hyperparameters of each model using 70 samples while 30 samples were held out for use as a test set. Wilcoxon signed-rank tests were employed to evaluate the statistical significance of differences among the nine models using their predicted residual sum of squares (PRESS) to make comparisons. Results show the advantages of linear models for this application, and lend insights into best practices for interpretation of data from ChemCam and other LIBS studies of geological samples elsewhere in the solar system.

2. Background

LIBS relies on quantized valence-electron transitions that occur when the electrons move to an excited state in the presence of an excitation source and subsequently decay back down to their ground states, emitting photons. When these transitions are detected by a spectrometer, emission lines are observed at wavelengths that are specific to the elemental or ionic electron source.

LIBS is challenging to use for geological sample analysis because peak intensities and areas are influenced by interactions in the plasma that are partially a function of the sample's chemical composition. These interactions are collectively referred to as matrix effects; they are chemical properties of a material that influence the extent to which a given wavelength emission is detected compared to the true abundance of the parent element. The matrix effects are related to the relative abundances of neutral and ionized species within the plasma, collisional interactions within the plasma, laser-to-sample coupling efficiency, and self-absorption [4]. Fortunately, advanced statistical analysis techniques can tease out relationships that may be obscured by matrix effects.

Multivariate analyses have been used increasingly for LIBS over the last decade, starting with the applications of principal components [5] and partial least squares [4–10]. A few other methods have been investigated, such as artificial neural networks (ANN) [11]; however, results showed that PLS was equivalent or superior to ANN. A few forays have been made into the sparser models (lasso) [12] and intelligent selection or rejection of training set spectra based on clustering methods [11]. Both of these show promise in improving results, particularly with clustering, and in more closely connecting the models with physical details, i.e., with lasso predominantly using the emission lines of the element of interest. Here we follow up on these works by comparing and contrasting additional methods for providing sparseness to the data.

An ideal regression model for LIBS should be sparse, interpretable, and well predicting. The property of sparsity, in which a small subset of predictor variables drives the prediction results, can be critical to instrument design because it may enable improved count rates and higher-resolution spectra by guiding sampling to fewer channels more frequently. It may not, however, enable model interpretation, because the chosen features are dependent upon a complicated convolution of end-member oxide spectra, experimental conditions, and measurement errors [13–15]. In this paper, several multivariate analysis techniques that meet these criteria to varying degrees are utilized and compared to assess the effectiveness of each model and the effects of training set size on the resultant predictions.

Table 1 provides a summary of the methods considered in this study. The following discussion provides some background on the techniques to be compared.

Table 1
Summary of models used.

Method	Summary	Tuning parameters	Advantage(s)	Disadvantage(s)	Other
PLS	Projects explanatory matrix, **X**, into a subspace of latent components that maximize the covariance of **X** and the response matrix, **Y**.	*k*, # of components	Used when **X** has many collinear features and when p> > N. Provides a stable multivariate model that can account for all oxides (PLS-2).	Provides a complex model in which all coefficients are linear combinations of the original channels. Involves a complex optimization problem with no simple, closed-form representation.	Linear, uses all channels (not sparse)
Lasso	Shrinks some coefficients and sets others equal to zero in accordance with shrinkage parameter. Provides a sparse model that can be used for both feature selection and composition predictions.	α, sparsity weight	Provides an interpretable model, selects subset of predictors with the strongest effects on the response variable. Can be used for feature selection when less data are available.	Arbitrarily chooses one covariate from a group of highly collinear covariates to use in the model and discards the rest [18].	Linear, sparse, eliminates noisy channels
Elastic net	Extends the lasso. Shrinks some coefficients and sets others equal to zero; averages highly correlated features and shrinks averages. Provides a sparse model that has more terms than the lasso and can be used for feature selection and composition predictions.	α and l_1 ratio	Performs well in the $p >> N$ case. Provides an interpretable model that is more stable than the lasso. Useful for feature selection.	Cannot be used for feature selection in situations when less data are available because it overwhelms the data with too many model variables.	Linear, sparse, eliminates noisy channels
PCR	Projects data to a low-dimensional uncorrelated subspace, then uses ordinary least squares to regress in the latent space.	*k*, # of components	De-correlates the data and reduces its dimensionality, combating the "curse of dimensionality"	Higher order polynomial kernels tend to over-fit the training set and poorly predict the testing set in this application.	May be linear or nonlinear; both use all channels
SVR	Uses only a subset of the training data (support vectors) to construct a model that is most generalizable. Can be linear or nonlinear depending on the kernel function used.	ϵ, sensitivity	Performs well with a linear kernel. Can be either linear or nonlinear depending on the kernel.	As above, polynomial kernels tend to over-fit the training set and poorly predict the testing set in this application.	May be linear or nonlinear; either uses all channels
*k*NN	A nonlinear regression model that predicts samples using a weighted interpolation of the *k* nearest training samples.	*k*, # of neighbors	Requires no model training other than choosing the number of neighbors, reducing run time and making it scale well to large data sets.	Tends to over-fit the training data and is only as effective as the distance metric used to compare samples.	Nonlinear, uses all channels

2.1. Partial least squares

Partial least squares (PLS), also known as projection to latent structures, was developed for use in situations where explanatory variables (p) significantly outnumber observations (N), such that $p \gg N$. It has been used to analyze data from a variety of types of spectroscopy, including, but not limited to, near infrared reflectance (NIR) spectroscopy, Fourier transform infrared (FTIR) spectroscopy, and Fourier transform-Raman (FT-Raman) spectroscopy. PLS calculates components that maximize the covariance between the feature and response matrices [16], and is especially well suited for problems with many, highly correlated features and multiple responses [17,18].

PLS sequentially chooses directions, or components, of maximal covariance from the feature matrix, **X**, and the response matrix, **Y**, to determine the model coefficients using a two-step process. The first step is the shrinkage step, in which the shrinkage penalty determines the number of factors to be included in the regression. This shrinks the feature matrix by projecting it down from p-dimensional space into a smaller q-dimensional vector space. In the context of this project, $p = 6144$, the number of channels (wavelength values) at which elemental intensity is measured, and $q \approx 10$, the number of components typically used. The second step follows ordinary least squares by regressing the response on the components generated in the first step to minimize the residual sum of squared error. For this project, a single response model, PLS-1, and a multiple response model, PLS-2, are trained for comparison. Each has only one hyperparameter, which is the number of components to include in the projection.

2.2. PCR and kernel PCR

Principal component regression (PCR) is a linear regression model that uses principal component analysis (PCA). PCA is a statistical technique that linearly transforms a data matrix with possibly correlated features into an orthogonal data matrix of uncorrelated features called the *principal components* [19]. The first principal component is the direction that explains the most variance of the data, and the successive components each explain the most variance under the condition that they are orthogonal to the previous components. PCA can also be used for dimensionality reduction by limiting the number of principal components used. Discarding the later principal components may eliminate noise present in the original feature matrix. The principal components are calculated using either an eigenvector decomposition of the feature covariance matrix or, equivalently, a singular-value decomposition of the feature matrix directly, and so it is recommended to mean center the original feature matrix before decomposition. PCR is an ordinary least squares method; however, instead of regressing directly on the feature matrix, PCR regresses on the principal components of the feature matrix. PCR has a single hyperparameter, which is the number of components to include in the model.

Kernel PCR (K-PCR) is a variation of PCR that uses *kernel PCA* instead of standard PCA to create a nonlinear kernel-based model [20]. K-PCA is useful when the variance of the feature matrix cannot be well explained with a linear hyperplane (e.g., concentric spheres of data). Instead of directly calculating nonlinear principal components, the feature matrix is implicitly mapped into a higher dimensional *kernel space* where a higher dimensional hyperplane can better fit the direction of highest variance. The kernel principal components are calculated as the eigenvectors of the kernel matrix K, where $K_{i,j} = (\Phi(x_i) \cdot \Phi(x_j))$ and Φ is the map to high dimensional space. Because $\Phi(x_i)$ is only used in the dot product, the *kernel trick* can be used, whereby the dot products between projected samples are replaced by a kernel function that maps samples to an inner product space without having to explicitly calculate Φ, so $K_{i,j} = k(x_i, x_j)$ [21]. It can be shown using functional analysis that the kernel function k computes the inner product in some high dimensional space. In this project, K-PCR was trained with a second order polynomial kernel, $K(x, y) = (\sum_i x_i y_i)^2$. Higher-ordered polynomial kernels and RBF kernels were not included in our experiments because preliminary trials showed that both tended to over-fit the training set and poorly predict the test set in this application. Similar to PCR, K-PCR is an ordinary least squares method that regresses on the K-PCA transformed feature matrix.

2.3. Lasso and elastic net

Least absolute shrinkage and selection operator (lasso) regression is an ordinary least squares regression model with an l_1 penalty on the model coefficients to induce sparsity. The lasso provides a sparse model by shrinking some coefficients and setting most other coefficients to zero. Under this principle, it is assumed that a smaller subset of the predictor variables is driving the prediction results. Thus, other coefficients can be excluded from the model (i.e., set to zero) with no significant performance loss. This reduces a sizable, largely uninterpretable model to a sparse, more interpretable model.

To calculate its model coefficients, the lasso solves the following optimization problem:

$$\underset{\beta}{argmin}\left\{\frac{1}{2}||y-X\beta-\beta_0||_2^2+\alpha||\beta||_1\right\}. \tag{1}$$

Adding a regularizer to ordinary least squares smooths the model coefficients and prevents the model from overfitting the training data. Unlike l_2-regularized models, i.e., ridge regression models, the l_1-regularizer is able to perform automatic feature selection by constricting feature coefficients to zero. For problems with many superfluous features, like broadband LIBS data, the lasso can eliminate these noisy features that may otherwise hinder the model. Parsimonious models have been shown to be effective for modeling in chemometrics [22,23]. The lasso has one hyperparameter α that controls the constriction level of the coefficient vector β.

Elastic net regression is a hybrid of lasso and ridge regression that retains the sparse properties of lasso regression and the stability of ridge regression in the $p \gg N$ case. It can also select groups of correlated variables. To calculate its model coefficients, the elastic net solves the following optimization problem:

$$\underset{\beta}{argmin}\left\{\frac{1}{2}||y-X\beta-\beta_0||_2^2+\alpha\left(\lambda||\beta||_1+\frac{1-\lambda}{2}||\beta||_2^2\right)\right\} \text{ where } 0\le\lambda\le 1. \tag{2}$$

In the equation above, α controls the strength of the combined regularizer penalty and λ controls the mixture of the two regularizers. At the extremes, if $\lambda = 0$ then the model is ridge regression [24], and if $\lambda = 1$ then the model is lasso. The elastic net aims to provide the best of both regularizers, the feature selection of l_1 and the predictive improvement of l_2. Unlike the lasso, which indiscriminately selects a feature from a group of highly correlated features to represent in the model, the elastic net tends to smooth the coefficient weights and better preserves information in groups of similar features by averaging them in the model.

2.4. Support vector regression

Support vector regression (SVR) is a kernel-based regression model from machine learning that uses only a subset of the training samples, called the support vectors, to define a model [25]. The coefficients of an SVR model are equal to a linear combination of the training samples, and to improve the model's generalizability, the model ignores training samples with a residual error smaller than ϵ. This is accomplished using an ϵ-insensitive loss function. Unlike the squared loss of ordinary least

squares, ϵ-insensitive loss is a piecewise function defined as

$$error\left(y,\hat{y}\right) = max\left(\left|y-\hat{y}\right|-\epsilon,0\right). \quad (3)$$

Using this loss, SVR creates an ϵ-tube around the regression function where samples with small residuals that fall within the ϵ-tube are ignored in the final model.

SVR uses only a linear combination of dot products with the support vectors to predict a sample, so like K-PCR it is able to use the kernel trick, whereby the dot products are replaced by a kernel function. SVR can be either a linear or nonlinear regression method based upon the kernel used. Although SVR was developed in the machine learning community, it is now becoming a popular tool in chemometrics [22,26,17]. For our experiments we trained SVR models with a linear kernel and a second order polynomial kernel. ϵ is the only hyperparameter that is relevant to our problem.

2.5. k-Nearest neighbors

Weighted k-nearest neighbor regression (kNN) is a nonlinear regression model that predicts samples using local interpolation of the k nearest training samples [27,28]. The response of an unknown sample x is predicted as the mean of its surrounding neighbors weighted by their distance from x,

$$f(x) = \left(\sum_{(x_n,y_n)\ \epsilon\ neighbors} \frac{y_n}{\|x-x_n\|}\right)\left(\sum_{(x_n,y_n)\ \epsilon\ neighbors} \frac{1}{\|x-x_n\|}\right)^{-1}. \quad (4)$$

The inverse of the neighbor distance is used in Eq. (4) so that samples close by are weighted more than samples far apart. Unlike all the other models tested, kNN has no model coefficients and requires only that k be tuned. Moreover, kNN requires no model training because all computation is done at the time of prediction. The only hyperparameter that requires tuning is the number of neighbors.

3. Model evaluation and comparisons

3.1. Model selection criteria

To compare the effectiveness of different multivariate analysis techniques, model selection criteria that compare the accuracies of the models are needed to minimize prediction error and assist in choice of models for optimal results. Mean squared error (MSE) is often used as a measure of the overall size of the measurement error [29]. MSE for a model coefficient has the following form:

$$MSE = E\left[\left(X\hat{\beta}-X\beta\right)^2\right] = Var\left(X\hat{\beta}\right) + bias^2\left(X\hat{\beta}\right), \quad (5)$$

where $\text{bias}\left(X\hat{\beta}\right) = E\left(X\hat{\beta}\right)-X\beta$. In these formulae, $\hat{\beta}$ is the calculated model coefficient and β is the true parameter.

Bias is the extent to which the model is overfit using the training set, with two components. Model bias is the difference (error) between the best-fitting linear approximation and the true function. Estimation bias is the difference (error) between the average estimates of the model components $x\hat{\beta}$ and the best-fitting linear approximation. Variance describes how much the model can deviate from the training set when it calculates the model coefficients.

When the MSE is minimized, so are the bias and the variance of the model. Thus, a reasonable bias–variance tradeoff can be achieved for each of the various regression methods. Prediction error results for regression methods are often reported as root mean squared errors of prediction (RMSEP) because these have the same units as the original measurements of sample compositions, which in this project are expressed as wt.% oxides.

3.2. Cross-validation

To compare and quantify the usefulness of various statistical methods to LIBS, we used K-fold cross-validation, which splits the data set into K approximately equal-sized parts, to train the model and tune its hyperparameters (e.g., the number of components used in partial least squares) before it is tested on a held-out dataset. When models are being fit for a sample in K_i, the other K-1 folds (all K_j folds, $i \neq j$) are used to train the model and the K_i fold is used to test the model. In the training set of 70 averaged spectra used here, 10-fold cross-validation has ten folds each containing seven samples. The models are trained using 63 samples and are then tested on the other seven samples.

3.3. Model comparison

The Wilcoxon signed-rank test can be used to compare the RMSEP values that result from two models tested on held-out data to determine whether a difference in their expected values exists. It is appropriate for this application because it does not assume an underlying distribution of the data. The hypotheses are H_0: $\Delta = 0$ versus H_1: $\Delta \neq 0$, where Δ denotes the location difference in the expected value for the populations. The two-sided alternative hypothesis is appropriate in this case because the question of interest is whether any difference between models exists. This comparison is performed for each element.

Three conditions must be met [29] for use of the Wilcoxon signed-ranked test, t. First, observations from each of the populations must be a random sample. Because models can be generated using randomly selected cross-validation folds, this assumption is valid for this application. Second, the observations from the two populations must be mutually independent. Because the models are generated using different algorithms, this is reasonable in this context. Finally, the two populations must be continuous. The RMSEP values can take on any value on the positive real line, so this last assumption is valid.

4. Methods

4.1. Samples and experimental methods

A suite of 100 igneous and meta-igneous rocks was selected for this study; the same samples were used in the work of Tucker et al. [4] and Dyar et al. [8]. These data were acquired at 9 m distance at Los Alamos National Laboratory (LANL) using conditions configured to mimic those on Mars so the results would be applicable to the ChemCam instrument [6]. The entire data set can be requested from the authors. However, we note that an even larger suite of data representing the actual calibration data used on Mars is available on the mission web site of the Planetary Data System at http://pds-geosciences.wustl.edu/missions/msl/chemcam.htm.

Approximately 150 g of each sample was crushed to particle sizes of <45 μm, roughly 10× smaller than the LIBS beam diameter, to minimize sample inhomogeneity and equalize grain size. Splits of these powders were used for X-ray Fluorescence (XRF) analyses of major, minor, and trace element analyses at the University of Massachusetts, Amherst in the laboratory of Michael Rhodes [30]. Samples were prepared as fused La-bearing lithium borate glass disks using a modification of the methods of Norrish and Hutton [31], though each sample was first ignited at 1000 °C for several hours in order to oxidize the iron to Fe^{3+} and remove volatiles. All elements (including Na_2O) were measured simultaneously using a Siemens MRS-400 spectrometer. Intensities were corrected for nonlinear background, inter-element interferences, and variations in mass absorption coefficient, using methods modified from those of Norrish and Chappell [32].

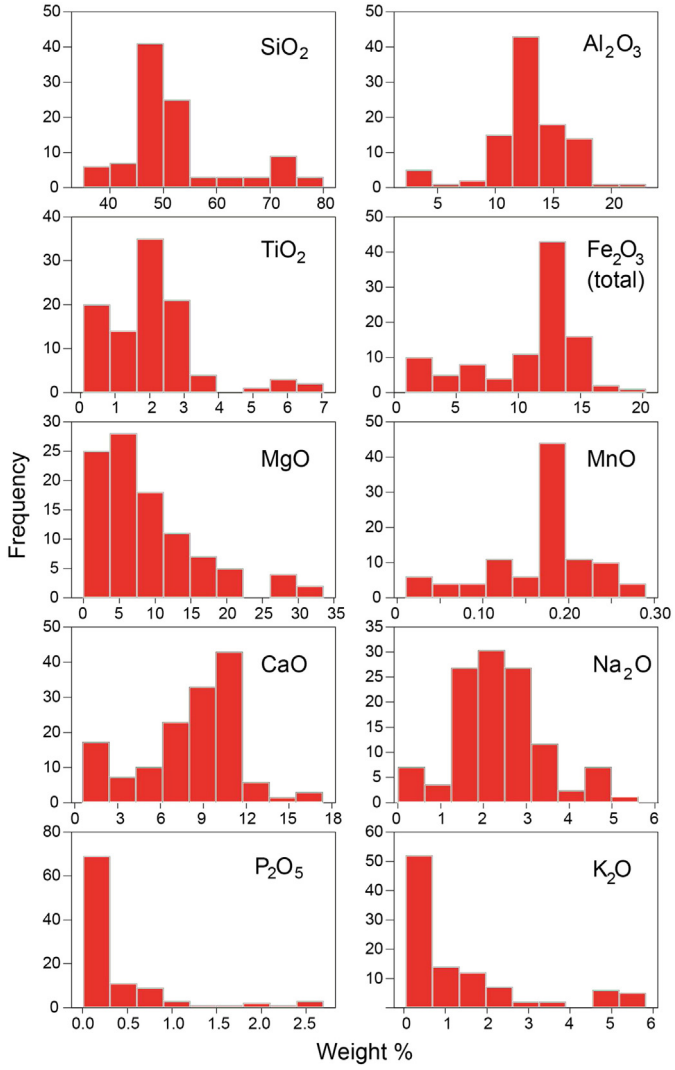

Fig. 1. Compositional ranges of samples used in this study, with elemental abundances expressed as oxide weight percents (wt.%) using the geological convention.

Mass absorption coefficients for elements with shorter characteristic wavelength than the Fe-absorption edge were estimated from the intensity of the Compton radiation of the appropriate X-ray tube [33]. Mass absorption coefficients of elements with longer characteristic radiation than the Fe absorption edge were calculated from the Compton-derived mass absorption coefficients, after allowance was made for Fe and Ti intensities [34]. Reference absolute errors on the XRF methods, expressed as the weight percent of each oxide, are 0.16 for SiO_2, 0.06 for Al_2O_3, 0.005 for TiO_2, 0.03 for Fe_2O_3, 0.04 for MgO, 0.005 for MnO, 0.03 for CaO, 0.09 for Na_2O, 0.003 for P_2O_5, and 0.006 for K_2O [30]. Compositional ranges of the 100 samples are shown in Fig. 1. For LIBS sample preparation, another split of 3 g from each sample was pressed into a pellet in an aluminum cup using 35 tons of pressure.

Because atmospheric pressure exerts known effects on LIBS spectra, samples were placed in a chamber filled with 7 Torr CO_2 to simulate the Mars surface pressure. They were observed from a standoff distance of 9 m using a 1064-nm Nd:YAG laser operating at 17 mJ/pulse; 50 laser pulses were taken at each of five locations per sample. The optical emission from resultant sample plasma was collected using three Ocean Optics HR2000 spectrometers with UV (223–326 nm), VIO (328–471 nm), and VNIR (495–927 nm) wavelength regions analogous to those in the ChemCam instrument (Fig. 2). Spectra of samples from this suite are exactly the same data used in [4] and [12].

4.2. Data pre-processing

Fifty plasmas per location on five locations per sample were averaged before smoothing. The baseline (A–D offset and ambient light background) was measured and subtracted. The bremsstrahlung continuum was modeled and removed by iteratively suppressing peaks by taking the average of each channel with its adjacent channels above and below it and then using the minimum value of those two quantities as the new value of the channel. This is repeated for every channel in every spectrum multiple times until changes become diminishingly small. Spectra were averaged to reduce noise. Channels with near-zero variance were removed prior to analysis [35].

4.3. Hyperparameter tuning

All hyperparameters were tuned using a grid search over the parameter space, and all settings were evaluated using 10-fold cross validation over the training set. A global minimum heuristic, based on the point at which the model MSE is at its smallest value, was used to choose the number of components for the model. A complete listing of the final hyperparameter values can be found in Table 2.

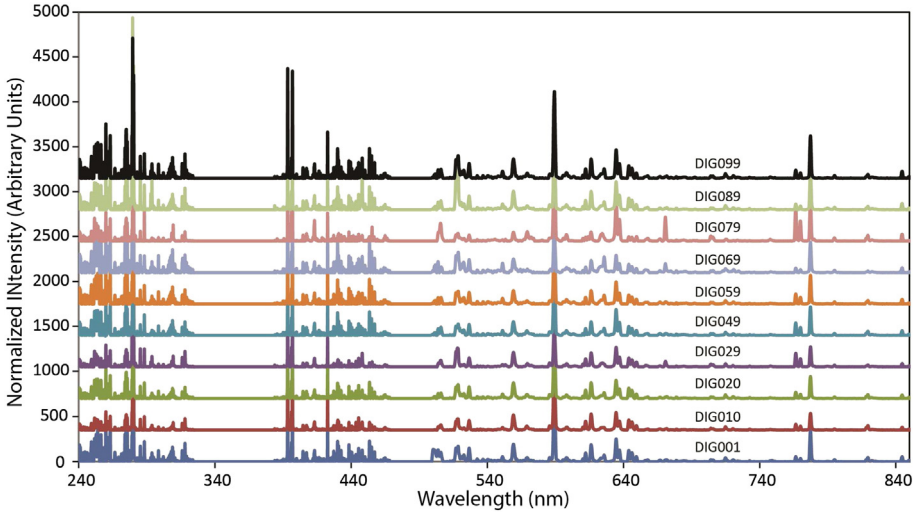

Fig. 2. Selected LIBS spectra of some of the 100 igneous and meta-igneous rocks used in this study.

Table 2
Hyperparameter settings.

		SiO_2	Al_2O_3	TiO_2	Fe_2O_3	MgO	MnO	CaO	K_2O	Na_2O	P_2O_5
PLS-2	Components	15	15	15	15	15	15	15	15	15	15
PLS-1	Components	3	7	12	11	14	3	7	6	6	9
PCR	Components	13	49	20	7	37	6	31	45	38	49
K-PCR	Components	22	46	40	8	39	26	48	48	38	50
Elastic net	α	2.91	1.31	1.01	0.11	0.91	2.51	0.41	0.41	2.91	2.91
	l_1 ratio	0.8	0	0.1	0.5	0	0	0.7	0.1	0.9	0
Lasso	α	2.395	0.73	0.114	0.057	0.12	0.069	0.297	0.038	2.99	0.33
SVR-Lin	ϵ	0	0.83	0.4	0.13	0.45	0.71	0.98	0.24	0.25	0.03
SVR-Py	ϵ	0	0	0	0	0	0.02	0.24	0.38	0.5	0.16
*k*NN	Neighbors	16	2	4	1	2	1	1	11	1	1

4.4. Code

The CRAN R [36] packages Peaks and hyperSpec were used for all data preprocessing. The open-source machine learning Python library Scikit-learn [37] was used to train and test all models except S-LLE. Scikit-learn uses the open-source applications LIBLINEAR [38] and LIBSVM [39] for its linear and polynomial SVR models, respectively.

5. Model comparison results

After tuning all hyperparameters, the models were fit using a 70-sample training set and evaluated using a 30 sample held-out testing set to provide accurate estimates of prediction error. The only requirement for this method is that all samples are from the same population. The models were compared based solely on their predictive performance, as indicated by the predicted residual sum of squares (PRESS), the mean squared error of prediction (MSEP), and the root mean squared error of prediction (RMSEP). Evaluation methods that make parametric assumptions about the data were strictly not used.

Fig. 3 compares the models' element-wise performance using the RMSEP of the nine models over the ten most abundant oxides. The oxides are sorted in descending order according to the average model error, where the scale of the first row is three times the scale of the second row. The five oxides in the first row are also the oxides with the highest abundances in the testing set.

No single model outperformed all others on every oxide. However, linear SVR had the lowest element-wise error, producing significantly better results for total Fe_2O_3 and Na_2O. In contrast, the nonlinear models SVR-Py and *k*NN had the markedly largest prediction errors on all oxides except Al_2O_3 and MnO. A complete listing of summed element-wise prediction errors is given in Table 3.

Fig. 4 uses PRESS to compare the sample-wise performance of the 10 models (~horizontal axis) over the 30 testing samples (receding axis). The vertical axis represents the magnitude of the total error associated with predicting each element in each sample. To better visualize the error, the samples are sorted in descending order according to the average model error. Samples with high PRESS are those for which there is a high mismatch between the predicted and the true compositions. Again, it is clear that the linear support vector results are the most accurate. However, all of the models suffer from some inconsistent performance on the test set with prediction errors much larger for some samples than others, but this was expected because of the non-homogeneity of the testing samples. The worst sample-wise performing models were kernel (polynomial) SVR and *k*NN regression; these models likely over-fitted the training samples. Kernel SVR and *k*NN had the lowest error on more samples than any other two models tested, a total of eight and five, respectively, but both had a cumulative

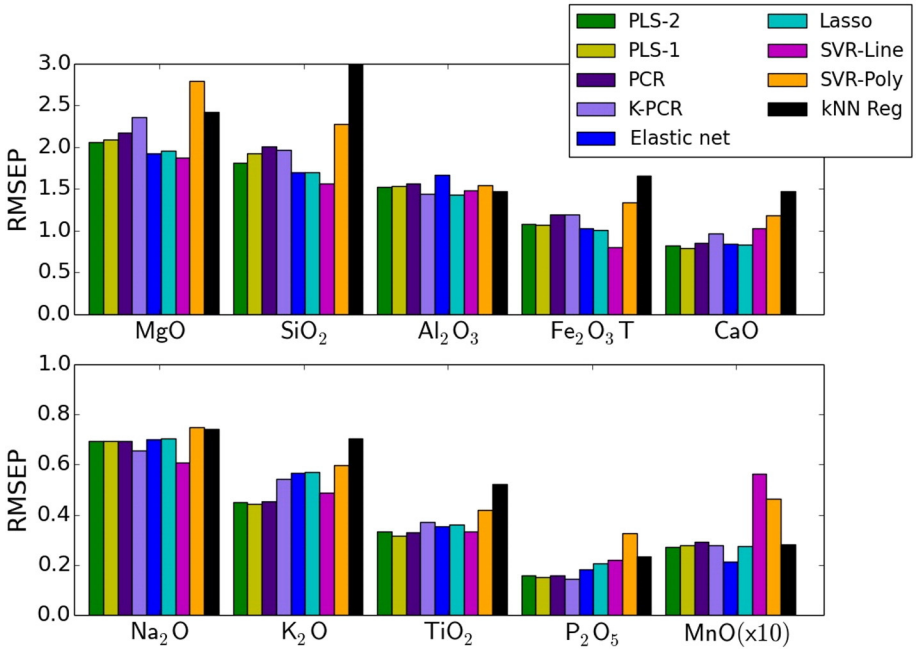

Fig. 3. RMSEP in units of wt.% oxide for the ten major elements of all models evaluated over the held-out testing set. The mean and standard deviations for concentrations of these elements in the 100-sample suite studied are 8.91 ± 7.37 wt.% MgO, 52.53 ± 9.64 wt.% SiO_2, 12.92 ± 3.42 wt.% Al_2O_3, 10.79 ± 4.21 wt.% total Fe_2O_3, 8.12 ± 3.64 wt.% CaO, 2.56 ± 1.14 wt.% Na_2O, 1.35 ± 1.59 wt.% K_2O, 2.04 ± 1.40 wt.% TiO_2, 0.43 ± 0.56 wt.% P_2O_5 and 0.17 ± 0.06 wt.% MnO.

Table 3
Model results.

	PLS-2	PLS-1	PCR	K-PCR	Elastic net	Lasso	SVR-Lin	SVR-Py	*k*NN
MSEP	12.55	13.07	14.22	14.79	12.14	11.51	10.63	19.79	23.40
RMSEP	8.97	9.07	9.47	9.68	8.99	8.81	8.46	11.28	12.27

error almost double the third worst performing model. This strongly suggests that these models are fitting a small portion of the training samples too tightly, namely the samples that are very similar in composition to these low error test samples. This overfitting prevents the models from predicting the majority of the test samples well.

The Wilcoxon signed-rank test [40] was used to determine whether the evaluated models produced significantly different results. Like all of our evaluation methods, this hypothesis test makes no parametric assumptions about the data. PRESS over the testing samples was used to compare each pair of models, and the results of the full pairwise model comparison are given in Table 4. Given a p value of <0.05, there are very few significant differences among models. However, the value of 0.033 for PCR vs. PLS-2 indicates that these models are not indistinguishable; their predicted compositions per sample are very different. The most distinctive model was *k*NN, which was expected because this model differs from the others by having no feature coefficients (nonparametric) and using solely a weighted interpolation of the k nearest samples for training. In fact, the construction of this algorithm is unique among those studied here because it is parameter-free. Apparently for this reason, *k*NN distinguished itself as the overall worst performing model.

6. Discussion

6.1. Sparse vs. non-sparse models

Results from sparse models (lasso and elastic net) versus models that use all channels of the data (Fig. 3) vary by oxide. For MgO, SiO_2, Fe_2O_3, CaO, and MnO, the sparse models outperform all the others except for linear SVR, requiring only between 20 and 50 channels (lines) per oxide. For Na_2O, K_2O, TiO_2, and P_2O_5, the sparse methods produce inferior results. This result may be explained by factors relating to the nature of the LIBS lines themselves.

The ability to make accurate predictions from sparse techniques that use only a few lines for any given oxide is dependent on the presence of strong lines associated with the oxide(s) being predicted. To show this effect, lines with intensities >300 that fell in the studied wavelength range for each of the ten oxides of interest were compiled from the NIST Atomic Spectra Database[1] using. Note that NIST database intensities are not standardized, so this is a relatively imperfect tool, but the results are quite informative (Fig. 5). The strongest emission lines for Na, K, Ti, P are all roughly an order of magnitude lower than those from the other six oxides in the energy range covered by the spectrometer used in these experiments. The lower transition probabilities result in weaker emission lines that are harder for the sparse techniques to separate from noise, whereas strong emission lines allow sparse models to eliminate much of the noise that would otherwise impede dense models.

The number of observable emission lines and the degree to which they are overlapped by lines from other oxides within the studied UV-VNIR wavelength range may also influence the functionality of the sparse methods. To pursue this hypothesis, the possibility of interference from nearby peaks for Na, K, Ti, P, Mg, Si, Al, Fe, and Ca was investigated using the same strong lines with intensities >300 as for the transition probabilities. Emissions within about 0.5 nm of those emission peaks were scrutinized, including any oxides that had a relatively high intensity (at least 20–30% of the intensity of the strong emission peak). Although there were some minor trace metals that might be abundant enough to cause interference, etc., there was no clear difference between the two sets of elements (Na, K, Ti, P vs. Mg, Si, Al, Fe, Ca).

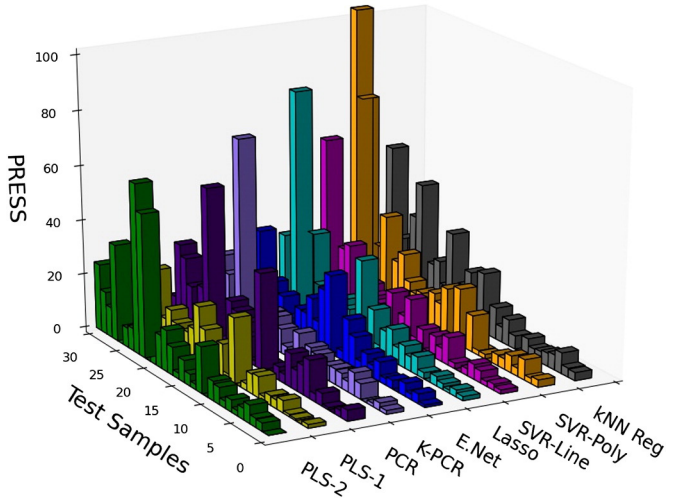

Fig. 4. The predicted residual sum of squares (PRESS) of all models (vertical axis) evaluated for each of the 30 held-out testing samples (receding axis) and all ten 10 models (horizontal axis). Color scheme is the same as in Fig. 3. The vertical axis represents the magnitude of the total error associated with predicting each oxide in each sample, so high bars indicate high errors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6.2. Linear vs. nonlinear models

In terms of predictive performance indicated by the predicted residual sum of squares (PRESS), the three worst-performing models are kernel PCR, polynomial SVR, and *k*NN regression (Fig. 4). This result suggests that such nonlinear methods tend to overfit the data and predict less accurately in this application. In contrast, the linear methods proved to be more generalizable with better predictive performance on a test set. This outcome is attributed to the high dimensionality of the data (6144 channels) relative to the small number of samples studied.

Theory from statistics shows that mean squared error can be decomposed into two terms that trade off: error due to the bias of the model (i.e., the distance of each prediction from the average of all predictions) and the variance error (the maximum distance between all possible models). As model complexity grows, variance error grows relative to bias error. So the high variance of the nonlinear models evaluated proved to be larger than the inherent high bias of the linear models.

6.3. Importance of training set selection for geological applications

Data in Fig. 3 can further be considered in terms of the abundances of each oxide, given in the figure caption and represented in Fig. 1. The average RMSEP represents a high percentage of the actual observed concentration for many oxides: 51% of the total concentration of P_2O_5, 41% of K_2O, 26% of Na_2O, and 24% of the average MgO. Ranking the oxides in terms of the percentage of relative error gives SiO_2 (lowest relative error at 4%) < Fe_2O_3 < CaO ≈ MnO ≈ Al_2O_3 < TiO_2 < MgO < Na_2O < K_2O < P_2O_5 (highest relative error).

Some of this observed variation from oxide to oxide may be due to the distribution and abundance (magnitude) of oxides in the data set overall. Inspection of Fig. 1 shows that the oxides are not uniformly distributed over the wt.% ranges in geological samples. Thus, any random procedure that draws a subset of samples may have very

[1] http://www.nist.gov/pml/data/asd.cfm.

Table 4
Wilcoxon test results.

	PLS-2	PLS-1	PCR	K-PCR	Elastic net	Lasso	SVR-Linear	SVR-Poly	*k*NN Reg
PLS-2		0.136	**0.033**	0.141	0.975	0.943	0.280	0.329	**0.022**
PLS-1	0.136		0.428	0.673	0.586	0.614	0.159	0.544	0.063
PCR	**0.033**	0.428		0.734	0.393	0.393	0.094	0.719	0.090
K-PCR	0.141	0.673	0.734		0.658	0.558	0.171	0.781	**0.030**
Elastic net	0.975	0.586	0.393	0.658		0.673	0.254	0.465	**0.021**
Lasso	0.943	0.614	0.393	0.558	0.673		0.704	0.465	**0.013**
SVR-Lin	0.280	0.159	0.094	0.171	0.254	0.704		0.393	**0.010**
SVR-Py	0.329	0.544	0.719	0.781	0.465	0.465	0.393		0.185
***k*NN**	**0.022**	0.063	0.090	**0.030**	**0.021**	**0.013**	**0.010**	0.185	

The bold values are those with a $p <= 0.05$, indicating a statistically significant value.

different summary statistics than the entire population. As an extreme example, consider that a majority of the samples contain $P_2O_5 < 0.5$ wt.%. So if a training set happens to draw only from that limited range of concentrations, then predicting samples in the test set with $P_2O_5 > 2.5$ wt.% is an extreme extrapolation. In such a situation, the test RMSEP values for P_2O_5 will be large. Conversely, the distribution of SiO_2 in our 100 samples is slightly more uniform, as shown in Fig. 1. Thus, it would be unlikely to randomly draw a test set that does not represent the full range of SiO_2 concentrations. If a training set does not contain a uniform distribution of each oxide in its geological composition space, then our results show that test set results will be greatly influenced. This result demonstrates the importance of careful construction of training sets for optimal accuracy.

6.4. Model selection for optimal performance in geological samples

PLS has become the conventional form of analysis for LIBS geological data in recent years. As such, researchers in the field are more familiar with the implementation and output of PLS. From an interpretation perspective, it does have two drawbacks. First, the dimensionality reduction algorithm does not have a closed form expression, so it is difficult to discern exactly what is happening during this process. This has been a subject of debate [41,42]. Second, because the model coefficients are made by taking linear combinations of channels from the original data set, these coefficients can only have implicit physical meaning, though they can be mapped onto emission lines using loading plots [43]. This matchup is not as convenient as the direct mapping from model coefficients to emission lines inherent in sparse methods like the lasso and elastic net, though that practice in itself may also be suspect, especially when there are many collinear features present in the data [13–15].

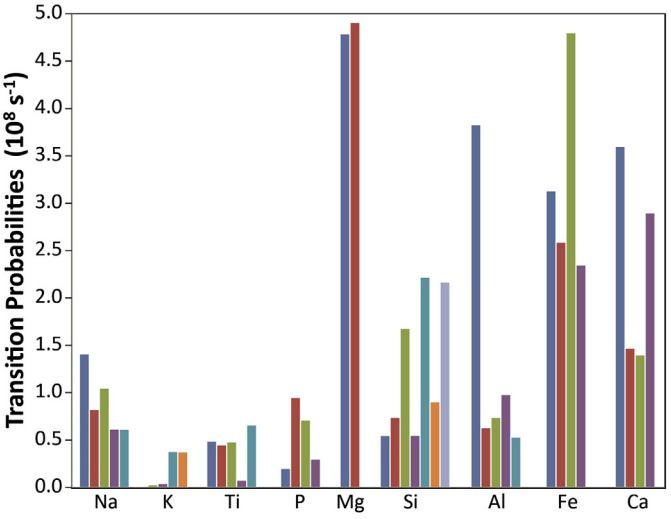

Fig. 5. Transition probabilities of emission lines with intensities >300 (ambiguous units using an arbitrary cut-off, as compiled from the NIST Atomic Spectra Database) that fell in the studied wavelength range for the nine most abundant oxides studied here for geological samples. Colors indicate the wavelength of the lines in order from lowest to highest. Emission lines for Mg, Si, Al, Fe, and Ca have probabilities significantly higher than those for Na, K, Ti, and P. Higher transition probabilities result in greater line strengths that likely make it easier for sparse models to outperform those that use the entire spectrum. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Model parsimony and explicit physical meaning of model coefficients may eventually enhance interpretability of sparser models, making them advantageous for future LIBS studies so that geochemistry can be better understood. Future studies may choose to employ sparse methods in situations where the data are noisy (i.e., when collected under difficult environmental conditions) or where interpretability is important. This suggests that alternative techniques that remove noisy features might be useful. Dimensionality reduction is a field in machine learning that focuses on simplified data representation by eliminating dimensions that do not inform the model. Such models show promise and are the subject of ongoing work [44].

However, among the methods used in the current study, there are also better (more accurate) alternatives to PLS that still employ the entire spectrum. For example, linear SVR can produce superior results because only a subset of the training set is used to construct the most generalizable model possible. This suggests that statistical bootstrapping methods, like bagging, or generating multiple training sets by uniformly sampling with replacement from the original training set [45], could be used as meta-algorithms to improve predictive performance. Future studies may choose to employ sparse methods in situations where the element being predicted has a small number of lines with large transition probabilities or where interpretability is important. The strong performance of the sparse methods in this study at predicting major elements suggests that the use of dimensionality-reduction techniques as a preprocessing step may improve the performance of the linear models.

Finally, the differences in model performance on an oxide basis might be taken to suggest that the model choice could be customized depending on the oxide of interest. The best-performing models as seen in Table 3 were SVR-Lin for SiO_2, MgO, Fe_2O_3total, and Na_2O, lasso for Al_2O_3, elastic net for MnO, and PLS-1 for CaO, TiO_2, and K_2O. However, none of these differences is statistically significant (Table 4) in our small test set of only 30 samples, except for the case of *k*NN (which is clearly worse). With a bigger test set, a greater difference in predictive performance might be expected, in which case using an ensemble of regression models might yield superior results. This supposition clearly merits additional work.

7. Conclusions

Our results show that the linear models evaluated better predicted the major elements of the out of sample rocks than the nonlinear models evaluated. Models that used a subset of the features or a subset of the training data proved to be the most accurate models for predicting the five major oxides. However, no superior algorithm has emerged in general, the methods should be used in tandem to

exploit their individual qualities, which include familiarity, sparsity, interpretability, and computational speed. This suggests that a combination of models could be used together to yield a lower total error of prediction, depending on the requirements of the user.

The extent to which these conclusions can be generalized surely depends on the size and composition of the data set, though this conclusion remains to be tested due to lack of availability of appropriate data. This data set is one of the largest, best-characterized calibration suites currently available for geological samples. The fact that it is deliberately restricted to igneous and meta-igneous rocks means that matrix effects caused by wide compositional variations are mitigated, allowing our work to be focused on model comparisons under ideal conditions. This study thus provides a useful benchmark against which other models that include more diverse samples can be compared. To date there are no larger data sets available for this analysis that were collected using similar rock types and identical instrumental conditions. With smaller data sets, certain test and training sets will not produce reasonable RMSEP values, but other randomly-generated combinations will. Results are expected to change as other data sets with even greater compositional variation are used. Thus, this work lays a foundation for future work as larger data sets are developed. It is likely that the future studies will further clarify the relative utility of the methods studied here.

In particular, as noted by Brown and Green [15], future experimental design may allow quantitative assessments of sensitivity and selectivity as figures of merit for LIBS analyses. However, acquiring such data may be difficult given the complexity of elemental interactions in LIBS plasmas caused by chemical matrix effects and the limitless combinations of elements in geological samples. As larger data sets become available, it is likely that an array of methods may be needed to extract optimal chemical analyses from LIBS data.

Acknowledgments

We are grateful for support from NSF grants CHE-1306133 and CHE-1307179 and NASA grants NNG06GH35G and NNX09AL21G, and for student support from the Massachusetts Space Grant Consortium. We thank Michael Vollinger and Michael Rhodes for contributing analyzed samples and good advice to this project.

References

[1] R.C. Wiens, S. Maurice, B. Barraclough, M. Saccoccio, W.C. Barkley, J.F. Bell III, S. Bender, J. Bernardin, D. Blaney, J. Blank, M. Bouye, N. Bridges, P. Cais, R.C. Clanton, B. Clark, S. Clegg, A. Cousin, D. Cremers, A. Cros, L. DeFlores, D. Delapp, R. Dingler, C. D'Uston, M.D. Dyar, T. Elliott, D. Enemark, C. Fabre, M. Flores, O. Forni, O. Gasnault, T. Hale, C. Hays, K. Herkenhoff, E. Kan, L. Kirkland, D. Kouach, D. Landis, Y. Langevin, N. Lanza, F. LaRocca, J. Lasue, J. Latino, D. Limonadi, C. Lindensmith, C. Little, N. Mangold, G. Manhes, P. Mauchien, C. McKay, E. Miller, J. Mooney, R.V. Morris, L. Morrison, T. Nelson, H. Newsom, A. Ollila, M. Ott, L. Pares, R. Perez, F. Poitrasson, C. Provost, J.W. Reiter, T. Roberts, F. Romero, V. Sautter, S. Salazar, J.J. Simmonds, R. Stiglich, S. Storms, N. Streibig, J.J. Thocaven, T. Trujillo, M. Ulibarri, D. Vaniman, N. Warner, R. Waterbury, R. Whitaker, J. Witt, B. Wong-Swanson, The ChemCam instruments on the Mars Science Laboratory (MSL) rover: body unit and combined system performance, Space Sci. Rev. 170 (2012) 167–227.

[2] S. Maurice, R.C. Wiens, M. Saccoccio, B. Barraclough, O. Gasnault, O. Forni, N. Mangold, D. Baratoux, S. Bender, G. Berger, J. Bernardin, M. Berthé, N. Bridges, D. Blaney, M. Bouyé, P. Cais, B. Clark, S. Clegg, A. Cousin, D. Cremers, A. Cros, L. DeFlores, C. Derycke, B. Dingler, G. Dromart, B. Dubois, M. Dupieux, E. Durand, L. d'Uston, C. Fabre, B. Faure, A. Gaboriaud, T. Gharsa, K. Herkenhoff, E. Kan, L. Kirkland, D. Kouach, J.L. Lacour, Y. Langevin, J. Lasue, S. Le Mouélic, M. Lescure, E. Lewin, D. Limonadi, G. Manhes, P. Mauchien, C. McKay, P.Y. Meslin, Y. Michel, E. Miller, H.E. Newsom, G. Orttner, A. Paillet, L. Pares, Y. Parot, R. Perez, P. Pinet, F. Poitrasson, B. Quertier, B. Sallé, C. Sotin, V. Sautter, H. Seran, J.J. Simmonds, J.B. Sirven, R. Stiglich, N. Streibig, J.J. Thocaven, M. Toplis, D. Vaniman, The ChemCam instruments on the Mars Science Laboratory (MSL) rover: science objectives and mast unit, Space Sci. Rev. 170 (2012) 95–166.

[3] S.M. Clegg, J. Lasue, O. Forni, S. Bender, R.C. Wiens, S. Maurice, B. Barraclough, D. Blaney, A. Cousin, L. DeFlores, D. Delapp, M.D. Dyar, C. Fabre, O. Gasnault, N. Lanza, R.V. Morris, T. Nelson, H. Newsom, A. Ollila, R. Perez, V. Sautter, D.T. Vaniman, ChemCam flight model calibration, Lunar Planet. Sci., Woodlands, TX, 2012, Abstract 2076, 2012.

[4] J.M. Tucker, M.D. Dyar, M.W. Schaefer, S.M. Clegg, R.C. Wiens, Optimization of laser-induced breakdown spectroscopy for rapid geochemical analysis, Chem. Geol. 277 (2010) 137–148.

[5] J.B. Sirven, B. Sallé, P. Mauchien, J.L. Lacour, S. Maurice, G. Manhès, Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods, J. Anal. At. Spectrom. 22 (2007) 1471–1480.

[6] S.M. Clegg, E. Skulte, M.D. Dyar, J.E. Barefield, R.C. Wiens, Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and related techniques, Spectrochim. Acta Part B 88 (2009) 79–88.

[7] R.M. Anderson, R.V. Morris, S.M. Clegg, J.F. Bell, R.C. Wiens, S.D. Humphries, S.A. Mertzman, T.G. Graff, R. McInroy, The influence of multivariate analysis methods and target analysis of rocks using laser induced breakdown spectroscopy, Icarus 215 (2011) 608–627.

[8] M.D. Dyar, M.L. Carmosino, J.M. Tucker, E.A. Brown, S.M. Clegg, R.C. Wiens, J.E. Barefield, J.S. Delaney, G.M. Ashley, S.G. Driese, Remote laser-induced breakdown spectroscopy analysis of East African Rift sedimentary samples under Mars conditions, Chem. Geol. 294–295 (2012) 135–151.

[9] L. Liang, T.L. Zhang, K. Wang, H.S. Tang, X.F. Yang, X.Q. Zhu, Y.X. Duan, H. Li, Classification of steel materials by laser-induced breakdown spectroscopy coupled with support vector machines, Appl. Opt. 53 (2014) 544–552.

[10] X.W. Li, Z. Wang, S.L. Lui, Y.T. Fu, Z. Li, J.M. Liu, W.D. Ni, A partial least squares based spectrum normalization method for uncertainty reduction for laser-induced breakdown spectroscopy measurements, Spectrochim. Acta Part B 88 (2013) 180–185.

[11] R.B. Anderson, J.F. Bell, R.C. Wiens, R.V. Morris, S.M. Clegg, Clustering and training set selection methods for improving the accuracy of quantitative laser induced breakdown spectroscopy, Spectrochim. Acta Part B 70 (2012) 24–32.

[12] M.D. Dyar, M.L. Carmosino, E.A. Breves, M.V. Ozanne, S.M. Clegg, R.C. Wiens, Comparison of partial least squares and lasso regression techniques as applied to laser-induced breakdown spectroscopy of geological samples, Spectrochim. Acta Part B 70 (2012) 51–67.

[13] M.B. Seasholtz, B.R. Kowalski, Qualitative information from multivariate calibration models, Appl. Spectrosc. 44 (2009) 1337–1348.

[14] O.M. Kvalheim, T.V. Karstang, Interpretation of latent-variable regression models, Chemometr. Intell. Lab. Des. 7 (1989) 39–51.

[15] C.D. Brown, R.L. Green, Critical factors limiting the interpretation of regression vectors in multivariate calibration, Trends Anal. Chem. 28 (2009) 506–519.

[16] J.A. Wegelin, A survey of partial least squares (pls) methods, with emphasis on the two-block case, Technical Report, University of Washington, USA, 2000.

[17] O. Erdas, E. Buyukbingol, F.N. Alpaslan, A. Adejare, Modeling and predicting binding affinity of phencyclidine-like compounds using machine learning methods, J. Chemometr. 24 (2010) 1–13.

[18] J.H. Kalivas, Interrelationships of multivariate regression methods using eigenvector basis sets, J. Chemometr. 13 (1999) 111–1329.

[19] I.T. Jolliffe, Principal Component Analysis, 2nd edition Springer, 2002.

[20] B. Scholkopf, A. Smola, K.-R. Müller, Kernel principal component analysis, Advances in Kernel Methods — Support Vector Learning, MIT Press, 1999, pp. 327–352.

[21] B. Scholkopf, A. Smola, Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, Cambridge, MA, USA, 2001.

[22] E. Andries, Sparse models by iteratively reweighted feature scaling: a framework for wavelength and sample selection, J. Chemometr. 27 (2013) 50–62.

[23] P. Filzmoser, M. Gschwandtner, V. Todorov, Review of sparse methods in regression and classification with application to chemometrics, J. Chemometr. 26 (2012) 42–51.

[24] A.E. Hoerl, R.W. Kennard, Ridge regression — applications to non-orthogonal problems, Technometrics 12 (1970) 69–76.

[25] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, V. Vapnik, Support vector regression machines, Advances in Neural Information Processing Systems, 9, MIT Press, 1997, pp. 155–161.

[26] C. Butnariu, C. Lisa, F. Leon, S. Curteanu, Prediction of liquid–crystalline property using support vector machine classification, J. Chemometr. 27 (2013) 179–188.

[27] J. Gertheissa, G. Tutza, Variable scaling and nearest neighbor methods, J. Chemometr. 23 (2009) 149–151.

[28] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, 2nd ed. Springer Verlag, New York, 2009.

[29] M. Hollander, D.A. Wolfe, Nonparametric Statistical Methods, 2nd ed. John Wiley & Sons, Inc., New York, 1999.

[30] J.M. Rhodes, M.J. Vollinger, Composition of basaltic lavas sampled by phase-2 of the Hawaii Scientific Drilling Project: geochemical stratigraphy and magma types, Geochem. Geophys. Geosyst. 5 (2004) Q03G13.

[31] K. Norrish, J.T. Hutton, An accurate X-ray spectrographic method for the analysis of a wide range of geological samples, Geochim. Cosmochim. Acta 33 (1969) 431–453.

[32] K. Norrish, B.W. Chappell, X-ray fluorescent spectrography, in: J. Zussman (Ed.), Physical Methods in Determinative Mineralogy, Academic, San Diego, Calif., 1967, pp. 161–214.

[33] R.C. Reynolds, Estimation of mass absorption coefficients by Compton scattering: improvements and extension of the method, Am. Mineral. 52 (1967) 1493–1502.

[34] D. Walker, Behavior of mass absorption coefficients near absorption edges: Reynolds' method revisited, Am. Mineral. 58 (1973) 1069–1072.

[35] M. Kuhn, Building predictive models in r using the caret package, J. Stat. Softw. 28 (2008) 1–26.

[36] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2013.

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[38] R. Fan, K. Chang, C. Hsieh, X. Wang, C. Lin, LIBLINEAR: a library for large linear classification, J. Mach. Learn. Res. 9 (2008) 1871–1874.
[39] C. Chang, C. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (27) (2011) 1–27.
[40] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bull. 1 (1945) 80–83.
[41] R. Rosipal, N. Krämer, Overview and recent advances in partial least squares, LNCS 3940 (2006) 34–51.
[42] C. Goutis, Partial least squares algorithm yields shrinkage estimators, Ann. Stat. 24 (1996) 816–824.
[43] B.H. Mevik, R. Wehrens, K.H. Liland KH, pls: partial least squares and principal component regression, R Package Version 2.3-02011.
[44] T. Boucher, M.D. Dyar, M.L. Carmosino, S. Mahadevan, S. Clegg, R. Wiens, Manifold regression of LIBS data from geological samples for application to ChemCam on Mars, Sci-X 2013, Milwaukee, Abstract #242013.
[45] L. Breiman, Bagging predictors, J. Mach. Learn. 24 (1996) 123–140.