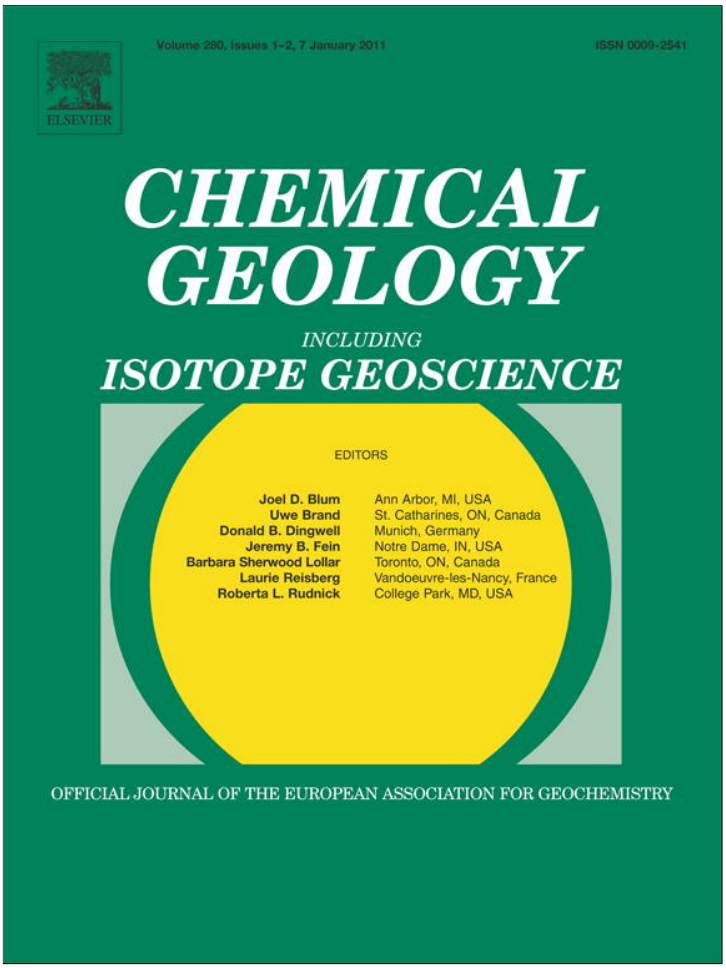

(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Chemical Geology 294-295 (2012) 135–151

Contents lists available at SciVerse ScienceDirect

Chemical Geology

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

Research paper

Remote laser-induced breakdown spectroscopy analysis of East African Rift sedimentary samples under Mars conditions

M.D. Dyar [a,*], M.L. Carmosino [b], J.M. Tucker [a], E.A. Brown [a], S.M. Clegg [c], R.C. Wiens [c], J.E. Barefield [c], J.S. Delaney [d], G.M. Ashley [d], S.G. Driese [e]

[a] Dept. of Astronomy, Mount Holyoke College, 50 College St., South Hadley, MA 01075, United States
[b] Dept. of Computer Science, Hampshire College, 893 West St., Amherst, MA 01002, United States
[c] Los Alamos National Laboratory, P.O. Box 1663, MS J565, Los Alamos, NM 87545, United States
[d] Dept. of Earth and Planetary Sciences, Rutgers University, 610 Taylor Rd., Piscataway, NJ 08854, United States
[e] Dept. of Geology, Baylor University, One Bear Place #97354, Waco, TX 76798, United States

ARTICLE INFO

Article history:
Received 22 March 2011
Received in revised form 1 October 2011
Accepted 17 November 2011
Available online 29 November 2011

Editor: D.B. Dingwell

Keywords:
LIBS
Quantitative analysis
Elemental analysis
Bulk analysis
Mars
East African Rift

ABSTRACT

Laser-induced breakdown spectroscopy will be used by the ChemCam instrument on the Curiosity rover to obtain chemical analyses of the martian surface. Surficial and deeper hydrothermal processes on Mars have produced a diverse family of chemical and clastic sedimentary lithologies from primary igneous rocks through physical and chemical transport, deposition, and diagenesis. This study uses 16 samples from the East African Rift (EAR) as martian analogues to assess use of LIBS to evaluate weathering reactions in sedimentary rocks. Data were acquired at 9 m distance with samples under conditions designed to simulate Mars. Use of external validation, in which a few sedimentary samples are added to the validation set, coupled with choice of the first local minimum in the root mean square value in all the components of the validation model, provides optimal results in this data set. Accuracy is measured using root mean square error predictions for major elements in the sedimentary rocks, as expressed in wt.% oxides. Even lower errors can be obtained by using a more focused training set. These results are sufficiently accurate to usefully characterize the four competing kinds of chemical weathering reactions in paleosols (hydrolysis, oxidation, hydration, and salinization).

1. Introduction

ChemCam, a remote sensing instrument package including a laser-induced breakdown spectrometer (LIBS) and remote micro-imager (RMI), will provide geochemical analyses and context imaging as part of the Mars Science Laboratory (MSL) Curiosity rover payload launched in 2011 (Maurice et al., 2005; Wiens et al., 2005). This will be the first deployment of LIBS for remote extra-terrestrial quantitative analysis. ChemCam will provide Curiosity with the innovative reconnaissance capability of near-complete chemical analyses on spot sizes ranging from ~175 μm to 500 μm at standoff distances of 1.5 m to 7 m. LIBS was also on the payload of the Surface and Atmosphere Geochemical Explorer (SAGE) mission to Venus that was selected as a finalist for the next New Frontiers Mission Program of space ventures to celestial bodies in our solar system. Several other countries are also considering or planning to use LIBS on upcoming missions. Thus LIBS is rapidly becoming part of the geochemists' toolkit for planetary exploration.

The LIBS technique builds on the fundamentals of atomic emission and inductively-coupled spectroscopy that are already familiar to most geochemists. In LIBS, a laser pulse is focused onto a sample to create plasma from which the optical emissions are recorded. The major emitting species of the plasma are neutral atoms and ions in the first and second ionization states of the elements comprising the samples. LIBS spectra from the near-UV to near-IR of geological samples typically consist of dozens to hundreds of atomic emission lines. The basis for qualitative and quantitative chemical analysis is the dependence of the peak height and area on the abundance of that element in the sample.

The simplicity and versatility of the LIBS technique have allowed it to be applied to a wide variety of materials and substances in commercial applications. However, because other robust microanalytical methods for elemental analysis are available to geologists, there has been only limited interest in geological applications of the LIBS technique (e.g., reviews by Harmon et al., 2005 and Pasquini et al., 2007). Until very recently, few published mineral (McMillan et al., 2006, 2007; McManus et al., 2008; Alvey et al., 2010) or rock (Remus et al., 2010) LIBS spectra existed, and the data were often acquired

* Corresponding author at: Department of Astronomy, Mount Holyoke College, 50 College St., South Hadley, MA 01075. Tel.: +1 413 538 3073 (office); fax: +1 413 538 2357.
E-mail address: mdyar@mtholyoke.edu (M.D. Dyar).

0009-2541/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemgeo.2011.11.019

under disparate conditions that precluded direct comparisons between spectra. Systematic calibration curves, matrix corrections, and procedures for spectral processing in complex mineral and rock samples did not yet exist. This situation is now changing as a result of the planetary missions, and because of the potential for using LIBS in real-time identification and discrimination of geological materials in the field (Harmon et al., 2009), as well as in non-geological applications (e.g. Ctvrtnickova et al., 2009; Colao et al., 2010; Fortes and Laserna, 2010), where its remote capabilities are advantageous.

LIBS applications have also benefitted from improvements in statistical analyses in recent years. Advanced statistical analyses are more necessary for geochemical analyses than for other LIBS applications, given the varieties of compositions and surface conditions of unprepared geological samples. Sirven et al. (2006) was among the first to apply multivariate analyses to LIBS spectra of geological samples, investigating the accuracies of Cr analyses in 30 soil and 30 kaolinite samples. Their study also compared partial least squares (PLS) regression with predictions from neural network analyses. It was followed (Sirven et al., 2007) with a study of classification of different rock types by multivariate analysis of LIBS data. Clegg et al. (2009) applied multivariate analysis techniques to analyze the LIBS spectra of 18 disparate igneous and highly-metamorphosed rock samples. Partial Least Squares (PLS) analysis was used to generate a calibration model from which compositions of samples extracted from that dataset could be analyzed. Techniques of classification and discrimination of geological materials by PLS-DA (partial least squares discriminant analysis), PCA (principal components analysis), and SIMCA (soft independent modeling of class analogy) were also investigated by Clegg et al. (2009), Gottfried et al. (2009), and Harmon et al. (2009). Tucker et al. (2010) were the first to use a large (100 sample) dataset of LIBS spectra of rocks to predict major element compositions in unknown samples.

As with most analytical techniques, LIBS works best when the standards used for calibration have compositions similar to those of the unknowns to be studied. For the multitude of possible geological compositions that might be encountered on Mars, this presents an analytical challenge. The useful predictive models developed by Tucker et al. (2010) were successful in part because they used igneous rocks to predict the compositions of other igneous rocks. However, little is known about the effects of bulk composition, mineralogy, and crystallinity (presence of glass vs. crystals of varying sizes) on the predictive abilities of LIBS. For example, can a calibration set built from a wide range of igneous rock compositions effectively predict the compositions of sedimentary rocks derived from igneous precursors?

These issues are of paramount importance given the preponderance of sedimentary rocks that may be present on Mars. The martian surface has abundant sedimentary materials, as shown, for example, by orbital imagery, especially of the Arabia region (Edgett and Malin, 2002; Bibring et al., 2005, 2006; Gendrin et al., 2005), and the striking evaporate deposits at Meridiani (e.g., Squyres et al., 2004; McLennan et al., 2005; Squyres and Knoll, 2005; Morris et al., 2006). Critical questions include the origin, extent, and duration of large bodies of water responsible for sediments in regions such as Arabia and Meridiani. Studies of martian meteorites (e.g., Gooding et al., 1991; Gooding, 1992), MER rover observations, and geochemical calculations (e.g., Griffith and Shock, 1997) make it clear that surface and deeper hydrothermal processes have produced a diverse family of chemical and clastic sedimentary lithologies through physical and chemical transport, deposition, and diagenesis. This has undoubtedly led to complex sedimentary mineralogy and chemical compositions.

So to enable LIBS data to best predict compositions of those wide-ranging martian lithologies, this paper extends the work of Tucker et al. (2010) by exploring various methods to predict the major element compositions of a genetically- and texturally-distinct set of sedimentary rocks derived from igneous precursors. For the current study, we selected a suite of sedimentary samples from the East African Rift (EAR) (Ashley and Driese, 2000) as good martian analogs. The master training set (sometimes called a calibration suite) with which our statistical models can be built includes the 100 igneous samples used by Tucker et al. (2010), a suite of phyllosilicate standards (many of which are logical end-products of alteration of igneous rocks), and the sedimentary rock spectra themselves. Our hypothesis is that the sedimentary rocks might be compositionally somewhat intermediate between the parent extrusive rocks and their ultimate breakdown assemblages, though interactions with water would obviously result in cation exchange into and out of the sedimentary rocks.

We here explore using variations of our training set to understand optimal model parameters and choice of training set, and to then determine the best possible predictions of sedimentary rock compositions using the data in hand. These data are then used to evaluate the extent of chemical weathering of the sedimentary rocks studied. This study thus selects an appropriate calibration (training) set to predict rock compositions, and begins to address the potential effects of bulk composition, mineralogy and paragenesis on LIBS spectra, which are fundamental issues in the application of remote LIBS spectroscopy to geological studies. The goal of this paper is thus to make quantitative geochemical analyses on bulk samples of sedimentary samples, in order to assess their ability to understand chemical weathering processes, as well as to provide the basis for and roadmap to expanding our work into a larger range of rock types and samples.

2. Background

Laser-induced breakdown spectroscopy (LIBS) is a type of atomic emission spectroscopy that is closely related to the more commonly-used geological technique of inductively-coupled plasma atomic emission spectroscopy (ICP-AES). In both techniques, as in any type of atomic spectroscopy, the analysis relies on quantized valence-electron transitions that occur in the UV, visible, and near-IR regions of the energy spectrum. To produce narrow emission lines with diagnostic energies, samples must be atomized, with no residual bonding remaining. After being excited, electrons decay from high-energy states to lower states, emitting photons with wavelengths that are characteristic to each atom or ion.

In traditional geological applications such as atomic absorption (AA) or ICP-AES, samples are commonly dissolved in solutions. LIBS differs from these methods in that a pulsed high power laser is used both to ablate and excite atoms without sample preparation. Emission from the ions and atoms in the plasma is collected by a fiber optic system and analyzed by a spectrometer and detector. The ChemCam instrument spectral resolution is ~0.2 nm FWHM in the UV (240–335 nm) and blue (385–465 nm) regions and 0.65 nm in the VNIR (500–850 nm), which was reproduced in the tests reported here.

LIBS instruments have several advantages that make them well-suited to planetary exploration, as well as routine field work on Earth. LIBS can be set up to operate semi-remotely, with a telescope collecting the plasma light up to hundreds of meters away (e.g., Palanco et al., 2006). No sample preparation is needed; in fact, the laser can ablate through rock coatings, making it possible to create depth profiles of elemental concentrations. The laser can profile to depths up to ~1 mm in a rock, easily "burning" away dust layers and chemical weathering coatings in a very short time. The laser can also be focused to an extremely small spot size ($\leq$0.5 mm at 7 m), allowing analysis of small objects such as or the "blueberries" observed at Meridiani Planum on Mars. All elements are simultaneously analyzed, and the entire analysis can be completed in seconds or minutes.

The main disadvantage of LIBS for geological samples lies in the variation of peak intensities and areas caused by interactions in the

plasma that are a function of chemical composition. These chemical matrix effects are material, optical, and plasma properties that influence the ratio of a given emission line to the abundance of the element producing that line. Chemical matrix effects are directly related to the elemental and molecular composition of the sample and ubiquitously perturb the LIBS plasma. They are associated with 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. Minor or trace elements in the sample may cause chemical matrix effects on major element emission lines and vice versa. Local atmospheric composition and pressure also significantly influence LIBS plasma intensity because the local atmosphere and the breakdown products from the atmospheric species interact with the ablated surface material in the plasma (Cremers and Radzeimski, 2006). There is as yet no theoretical model that allows prediction of matrix effects (i.e., there is nothing analogous to the ZAF correction commonly-used in electron microprobe microanalyses, which adjusts for matrix effects based on atomic number effects, X-ray absorption, and secondary fluorescence), so correction schemes must be developed empirically (more closely akin to the Bence and Albee (1968) correction) and/or overcome using statistical methods and/or large data sets to derive predictive algorithms. The one possible exception to this is the calibration-free LIBS technique, which attempts calibration from first principles (e.g., Ciucci et al., 1996). However, in addition to making the assumption of local thermal equilibrium during the brief life of the plasma, this requires experimental details that are not available from the ChemCam experiment.

Moreover, the extent of the dependence of mineralogy, rock texture, and crystallinity on LIBS spectra is poorly known at this time, but is the subject of ongoing work. The current study contributes to this body of work by examining the importance of rock type on predicting chemical compositions using LIBS.

3. Geological context

For this work, we compare and interrelate three suites of rocks representing a potential continuum of fresh to completely-weathered rocks on Mars. Because most of the martian surface was originally igneous (probably basaltic, cf. McSween et al., 2009), the first suite comprises various igneous rock types with disproportionate representation of basalts. A second suite is composed of various phyllosilicate minerals that are the end products of complete hydrolysis (chemical weathering of basalts through interactions with water) of basaltic silicate minerals. Given these two "end members," we seek to predict the compositions of samples in the third suite, which is a set of sedimentary rocks that formed as a result of the alteration of basaltic precursors in the East African Rift.

The "century set" consists of 100 igneous rocks with a wide range of compositions from Tucker et al. (2010). In that suite, the majority of samples are basalts by composition, but samples of higher and lower silica content are included to extend the range of calibration to the entire range of naturally-occurring igneous rock compositions. These samples are reasonable proxies for igneous rocks that might be found on Mars based on meteorite, orbital, and rover data (McSween et al., 2009). The most notable difference is that martian basalts are slightly more rich in FeO than terrestrial occurrences (16–18 wt.% vs. 10 wt.%), but this data set encompasses samples with FeO covering that range (see Fig. 4 in Tucker et al., 2010).

Spectra of a suite of 17 phyllosilicate standards from Tucker et al. (2008) were also acquired; these are a combination of Clay Minerals Society standards and standards available from Brammer Standard Company, Inc. A summary of the compositions of these two data sets, along with those of the sedimentary samples introduced here, is given in Table 1.

The third suite of samples, and the focus of this study, consists of sedimentary samples that were selected from the East African Rift (EAR) suite studied by Ashley and Driese (2000) for a number of reasons. Rift valleys are prominent features on the surfaces of both Mars and many continents of Earth. Valles Marineris (~8000×200×5 km) dominates the face of Mars, and on Earth the EAR (~6000×150×3 km) is by far the largest modern rift system. Both rifts are formed by extensional tectonics producing thick sequences of basalt flows and volcaniclastic sediments and expose significant sections of planetary crusts. The EAR is centered in the low latitudes (tropics) and because of the low elevation of the EAR floor, the rift has a localized climate system that is dominated by arid weathering and erosion where evaporation is 4–5 times higher than precipitation. The regional settings of both EAR and Valles Marineris show morphological evidence of arid to semi-arid landscapes: playa lakes and ephemeral rivers and groundwater-fed discharges. The EAR soils (Inceptisols to Aridisols) that form in evaporative basins precipitate sulfates (gypsum and anhydrite), chlorides (halite and sylvite), and trona. In Valles Marineris, and the near-equatorial region of Mars around it, the paleo-climate is known in less detail, but appears to have been dominated by dryland processes. The EAR system may be a good Earth analogue for areas on Mars suggesting rifting, as well as martian surface sediments, in general, in terms of tectonics, petrology, and climate.

Volcanism in the EAR ranges in composition from basalt to andesite to trachyte and their alkaline/peralkaline equivalents; SiO_2 (49–53%), FeO (10–20%) MgO (6–12%) and CaO (~20%) (McHenry, 2005). Compositions inferred from orbital observatories (i.e. spectral data) suggest the predominance of basalts on the surface of Mars within 40° of the equator. Spectral data from Mars Express and the Mars Reconnaissance Observer (MRO) suggest large areas of phyllosilicates (clay) and evaporites (Fairen et al., 2010). Analyses of the soil samples collected by the Viking landers in 1976 indicate iron-rich clays consistent with weathering of basaltic rocks (Francis, 1993). Weathering and erosion in both sites have produced fans of volcaniclastic debris derived from these lavas.

Table 1
Means and standard deviations of three data sets used.

	Century set				Phyllosilicate set				Sedimentary set			
	Mean	1σ	Min	Max	Mean	1σ	Min	Max	Mean	1σ	Min	Max
SiO_2	52.53	9.69	38.23	76.99	53.22	9.41	34.70	70.10	51.89	11.53	12.85	58.38
Al_2O_3	12.92	3.43	2.28	22.96	16.48	12.28	0.69	39.70	13.54	3.36	2.59	15.74
TiO_2	2.04	1.40	0.08	7.04	0.42	0.61	0.00	2.08	1.10	0.31	0.37	1.43
Fe_2O_3T	10.79	4.23	0.94	20.24	9.51	13.86	0.22	37.42	7.68	1.97	1.97	9.67
MgO	8.91	7.40	0.00	33.49	5.38	6.78	0.03	23.30	5.10	1.45	3.74	7.60
MnO	0.17	0.06	0.01	0.29	0.03	0.04	0.00	0.11	0.24	0.05	0.17	0.34
CaO	8.12	3.66	0.31	17.79	3.18	5.92	0.00	23.40	3.87	8.97	0.38	34.52
K_2O	1.35	1.60	0.05	5.82	1.29	2.78	0.02	7.83	2.07	0.54	0.59	2.75
Na_2O	2.56	1.15	0.00	6.08	0.40	0.69	0.00	2.10	4.68	1.59	2.65	8.28
P_2O_5	0.43	0.56	0.01	2.69	0.09	0.21	0.00	0.80	0.20	0.04	0.08	0.27

1σ = standard deviation.

In the EAR, smectite clay and zeolites develop during weathering, pedogenesis and early diagenesis, after burial in alluvial/debris fans, soils, and playa saline/alkaline lakes (Hay and Kyser, 2001). Research in Plio-Pleistocene rocks of Olduvai Gorge (3°S, on the margin of the EAR) reported that compositional signatures from clay minerals produced during early pedogenesis are inherited from their parent source rocks (Hover and Ashley, 2003). Sediments sourced from volcanics contain highly-disordered dioctahedral smectite. Smectite in soils developed further from volcanic source rocks that have more Fe (III) and Mg than soils closer to source rocks. Diagenesis in surface sediments involves replacement of original Al-rich smectite by neoformed trioctohedral Mg-rich smectite (stevensite). Neoformed celadonite replaces smectite in the most saline lake sediments. Zeolites (mainly Na-rich analcime) also appear to be primary pedogenic to early diagenetic minerals formed by replacement of volcanic rock fragments and as pore-fillings in root and animal biopores (Ashley and Driese, 2000). In that study, whole-rock X-ray fluorescence (XRF) data were used to evaluate chemical variations in the soil following the molecular ratios approach of Retallack (2001) and a mass-balance approach (cf. Brimhall et al., 1991a,b).

A suite of 18 EAR rocks representing six EAR lithologies was studied (Table 2). The parent rock type is represented by the red paleosol, and the derived lithologies are silty clay, lapilli-rich clay, waxy clay, calcareous waxy clay, and tuff and tuffaceous clay. Ashley and Driese (2000) powdered their samples to obtain bulk chemical analyses using XRF; those XRF results were used as the "true" composition for each sample. For 16 of their samples, there was enough leftover rock powder to provide the 3 g of sample needed to press powder pellets for the LIBS analyses.

4. Experimental methods

About 150 g of each sample was crushed to <45 μm particle size (about an order of magnitude smaller than the LIBS beam diameter) in a Spex tungsten carbide shatterbox in order to mitigate inhomogeneity and equalize grain size and porosity. Aliquots of this powder were used for both XRF in the original study (Ashley and Driese, 2000), and LIBS for the current study. Major and minor elements were measured in the University of Tennessee (UT)-Knoxville, Earth and Planetary Sciences Department. Powdered soils were oven dried at 608 °C and were pressed into pellets using crystalline boric acid as a binder. Paleosol samples were analyzed for selected major, minor, and trace elements using an EG&G ORTEC TEFA III tube-excited X-ray fluorescence (XRF) analyzer (Singer and Janitzky, 1986). XRF analyses and errors are reported in Ashley and Driese (2000). A clays protocol (prepared at UT-Knoxville) that employed appropriate high-iron clay soil and shale standards was used for quantitative analyses; analytical uncertainties for XRF analyses are also reported in Ashley and Driese (2000).

Table 2
Rock types for sedimentary rocks used in this study.

Sample name	Location	Depth in core (m)	Rock group
46A-2	Trench 46A	−140	Red paleosol
46A-9	Trench 46A	0	Red paleosol
46B-1	Trench 46B	−160	Tuff and tuffaceous clay
46B-2	Trench 46B	−150	Tuff and tuffaceous clay
46B-3	Trench 46B	−140	Waxy clay
46B-5	Trench 46B	−120	Red paleosol
46B-16	Trench 46B	−10	Red paleosol
46C-2	Trench 46C	−195	Waxy clay
46C-3	Trench 46C	−170	Waxy clay
46C-4	Trench 46C	−160	Lapilli-rich clay
46C-5	Trench 46C	−150	Calcareous waxy clay
46C-7	Trench 46C	−130	Red paleosol
86-1	Trench 86	−170	Calcareous waxy clay
86-2	Trench 86	−160	Silty claystone
86-3	Trench 86	−150	Silty claystone
86-4	Trench 86	−140	Silty claystone

For LIBS sample preparation, 3–5 g of sample was poured into an aluminum cup and pressed under ~15 metric tons of pressure. No binders were added. LIBS analyses of the pellets were performed at Los Alamos National Laboratory (LANL) using instrumental and environmental conditions configured to mimic those on Mars with the ChemCam instrument. Because of known atmospheric pressure effects on LIBS spectra, samples were placed in an evacuated sealed chamber filled with 7 Torr (0.009 atm) CO_2 to simulate the martian atmosphere. Samples were probed with a Nd:YAG laser operating at 1064 nm and 10 Hz, set to 17 ± 1 mJ per shot at a 9 m standoff distance. Spectra were recorded with three Ocean Optics HR2000 spectrometers covering the UV, blue, and visible/near-infrared ranges nearly continuously from 220 nm to 930 nm. The spectrometer integration was set to 1 s and five averages were collected from each sample spot such that each spectrum represented the emission from 50 laser shots. Even though the grain size is smaller than the laser spot, we probed the five spots to further compensate for possible inhomogeneities. Sample spectra of each of the six rock types are shown in Fig. 1.

5. Data pre-processing protocols

The Ocean Optics software summed the 50 shots for each spot to correct for variations in shot intensity over time, and also subtracted dark (non-laser) backgrounds. Hot pixels were then removed from each of the five raw data files acquired from different spots on each sample through a combination of smoothing and manual inspection of each spectrum. Three subsequent pre-processing steps were undertaken, and are described in Appendix A. From them, we obtained one wavelength-registered, background-subtracted, mean spectrum for each sample, which was then used for subsequent statistical analyses.

Elemental analyses were re-scaled to ensure that the PLS assigns equal importance to all elements. As recommended in Tucker et al. (2010), each elemental concentration in oxide wt.% was divided by the standard deviation of that element's distribution: $Y_i' = Y_i/\sigma_y$. This procedure rescales all elemental distributions to have a variance of 1. It must be noted that re-scaling is only appropriate when the original elemental distribution is approximately Gaussian.

6. Statistical analysis

6.1. Overview

Data were analyzed using software written in GNU R (R Development Core Team, 2010). This customized R software wraps routines from several packages including hyperSpec (Beleites and Sergo, 2011), Peaks (Morháč, 2008, 2009), and PLS (Wehrens and Mevik, 2007) and specifically applies them to LIBS data sets. HyperSpec is an unpublished software package used for convenience here to make programming with spectral data types more tractable.

This study uses multivariate regression to model the relationship between the LIBS spectrum and elemental composition of a geological sample. The spectra are recorded in a matrix **X** with *n* rows (one for each sample) and *p* columns (one for each wavelength bin). The response variables are stored in a matrix **Y**, with one row per sample and one column per major element. The familiar equation for multiple linear regression models elemental composition as a linear combination of channel intensities. The equation below is the standard formulation of this model (sometimes called multiple linear regression) where each β_i is learned from the training set:

$$\mathbf{Y}_j = \beta_0 + \beta_1\mathbf{X}_1 + \beta_2\mathbf{X}_2 + \ldots + \beta_n\mathbf{X}_p. \quad (1)$$

Normally, a linear regression model like this is fit using Ordinary Least Squares (OLS), which is the standard procedure for finding a

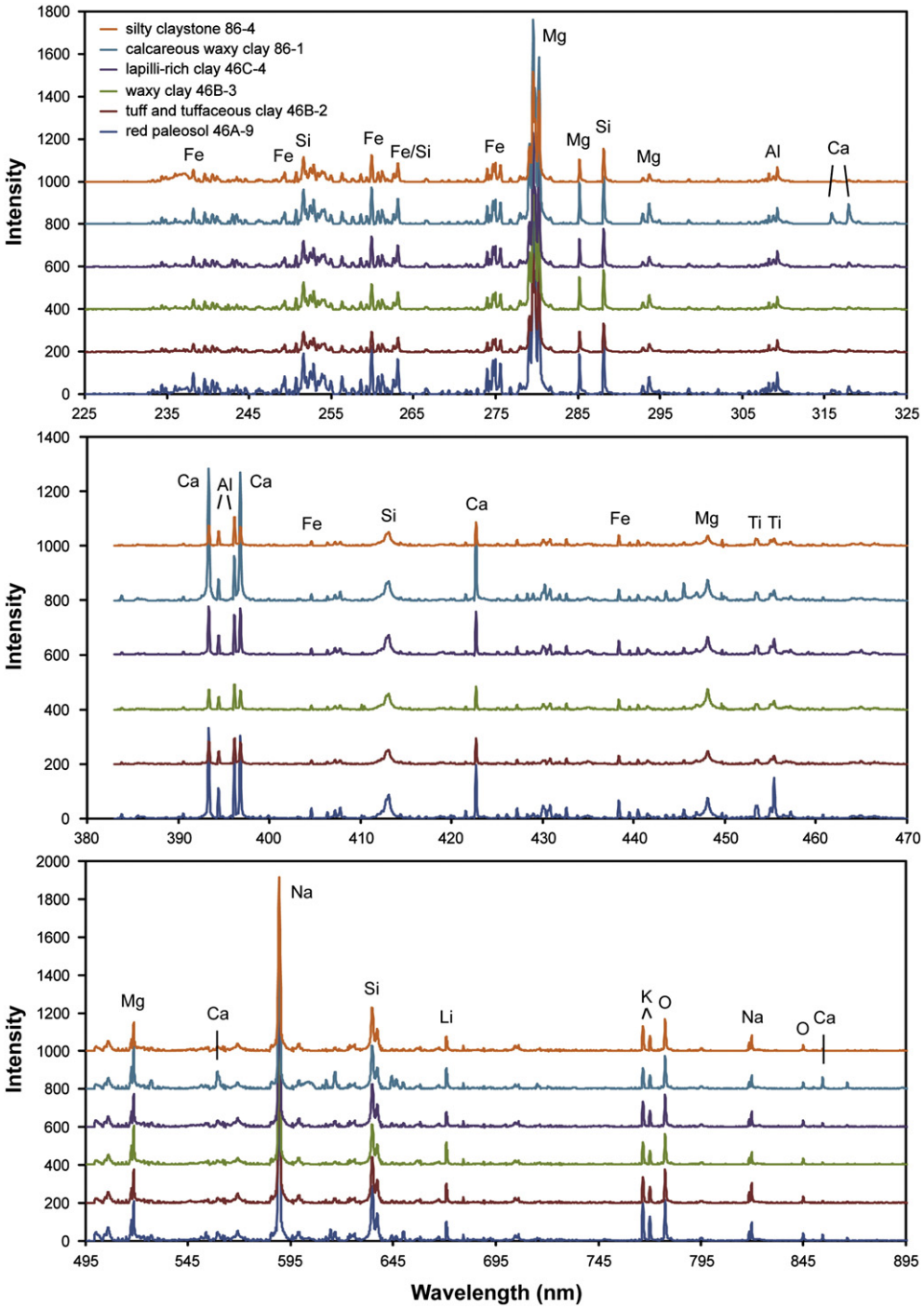

Fig. 1. LIBS spectra of sedimentary rocks representing each of the six rock types studied. Spectra were recorded with three Ocean Optics HR2000 spectrometers covering the UV (top), blue (middle), and visible/near-infrared (bottom) ranges nearly continuously from 220 nm to 930 nm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

coefficient vector β. Unfortunately, in the case where $p \gg n$, the coefficient vector β produced by an OLS fit is far from optimal, and can lead to very poor predictions of unknown samples. Because our spectra have 6144 channels (p) and we have only 100 geological samples (n), OLS is not an appropriate technique for estimating β.

The weakness of OLS in this situation is two-fold: first, many of the predictors (channels) are highly correlated, violating the assumption of independence that is inherent in multiple linear regression. Spectral peaks tend to be composed of several (5–20) channels, so our data contain many small groups of highly correlated predictors. Second, there are very many predictors, most of which may simply be irrelevant for a given element. The OLS procedure is not well-suited at giving high correlation values to the best predictors when the ratio of irrelevant to relevant predictors is very large.

To solve this problem, shrinkage methods are used to transform the standard multiple linear regression problem into a smaller equation. LIBS data can be viewed as a collection of n points in p-dimensional space. If the dimensionality of this space can be shrunken while maintaining a high-fidelity representation of the original data, OLS can be used to generate a good β coefficient vector, not for the original p columns of $\mathbf{X}$, but for some $\mathbf{X}'$ with a small number of columns that still adequately represents the data.

PLS is a widely-used shrinkage technique that implements this idea. The PLS transformation is not simple to describe mathematically, and there are numerous subtleties involved in developing algorithms to compute it. PLS is fundamentally based on principle components analysis, a similar but much simpler dimensionality-reducing transformation, which is described in detail in Hastie et al. (2009) and Jolliffe (2002). Based on this foundation, additional specifics for PLS are given in Hastie et al. (2009), Frank and Friedman (1993), and the overview paper by the inventor of the technique (Wold et al., 2001).

Conceptually, PLS is a transformation that takes a matrix **X** and a desired number of dimensions k, and then projects **X** down into that number of dimensions. Thus, each point in p-dimensional space is mapped to a point in k-dimensional space $k \ll p$. Though it seems that much data are being discarded, this is exactly what is desired. It is analogous to the compression that digital music files undergo to save space on consumer MP3 players: the important part of the signal remains in the reduced data. The PLS transformation also guarantees that the new dimensions are orthogonal to each other: in statistics terminology, this means that they are uncorrelated despite the high co-linearity of the original data. Finally, PLS outputs a vector v that can be multiplied by any p-dimensional point to embed that point into the k-dimensional space. This allows use of an OLS fit that produces a value for the β coefficient vector in the reduced space to predict unknown samples.

The dimensions of the reduced space are called components. The amount of compression that is ideal for any given data set is a characteristic of that data set, and must be determined by testing, as will be explained in Section 7 below.

This study utilizes two different types of PLS (Hastie et al., 2009). PLS-1 regresses a single response variable (concentration of a single element) against the predictor variables (spectra). PLS-2 simultaneously regresses multiple response variables (elemental concentrations of the ten major elements in rocks) against the predictors, taking advantage of natural correlations between elements. Both techniques explain the variance in both **X** and **Y**. The tests described herein use only the major elements Si, Al, Ti, Fe, Mg, Mn, Ca, Na, P, and K for the **Y** variables; the **X** variables are the 6144 channels of the three detectors.

Three different data sets were used to predict the sedimentary rock compositions: C (the century set of 100 igneous samples), P (the 17 phyllosilicates), and C + P (both data sets combined). The sedimentary data set (S), which consisted of 16 samples total, was randomly split into two groups of eight samples each: one for validation to tune parameters of the model (SV), and one for use as a test set to apply derived parameters and provide an estimate of how good the generalization of the data will be with those fixed parameters (ST). The C, P, and C + P data sets were used to train both PLS-1 and PLS-2 models. A summary of our data processing steps is given in Fig. 2.

For both PLS-1 and PLS-2 models, validation was done externally using the SV set (step C2 in Fig. 2) and internally (step C1 in Fig. 2) using a jackknife or "leave one out" full cross-validation as recommended in Mevik and Cederkvist (2004) for PLS (k-fold validation). In the latter kind of validation, a PLS model is built from $n-1$ spectra out of n training set samples, and the composition of the sample left out is predicted from the reduced PLS model. This procedure is repeated, leaving out and predicting each of the samples one at a time. To provide an effective means of comparing results from different models, we use the root mean square error of prediction (RMSEP), which guesses the predictive power of the model on unseen data acquired on remote samples (Hastie et al., 2009).

7. Effects of selection of model parameters from internal vs. external validations

To select the number of components that should be used to predict elemental composition, we created validation plots for the C, P, and C + P data sets (steps C1 and C2 in Fig. 2) and used those values to select the number of components to be used in subsequent testing of the models. Figs. 3–5 show the three sets of validation plots, each comparing internal and external validation results.

Careful selection of the number of components to be used in any predictive model is highly dependent on the characteristics of the training and validation sets. In our situation, we have two training sets, one large (C) and one small (P), to predict the compositions of a validation set (SV). All RMSEP values for external validations are directly comparable because they are being tested on the same prediction set (SV). These calculations were done using both PLS-1 and PLS-2, but the resultant shapes of the plots are indistinguishable, with the RMSEP values for PLS-1 being slightly lower. Thus the following discussion refers to either type of PLS.

1. For C only, the internally-validated plots (Fig. 3) generally show a monotonic decrease in RMSEP as the number of components increases, presumably because all the samples in that data set are compositionally similar. Alternatively, the externally-validated plots, which use the C set to predict the sedimentary rock SV compositions, consistently suggest the use of a lower number of components than the internally-validated plots. This makes sense because the sedimentary rocks are fundamentally different from those in the C set, and this reflects a cut-off point past which the C-based model is too biased (specialized) towards igneous rock compositions to be useful for predicting the SV rocks. The lesson learned from this comparison is that no single number of components works for every compositional variable, suggesting that the models must be finely-tuned. Moreover, when using a training set to predict an unknown of a distinct rock type or composition, different model parameters would be implied for internal and external validation.
2. If we use the 17-sample phyllosilicate (P) data set for the same comparison (Fig. 4), results demonstrate the instability of predictive models based on small training sets, as the numbers of components increases in both the internally- and externally-validated runs relative to the C set (Fig. 3). This result means that the phyllosilicate data set alone cannot even predict its own constituent sample compositions very well, and thus is unlikely to be very useful in predicting compositions of anything else. This problem likely arises from the small size of the training set and the disparate compositions represented in it (Table 1). Note that the samples in all three data sets used here cover approximately the same range of composition space, but they do so with very different distributions. These results raise a cautionary flag for interpretation of other previous studies that used very small sample sets and less well-chosen methods for selecting the number of components in their models, especially in situations where the training set samples have very different compositions from the unknowns.
3. The C + P data set has different compositional diversity, but neither C nor P is very similar to SV for all variables. Combining the phyllosilicates with the igneous rocks does not have any really significant impact on the parameters of the validation models, and the same trends are seen as for the C set alone. Again, the internally-validated plots (Fig. 5) show a general decrease in RMSEP as the number of components increases and the externally-validated plots suggest that use of a lower number of components could be advantageous on test data. Also, it is apparent from the external validation set that the number of components for all elements is not the same, so that a specific number of components should be chosen individually for each element.

8. Test results

The most important goal of this paper is to determine how well compositions (and thus parageneses) of unknown samples on Mars

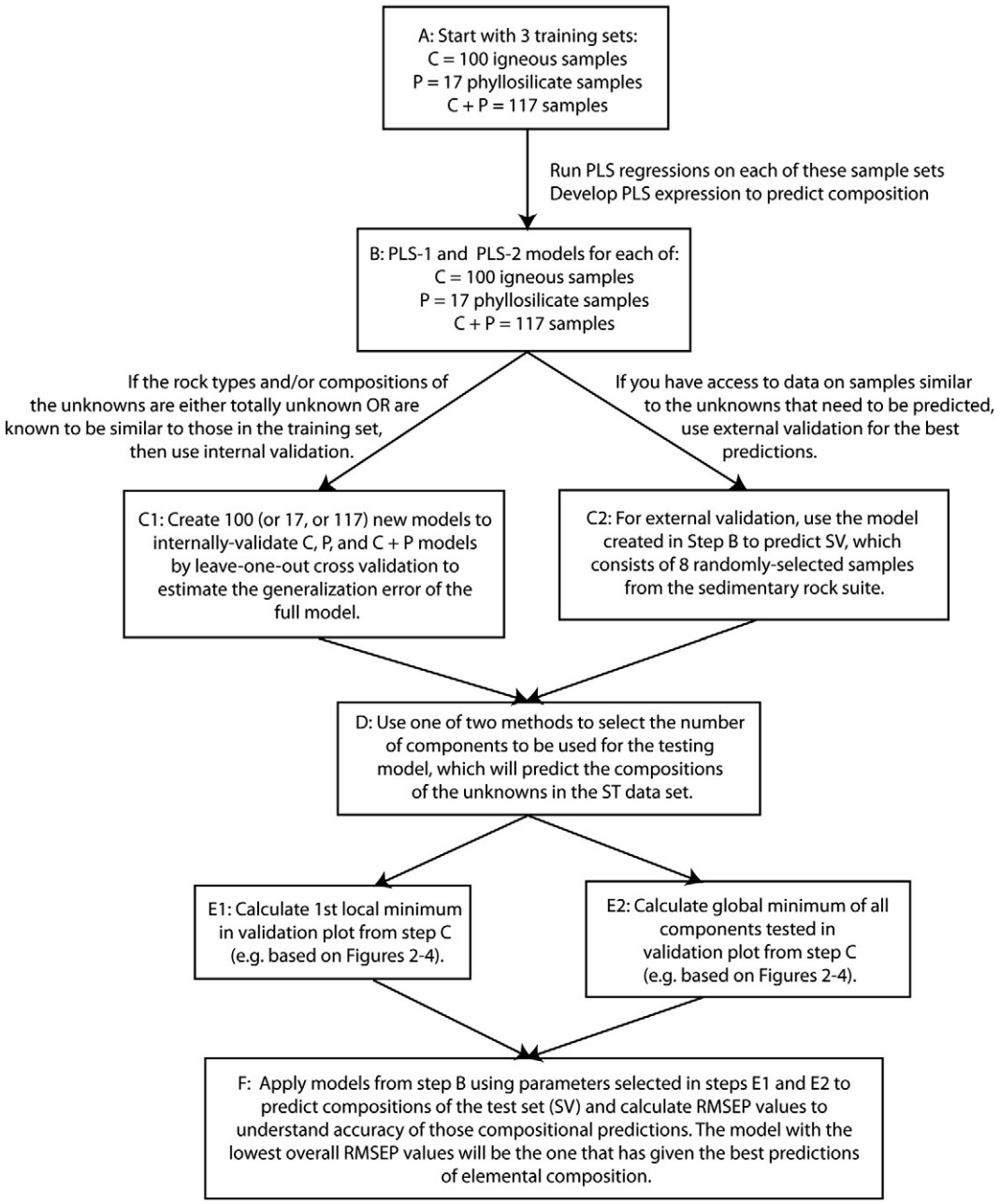

Fig. 2. Flowchart of data analysis steps used in this study. Sedimentary rock compositions are best predicted using the models created in step B (particularly C + P, which uses both the 100 igneous rocks and 17 phyllosilicates in the training set) and a number of components chosen by externally-validated (step C2), global-minima in the validation plots (steps D, E2, and F).

that happen to be sedimentary rocks could be predicted. As described in Section 6 above, we will base these calculations on the compositions of the eight samples in the sedimentary rock test set (ST) that has consistently been excluded from any training set.

RMSEP values obtained by use of varying numbers of components are used to predict each element (steps D and E in Fig. 2). We here test two basic methods for choosing the number of components based on the validation results:

A. Starting from the left (in Figs. 3–5) at the lowest number of components (0, which is the intercept), choose the first local minimum encountered. For example, in Fig. 4, this would be at 6 components for SiO_2 (1st local minimum model). This procedure is recommended by Mevik and Wehrens (2007) to avoid overfitting.
B. Use the number of components that gives the absolute lowest value for RMSEP for each element (global model).

We arbitrarily chose to evaluate 1–15 components only because that is the maximum possible for the 17-sample phyllosilicate data set. This choice was made because in general, RMSEP values are either nearly constant at values higher than 15, or they start to increase again with higher numbers of components due to overfitting the data. Fig. 6, which shows an internal validation based on the century set alone, provides an uncomplicated example of this effect.

These tests compare how the various methods of validation would perform on totally unseen data. Results are given in Tables 4 and 5, which show the minimum number of components and RMSEP values for PLS-1 vs. PLS-2 for internally vs. externally-validated data using the C, P, and C + P validation models. Note that comparisons between PLS-1 and PLS-2 in terms of numbers of components must be made with care to avoid over-interpretation, because we are projecting along different directions using distinctive criteria.

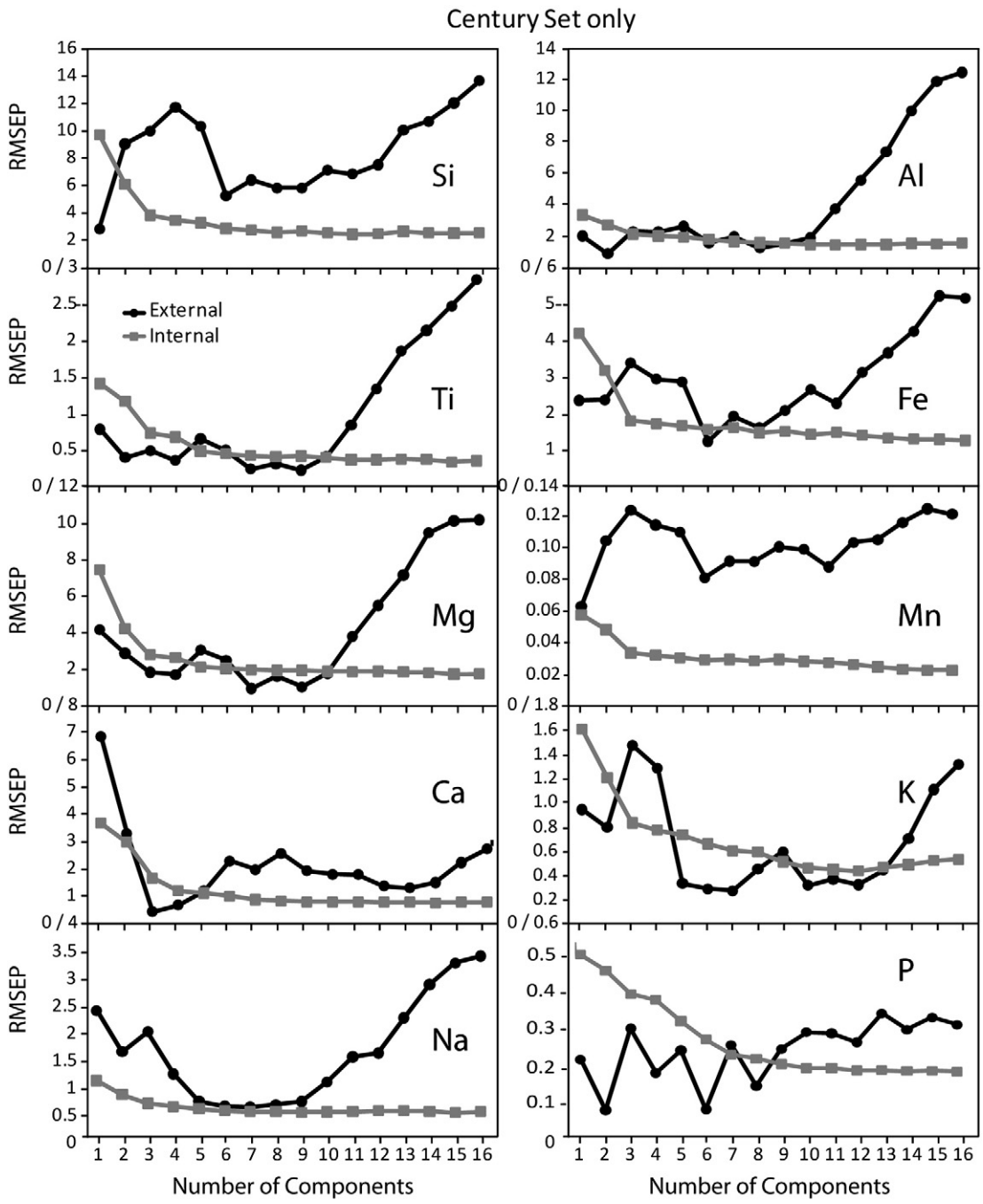

Fig. 3. RMSEP plots comparing the errors resulting from predictive models that use either internal (gray) or external (black) validation based on use of the 100-sample igneous rock suite alone. Internal validation is based on only the 100 samples and uses leave-one-out cross-validation. External validation is built from the 100 igneous samples and predicts the compositions of eight randomly-selected sediments. Note that the minima vary from element to element.

Consider the case of CaO as an example, and compare results for the externally vs. internally-validated global model for PLS-2 in Table 4. The internal model (validated without using data from any of the sedimentary rocks) selects 13, 14, and 16 components to predict CaO in the test set using the C, P, and P + S models. The high number of components here reflects a highly-customized fit to the specific characteristics of the C, P, and C + P data sets that might not be useful over a broader range of compositions. However, the associated RMSEP values for predictions of CaO in the test set samples are still reasonable, with values of 2.64, 0.80, and 0.83 (Table 5; in units of wt.% oxide). In contrast, the externally-validated model that included the SV data as part of the validation needs only 9, 5, and 5 components for predictions of the test set based on the C, P, and P + S validation models, resulting in even lower RMSEP values of 0.64, 0.80, and 0.72 (again in units of wt.% oxide). Here, the low numbers of components for C and C + P imply that the resultant PLS expressions should generalize well to the sedimentary rock data. The 15 components needed by the P model suggest that the phyllosilicate models are too self-specific to be useful in predicting the sedimentary rocks. Higher numbers of components reflect increasing specialization that allows them to predict well only unknowns that closely resemble the samples in the training set. This effect is also seen for Al_2O_3, Fe_2O_3T, MgO, and P_2O_5. For the other elements, results from internal and external validation are very similar.

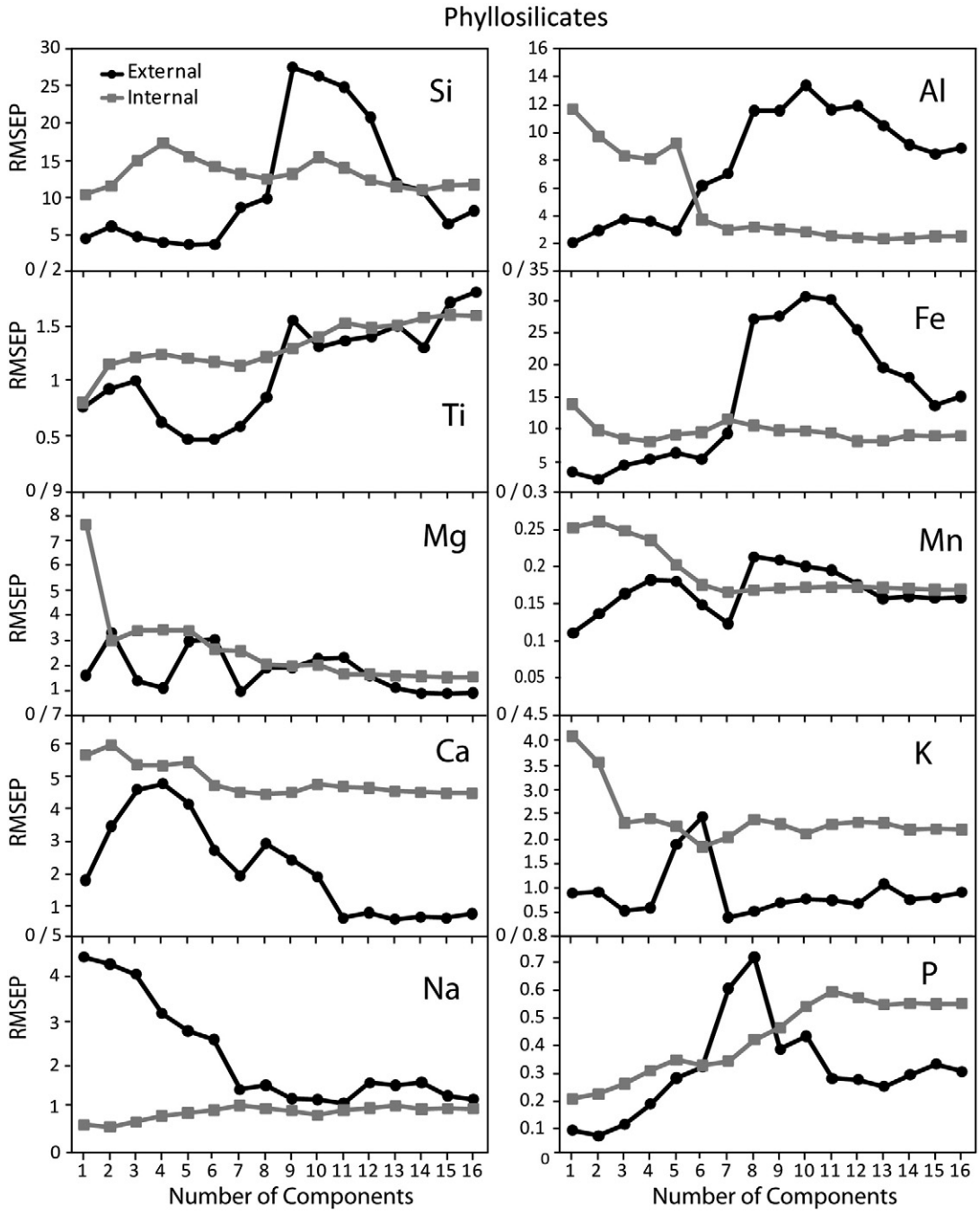

Fig. 4. Validation plots comparing internally- and externally-validated predictive models based on use of the 17 phyllosilicate samples only. Internal validation uses only the 17 samples along with leave-one-out cross-validation. External validation uses the model built from the 17 phyllosilicates to predict the compositions of eight randomly-selected sediments. For most elements, these phyllosilicate predictions are much worse (i.e. RMSEP values are significantly larger). Compared to the plots from internal validation of the C set in Fig. 3, these internal predictions are very unstable.

Of all the models shown in Table 5, the lowest overall values of RMSEP (which are the confidence limits of the predictions in the units of the original measurements) can be arbitrarily evaluated from the *sums* of the RMSEP values for each of the elements in any column. By this metric, the externally-validated models using both the century set of igneous rocks and the phyllosilicates along with the first local minimum to select an appropriate number of components for each element yield the best results for predicting the composition of the sedimentary rocks (other than using the sedimentary rocks themselves; see below). PLS-1 and PLS-2 are roughly comparable here, with PLS-1 slightly better than PLS-2 for this specific comparison.

Overall, results of this study do not suggest a clear preference for either PLS-1 or PLS-2. There are (at least) two competing factors that determine the relative efficacy of PLS-1 vs. PLS-2. Elements in the plasma are undoubtedly interacting, which would theoretically suggest that PLS-2, which can account for interactions between elemental concentrations, should yield better results. On the other hand, PLS-1 is not affected by the aggregate summed errors on the individual elemental analyses. In our results, the far-right columns of Table 5 suggest that PLS-1 yields slightly better results overall when the global minimum number of components is used, while PLS-2 generally does a better job employing the first local minimum number of

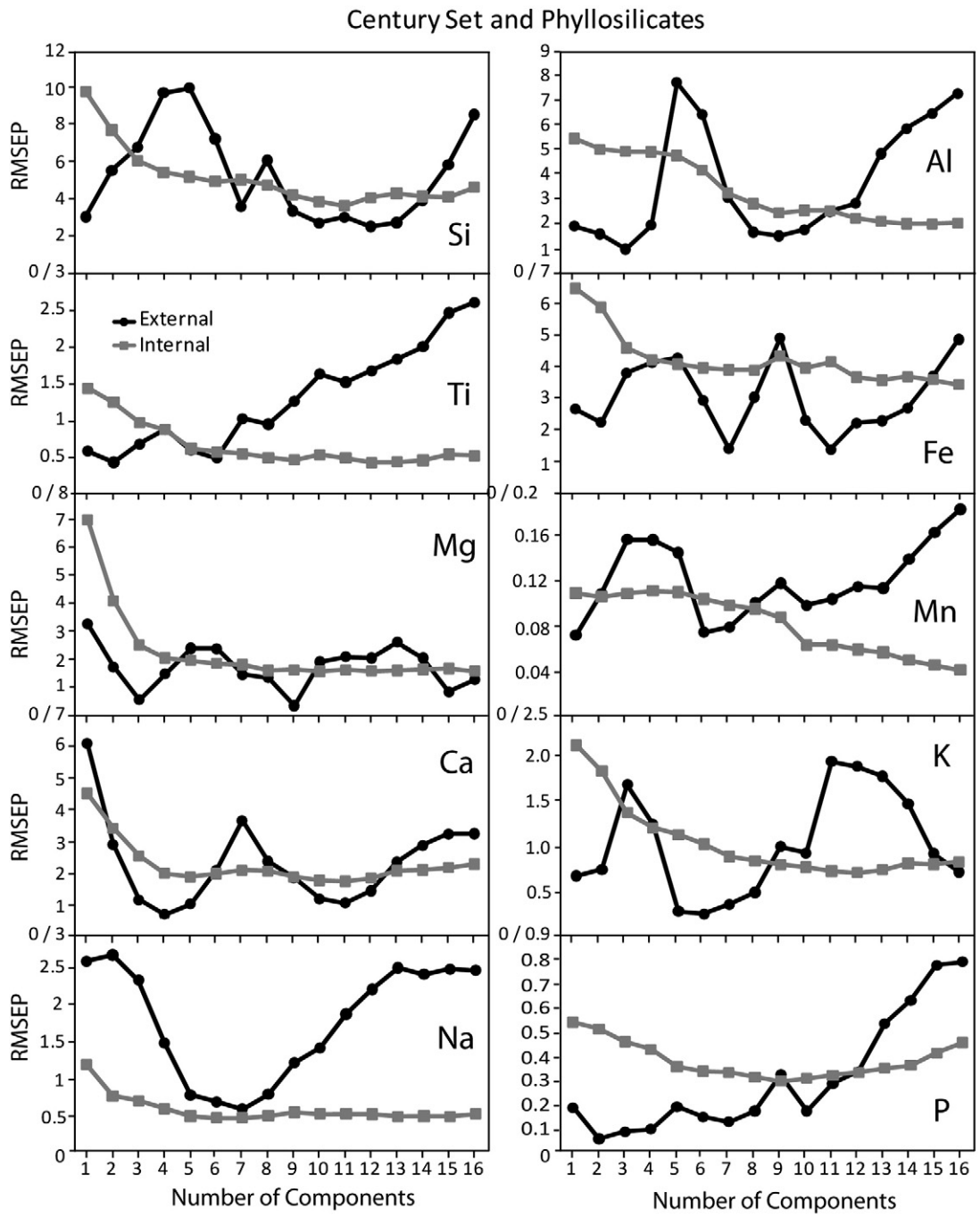

Fig. 5. Validation plots comparing internally- and externally-validated predictive models based on use of the 100 igneous samples combined with the 17 phyllosilicate samples. Internal validation uses only the 117 samples along with leave-one-out cross-validation. External validation uses the model built from the 117 spectra to predict the compositions of eight randomly-selected sediments. For most elements, these predictions have the lowest values of RMSEP compared with Figs. 3 and 4.

components (Fig. 7). However, these relative improvements may be element-specific or the result of random fluctuations.

9. Discussion

This project represents a test case for a difficult statistical problem relating to this application. The general problem is the situation where there are several sets of data that differ in some important aspect (in this case, geological provenance), but are in many ways similar (here, LIBS spectra of rocks composed mostly of the same ten elements). The simplest instance of this problem would be a case where there is one giant data set A (e.g., our 100-sample C set) and one small data set B (the 18), and a prediction task is being undertaken on B. If A and B are mostly similar, is it possible to leverage A to do better at predicting B than just using B alone? Whether or not it is possible to do this is highly dependent on the specifics of any given problem.

Thus this project addresses the issue of whether or not it is possible to cleverly combine multiple similar (yet different) data sets to boost performance on each individual task of predicting chemical composition. This paper is a necessary test of feasibility for this approach in the geologic LIBS setting. The tests presented in this paper are important because there is no theoretical/analytical way to show that this methodology is possible for any given statistical problem. Our results demonstrate that combining multiple data sets for the purpose of predicting subsets is not only feasible, but

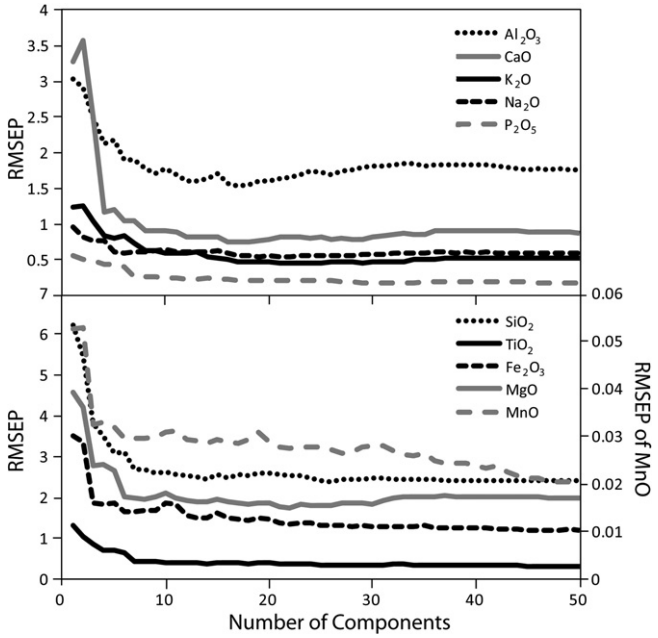

Fig. 6. Internal validation plot based on 100 igneous samples showing the futility of using high numbers of components in predictive models. In most cases, the minimum occurs within the first 15 components; in a few others, the number of components asymptotically approaches a minimum at high numbers of components, systematically over-fitting the data as it does so.

probably necessary to routinely obtain accurate predictive results for small subsets from very large, diverse training sets.

The applied goal of this project is to determine how best to use remote LIBS data to predict compositions of sedimentary rocks and understand the extent of their chemical weathering when they are encountered on Mars. To that end, Table 5 shows that the efficacy of using the first local minimum in the validation plots (model A from above) vs. the global minimum method for determining the number of components to be used (model B from above) is highly model-dependent. The first local minimum infers the number of components by effectively customizing or reducing the specificity (bias) of the model to the data set on which it is based. Overall, the first local minimum does a better job of predicting the compositions of the sedimentary rocks, especially when externally-validated, than using the global minimum to determine the proper number of components. For these samples, results from PLS-1 and PLS-2 are roughly comparable. In general, the externally-validated first local minimum model does the best job of producing predictions on the test set with minimum RMSEP values, so this can be followed as a general rule for spectra of geological samples like those used herein. However, use of external validation presumes prior knowledge of the rock type of the unknowns. Comparing only the internal validation models, the PLS-2 model using the first local minimum for component number selection and both the igneous rocks and the phyllosilicates in the training set gives the best (lowest) values for RMSEP. These RMSEP errors apply to a scenario on Mars in which the rocks being analyzed have completely unknown compositions.

What happens if only sedimentary rock compositions, rather than the more universal igneous plus phyllosilicate-based training set, are used to predict sedimentary rock compositions? This question probes the hypothesis that using a refined training set of samples intentionally chosen to resemble the unknown in terms of rock type (i.e., after an initial investigation as discussed above) will yield improved elemental predictions over that of a more universal training set. We do not at this time have an additional suite of well-characterized sedimentary rocks to use for this test. However, using the sedimentary rocks to predict their own compositions using a simple leave-one-out, internally-validated cross-validated model with PLS-1 and the first local minimum number of components yields the results shown in the lower half of Table 5 in the fifth column. This calculation shows elemental RMSEP values for the sedimentary rocks as a stand-alone data set without keeping any samples out (includes both ST and SV above). As expected with similar samples and internal validation, errors on elemental predictions are excellent. Interestingly, they are comparable to the best RMSEP predictions in the rest of the table, suggesting that a global training set actually does a fairly good job even when the rock type or composition of the samples being analyzed is completely unconstrained.

The ability of a model to make accurate predictions of elemental composition is a function of many variables, especially the characteristics of the training set used in the predictions. The spread of abundances of each element in the training set vs. the true composition of the unknowns is one of those values. For example, standard deviations on Al_2O_3 in the century set, phyllosilicates, and sediments are 3.43, 12.28, and 3.36 wt.%, respectively while the corresponding RMSEP values for sedimentary rock predictions using each of those rock types are 0.71, 1.78, and 0.62 respectively, suggesting that the higher standard deviation among the Al_2O_3 abundances in the phyllosilicates, when used as a training set, led to a higher RMSEP, that is, a less accurate prediction. However, the same logic does not apply to SiO_2 predictions, which are $1\sigma = 9.69$, 9.41, and 11.53 but have variable RMSEP values of 3.26, 4.02, and 2.18. This point is shown graphically in Fig. 8, which plots the standard deviations of all XRF elemental abundances in each of the data sets used here (century set, sediments, and phyllosilicates) on the x axis vs. the RMSEP

Table 3
Peaks used for wavelength calibration.

UV spectrometer		Blue spectrometer		VNIR spectrometer	
NIST wavelength	Element	NIST wavelength	Element	NIST wavelength	Element
234.350	Fe	383.829	Mg	505.598	Si
238.204	Fe	385.602	Si	516.732	Mg
239.563	Fe	386.260	Si	518.360	Mg
239.924	Fe	390.056	Ti	534.947	Ca
240.489	Fe	390.552	Si	553.548	Ba
241.052	Fe	391.345	Ti	588.995	Na
243.515	Si	393.366	Ca	610.272	Ca
248.327	Fe	394.401	Al	612.222	Ca
250.690	Si	394.867	Ti	614.171	Ba
251.920	Si	396.152	Al	616.217	Ca
258.588	Fe	396.847	Ca	634.710	Si
260.709	Fe	398.176	Ti	637.136	Si
261.762	Fe	398.976	Ti	643.900	Ca
262.567	Fe	399.864	Ti	646.257	Ca
271.441	Fe	404.581	Fe	670.779	Li
271.903	Fe	406.359	Fe	714.815	Ca
272.754	Fe	422.673	Ca	720.219	Ca
273.955	Fe	427.176	Fe	720.943	Ti
274.320	Fe	428.301	Ca	732.615	Ca
276.181	Fe	430.253	Ca	766.490	K
279.078	Mg	430.591	Ti	769.897	K
279.553	Mg	430.774	Ca	787.705	Mg
280.271	Mg	431.865	Ca	789.637	Mg
285.213	Mg	432.576	Fe	818.326	Na
288.158	Si	438.355	Fe	819.482	Na
292.863	Mg	440.475	Fe	849.802	Ca
293.651	Mg	455.403	Ba	854.209	Ca
298.765	Si	459.618	O	866.214	Ca
308.215	Al	460.733	Sr	891.207	Ca
309.271	Al	461.727	Ti	892.736	Ca
315.887	Ca	462.310	Ti		
317.933	Ca	462.934	Ti		
319.091	Ti	464.181	O		
320.256	Ti	464.914	O		
321.707	Ti	466.306	Al		
322.284	Ti	468.191	Ti		

Table 4
Number of components selected for all models based on Figs. 3–5.

	SiO_2	Al_2O_3	TiO_2	Fe_2O_3T	MgO	MnO	CaO	K_2O	Na_2O	P_2O_5
Global minimum										
Internal validation										
Phyllosilicates PLS-1	1	13	1	4	15	7	8	6	2	1
Century set PLS-1	11	11	15	16	15	16	14	12	15	16
Both sets PLS-1	11	15	12	16	10	16	11	12	7	9
Phyllosilicates PLS-2	1	15	1	4	16	13	13	14	2	1
Century set PLS-2	15	14	15	14	15	15	14	16	7	16
Both sets PLS-2	13	16	16	16	12	16	16	16	9	16
External validation										
Phyllosilicates PLS-1	5	1	5	2	14	1	13	7	11	2
Century set PLS-1	1	2	9	6	7	1	3	7	7	2
Both sets PLS-1	12	3	2	11	3	6	4	6	7	2
Phyllosilicates PLS-2	4	2	9	7	16	3	9	10	14	3
Century set PLS-2	1	3	2	10	12	1	5	10	12	2
Both sets PLS-2	13	6	2	2	4	7	5	14	9	2
1st local minimum										
Internal validation										
Phyllosilicates PLS-1	8	4	7	4	9	7	4	3	10	6
Century set PLS-1	8	11	8	6	11	6	9	12	9	12
Both sets PLS-1	6	9	9	8	8	10	5	12	7	9
Phyllosilicates PLS-2	4	4	4	4	5	6	7	5	14	7
Century set PLS-2	6	5	5	5	4	4	5	6	7	9
Both sets PLS-2	5	5	4	6	4	4	6	8	3	7
External validation										
Phyllosilicates PLS-1	3	5	5	6	4	7	7	4	7	9
Century set PLS-1	6	4	4	6	4	6	3	7	7	4
Both sets PLS-1	7	3	6	7	3	6	4	6	7	4
Phyllosilicates PLS-2	4	6	6	7	6	3	9	4	5	3
Century set PLS-2	3	3	5	7	5	3	5	3	8	4
Both sets PLS-2	7	3	7	5	4	7	5	6	9	5

error on prediction of that element in the sedimentary rocks on the y axis. For some elements, a small spread of composition (i.e., as represented by 1σ) in any given element results in a smaller error on the prediction to a first order, but this is not a universal conclusion. This suggests that the distribution of values within each training set is varying (i.e., 1σ is an imperfect representation of the diversity in that element) and/or that some underlying physical process is not being accounted for in our analysis.

Table 5
RMSEP values for PLS-1 vs. PLS-2 resulting from use of numbers of components in Table 4.

	Internal validation, PLS-1				Internal validation, PLS-2			External validation, PLS-1			External validation, PLS-2			Sum	
	Phy	Cset	C + P	Seds	Phy	Cset	C + P	Phy	Cset	Ct + P	Phy	Cset	C + P	PLS-1	PLS-2
Global minimum															
SiO_2	10.31	5.96	7.74		4.57	7.04	10.61	4.12	5.61	3.92	4.57	7.27	8.82	**37.66**	**42.89**
Al_2O_3	2.80	3.98	1.76		4.42	2.19	0.91	3.05	2.04	0.57	2.37	0.76	1.33	**14.19**	**11.98**
TiO_2	0.54	0.31	1.23		0.94	0.45	0.88	0.41	0.40	0.47	0.82	0.45	0.39	**3.37**	**3.94**
Fe_2O_3T	5.30	1.33	3.14		5.21	3.76	5.32	5.45	1.33	1.48	1.95	1.43	4.13	**18.04**	**21.80**
MgO	1.92	3.92	1.80		2.79	1.70	1.06	1.09	1.69	1.00	2.39	1.69	1.06	**11.41**	**10.70**
MnO	0.12	0.09	0.10		0.20	0.14	0.14	0.12	0.09	0.07	0.11	0.09	0.07	**0.59**	**0.74**
CaO	4.49	1.95	0.75		2.64	0.80	0.83	1.68	0.82	0.64	0.64	0.80	0.72	**10.33**	**6.42**
K_2O	0.81	0.43	1.91		1.04	0.63	1.22	0.85	0.34	0.46	0.99	0.53	1.45	**4.79**	**5.87**
Na_2O	1.17	0.50	0.71		0.84	0.67	2.77	1.48	0.54	0.71	4.55	0.67	0.72	**5.10**	**10.23**
P_2O_5	0.35	0.22	0.35		0.18	0.12	0.32	0.43	0.16	0.10	0.09	0.22	0.10	**1.62**	**1.02**
Sum*	**27.80**	**18.70**	**19.49**		**22.84**	**17.50**	**24.05**	**18.67**	**13.03**	**9.41**	**18.48**	**13.91**	**18.80**	**107.11**	**115.59**
First local minimum after intercept and first component															
SiO_2	4.65	6.99	2.21	2.18†	4.65	6.92	2.02	4.02	3.26	1.88	4.57	3.26	2.02	**23.01**	**23.45**
Al_2O_3	10.65	3.98	6.64	0.62	8.98	2.41	2.64	1.78	0.71	0.57	0.53	0.76	0.72	**24.33**	**16.04**
TiO_2	0.72	2.45	1.62	0.23	0.72	0.35	0.85	0.41	0.25	0.37	0.44	0.28	0.29	**5.84**	**2.94**
Fe_2O_3T	5.30	5.02	4.26	1.02	5.21	1.56	2.69	2.36	1.33	1.21	1.95	1.07	1.42	**19.47**	**13.91**
MgO	0.87	10.11	2.33	0.74	0.90	4.14	1.67	0.87	0.94	1.00	0.90	1.42	1.06	**16.12**	**10.10**
MnO	0.12	0.12	0.19	0.04	0.25	0.08	0.09	0.11	0.07	0.07	0.11	0.07	0.07	**0.69**	**0.66**
CaO	2.77	1.06	0.84	0.43	2.31	2.42	2.25	0.48	0.82	0.64	0.64	0.80	0.72	**6.60**	**9.15**
K_2O	2.49	0.43	1.91	0.37	1.25	0.89	0.56	0.55	0.34	0.46	0.40	0.32	0.44	**6.18**	**3.87**
Na_2O	4.54	2.95	0.71	0.72	4.60	0.67	0.72	1.07	0.54	0.71	0.84	0.53	0.72	**10.52**	**8.08**
P_2O_5	0.11	0.24	0.35	0.02	0.11	0.41	0.16	0.09	0.07	0.06	0.09	0.06	0.03	**0.91**	**0.85**
Sum*	**32.23**	**33.35**	**21.07**	**6.37**	**28.98**	**19.84**	**13.68**	**11.74**	**8.33**	**6.96**	**10.47**	**8.57**	**7.50**	**113.67**	**89.04**

*The sum of RMSEP values is calculated for each model to provide an arbitrary means of comparing different models, though it has no statistical use.
†These data are 1st local minima from a separate calculation, and are not from Table 4.

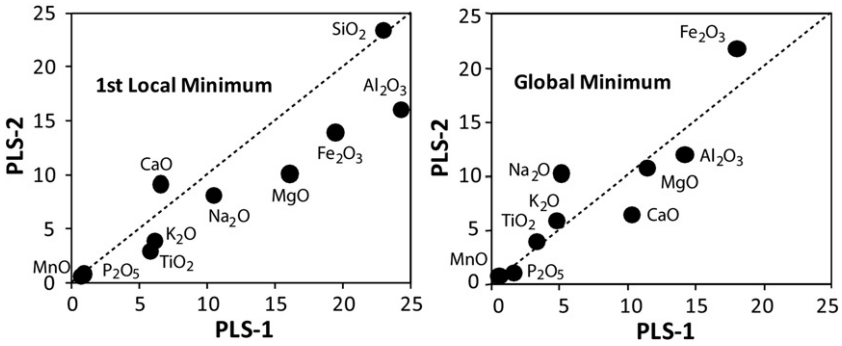

Fig. 7. Comparison of summed RMSEP values for all types of validation and combinations of different samples in the training sets for PLS-1 and PLS-2 models (data from right-most two columns in Table 4). SiO_2 is not shown for global minimum because it lies far outside the region plotted. A dashed 1:1 line indicating where the two models would produce identical results is shown diagonally in each plot. Neither PLS-1 nor PLS-2 is a clear favorite for producing low prediction errors.

10. Geochemical prognosis

Within these errors, we can finally address the question of how useful LIBS results would be for interpretation of sedimentary rock types. On Mars as on Earth, source rock type, weathering, transport, sorting, redox changes, and diagenesis are all represented in the chemistry of sedimentary rocks. Provenance information can be obtained from sedimentary rocks only if modification by secondary processes is minor. For example, Retallack (2001) suggests there are four competing kinds of chemical reactions in paleosols that provide information about the extent of chemical weathering; those that involve major elements analyzed here include increases in:

1. Hydrolysis: $Al_2O_3/(CaO + MgO + K_2O + Na_2O)$ and Al_2O_3/SiO_2
2. Oxidation: $(Fe_2O_3 + FeO)/Al_2O_3$ and $(Fe_2O_3 + FeO + MnO)/Al_2O_3$
3. Hydration: $SiO_2/(Fe_2O_3 + Al_2O_3)$, and
4. Salinization: $(K_2O + Na_2O)/(Al_2O_3)$, Na_2O/K_2O, and Na_2O/Al_2O_3.

These possibilities cover common types of chemical alteration that the rocks might have experienced during weathering. As summarized by Retallack (2001), hydrolysis involves the reaction of a mineral grain with carbonic acid in a process that removes exchangeable cations such as Ca^{2+}, Mg^{2+}, K^{1+}, and Na^{1+}. The oxidation reactions involve electron loss, generally from a Fe cation. Hydration (or dehydration) involves the addition or loss of water that is structurally part of a mineral, and salinization describes the extent to which minerals disappear from the rocks by dissolving into solution.

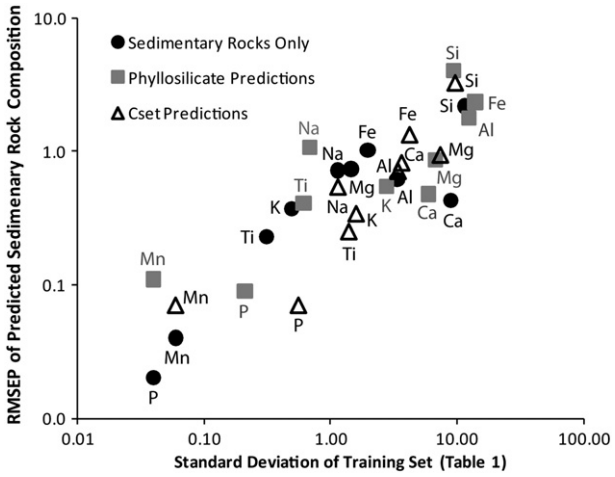

Fig. 8. The 1σ standard deviations of all XRF elemental abundances in each of the data sets used here (century set, sediments, and phyllosilicates) are plotted on the x axis vs. the RMSEP error on prediction of that element in the sedimentary rocks on the y axis using a log scale. Elements with a wide range (large 1σ) are not well-predicted, probably because the training set is more dissimilar to the samples being predicted. Elements for which the training set has only a small range that is also similar to that of the unknowns do a better job in many cases. SiO_2 is an important exception to this rule, suggesting that the distribution of values in the training sets must be fundamentally different from all the unknowns, and/or that some other underlying fundamental process is not accounted for.

In this paper, we wish to know if the prediction errors (expressed as RMSEP in Table 5) on these elements are small enough to allow these ratios to be used to understand the chemical weathering processes in a martian regolith. This question is addressed in the plots shown in Fig. 9, in which oxide units have been converted to atoms following the procedure in Retallack (2001). If a rock on Mars was to be analyzed with LIBS using the current RMSEP values as error bars, it is likely that some useful information would result. For example, the hydrolysis plot shows that the silty clay is quite similar to its parent red paleosol, and has not undergone much hydrolysis. The waxy clay, however, has relatively high Si and low Al, indicating loss of Al.

The other plots in Fig. 9 lead to similar conclusions. Most of the samples studied here have Fe/Al ratios of ~0.30, corresponding to fairly reducing environments. The hydration plot suggests an increase in Si (relative to Al and Fe) with chemical weathering, with few hydrated minerals. The salinization plots show a nearly constant of Na/Al ratio and some slight variations in Na/K, with the latter having ratios much less than the value of Na/K = 1 that would imply chemical weathering via salinization (mainly Na-zeolite precipitation).

In all of these plots, the distance between the data points is at least slightly greater than the size of the error bars. So some idea of the degree of hydrolysis, oxidation, hydration, and salinization would be obtained from the LIBS analyses, though it could not be considered as robust evidence of one mechanism over another without further data. Note also that the spectra used here were acquired at a 9 m standoff distance, essentially equivalent to near the outer limit to ChemCam's range. Signal to noise ratios are a strong function of distance (d), nearly $1/d^3$, so that even a small decrease in the distance will bring rapid improvements.

Moreover, major elements used in combination with minor elements may be even more useful in predicting provenance, following discriminant plots by such workers as Bhatia and Crook (1986), Roser and Korsch (1988) and Floyd and Leveridge (1987). Development of procedures to optimize trace and minor element analyses of LIBS data are critically important for planetary exploration (Wiens et al., 2002). Minor and trace element analyses present special challenges to PLS because of the well-known correlations between major and minor elements in geological samples that result from chemical "camouflage." PLS does not depend on knowledge of

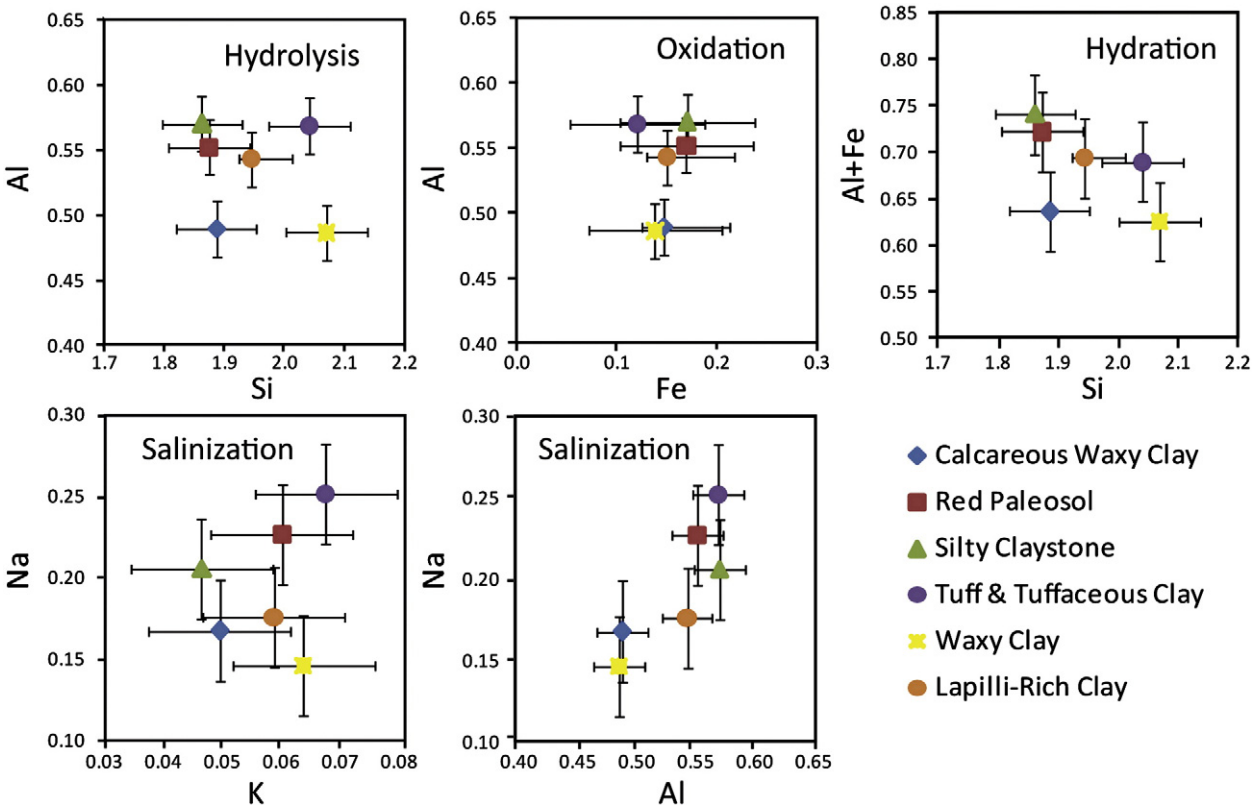

Fig. 9. Elemental ratio plots (converted to atoms as in Retallack, 2001) of the averages of samples in each rock type (from Table 2) with errors from the externally-validated PLS-1 model using the first local minimum as the number of components and the model built from the 100 igneous samples and the 17 phyllosilicates. These plots show trends associated with different types of chemical weathering in paleosols, including hydrolysis, oxidation, hydration, and salinization, after Retallack (2001).

which element produces a signal, so long as any given line correlates with an elemental abundance, and the highest regression coefficients. For example, because Rb readily assumes the crystallographic role of K via a simple substitution, PLS predictions of Rb may make use of K lines. Work by Tucker et al. (2009) shows that for many trace elements, useful lines corresponding to emission from minor elements do exist and may be used for qualitative identifications and for abundances calibrated using univariate analyses. Preliminary work by Speicher et al. (2011) also shows that some minor elements are predicted very well by PLS-1, including Ba, Cr, Ni, Rb, and Sr.

11. Implications for remote LIBS

For remote LIBS instruments such as ChemCam, this study also has broader implications. The ChemCam flight instrument was tested under Mars conditions pre-flight; spectra were acquired from ~70 standards. However, PLS results of the current study make it clear that successful quantitative predictions of elemental compositions on Mars might be improved by the use of a larger calibration set spanning a broad compositional range if the unknown samples on Mars cover a broad compositional range.

This project shows that training set selection is complicated and is dependent on the samples being studied. For our sedimentary sample suite, the lowest prediction errors result from using a training set with the broadest possible range of compositions but with external validation of samples that are similar to the unknowns. This data set thus yields results somewhat in contrast to those of Tucker et al. (2010), who obtained the best results on the Cset samples (studied herein) by using focused training sets bracketing the composition of particular rock types within that data set with internal validation. However, there are several important differences between Tucker et al. (2010) and this study that reflect our evolving ability to perform these analyses in increasingly sophisticated ways. Tucker et al. (2010) utilized the Unscrambler® software package to automatically pick the same number of regression components for all elements in each regression, in contrast with the more customized procedures (picking component numbers for each element based on internal and external validation procedures) used herein. If the Tucker et al. (2010) analyses had been done using the procedures in this paper, the results might have been different. Moreover, Tucker et al. (2010) reported precision in terms of 1σ errors, rather than using RMSEP as in the current study. These parameters are related, but not identical. The formula for standard deviation is

$$\text{Standard deviation} = \sigma = \sqrt{\frac{\sum_i (y_i - \bar{y})^2}{n-1}},$$

where y_i is the residual (true-predicted value), $\bar{y}$ is the average residual, and n is the number of samples. In Tucker et al. (2010), y_i is the residual (true-predicted value) for each sample and $\bar{y}$ is again the average residual, i.e., the difference between each of the single measurement residuals and the mean of all of them. When the average residual ($\bar{y}^i$) is zero, then the equation for σ reduces to the same equation as that for root mean square:

$$\text{Root mean square} = \sqrt{\frac{\sum_i (y_i)^2}{n-1}}.$$

The philosophical difference between these two parameters is that σ is a measure of how tightly bound the residuals around the average

residual, while RMSEP is a measure of how tightly bound the residuals are around zero. Thus, RMSEP is a measure of accuracy and σ is a measure of precision. If the average residual happens to be exactly zero, then these two parameters are exactly equivalent. For all these reasons, direct comparisons between the results of Tucker et al. (2010) and the current study are inappropriate, though the macroscopic conclusions will remain the same.

However, all these results do underscore the importance of training set selection in producing optimal LIBS prediction errors for individual elements, and suggest a need for an automated methodology to choose specialized training sets for predictions of individual elements in individual unknowns. The subjective attempts presented here and in Tucker et al. (2010) and Dyar et al. (2011a,b) show that using geological reasoning to manually select training set samples can only go so far in optimization of prediction errors. More rigorous (less subjective) statistical procedures for training set selection are needed. Unfortunately, PLS analysis is not designed to select specific samples for training sets. PLS provides high-dimensional regression capabilities and is thus useful for LIBS analysis in some applications, but it provides no framework for taking advantage of structural similarities in the data. Other high-dimensional regression techniques that combine the ability to shrink the number of input variables using projection with automatic selection of similar samples for predictions will be needed.

Ultimately, the best geological training set for predictions of chemistry in samples of completely unknown composition will be one that covers the broadest possible range of rock types and compositions. If the compositional space of the training set has no gaps, it will always have rock types similar to the unknowns being probed. From such a training set, samples that are similar to the unknown can then be selected and used for optimal predictions.

For analyses of sedimentary rocks, this work demonstrates the potential of LIBS for quantification of major elements at a significant distance from the sample. Improvements in the size and scope of laboratory calibration data along with automation of training set selection should ultimately reduce prediction errors on LIBS analyses, making this a very attractive analytical technique not only for Mars for terrestrial field instrumentation. Portable backpack LIBS units have been tested for a wide variety of applications (e.g. Cunat et al., 2005; Harmon et al., 2005; Barefield et al., 2010), and these units have the potential to obtain high quality in-situ analyses of geological materials. This study demonstrates that LIBS can discern and quantify the four competing kinds of chemical reactions in paleosols (hydrolysis, oxidation, hydration, and salinization) that provide information about the extent of chemical weathering. Results of this study show that the LIBS technique has a promising future for characterizing both chemical/biochemical and siliciclastic sedimentary rocks and understanding their evolution.

Acknowledgments

We are grateful for support from NASA grants NNG06GH35G and NNX09AL21G from the Mars Fundamental Research Program, as well as funding from the Massachusetts Space Grant Consortium for support of M.L.C and E.A.B. We thank Elly Breves for assistance with figure preparation and Steven Bender for helpful discussions.

Appendix A. Evaluation of data pre-processing steps

To confirm that data pre-processing was performed in the most optimal fashion possible, the order and need for three different steps were tested.

a. Wavelength calibration and registration. Because of non-linearity in the detectors' pixel to wavelength relationship, data output from the spectrometers does not have a consistent increment in wavelength. To correct for this, a set of intense, isolated peaks that could be clearly assigned to specific elements with known peak energies was identified in each of the three wavelength regions (Table 3). Comparison of the known peak positions with their pixel index in the experimental data enabled derivation of a 2nd-order polynomial calibration function for each of the three wavelength regions. To allow for comparison between multiple data sets that use slightly different calibration functions, we interpolated the spectral data to a standard wavelength axis with consistent intervals. The interpolation also involved smoothing, for which we chose smoothing parameters as benign as possible so as not to either lose data or introduce noise.
b. Background removal. Treating each of the three wavelength regions separately, the R package Peaks (Morháč, 2008, 2009) was used to fit and remove the baseline (mostly Bremsstrahlung) from the spectrum of each spot (five spots per sample). This step was undertaken because we tested results before and after removing the baseline and found that better predictions result from analyses done after baseline subtraction, unlike what was found with the analyses of Tucker et al. (2010). The order of these steps was determined through testing of various permutations to understand which sequent would produce the lowest possible errors on the eventual predictions of chemical compositions (Fig. A1). The procedure of normalizing the data to total intensity before analysis was also tested, and it was found to be unnecessary to achieve desirable predictions for these major element data. This result differs from those reported in Tucker et al. (2010) because that paper did not test all these permutations together in varying sequences.
c. Averaging of the five spots into one spectrum for each sample. All samples in this study, both those in the prospective training sets and the sedimentary samples themselves, have previously been analyzed by XRF as noted above as bulk samples. For this reason, we averaged the five spectra from five spots on each sample, yielding one spectrum for each sample.

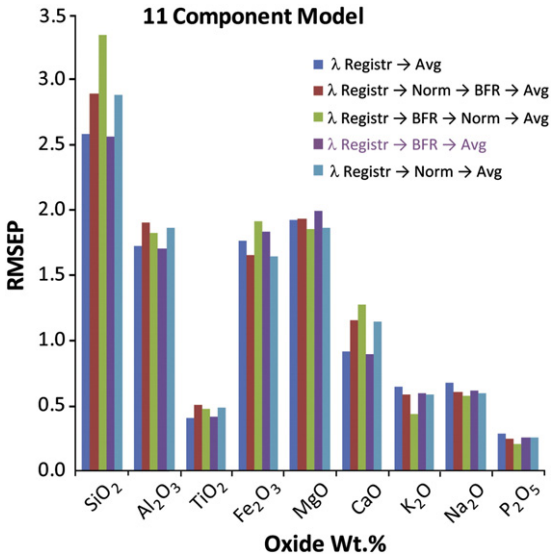

Fig. A1. Comparisons of varying steps in LIBS data pre-processing sequence using an 11 component model. The steps include registration of wavelength (λ Registr) correct for non-linearity in the detectors, normalization (Norm) to scale all variables as described in Tucker et al. (2010), baseline fitting and removal (BFR), and averaging of the five spots on each sample (Avg). Of these, the sequence that consistently gives the lowest RMSEP values is wavelength registration followed by baseline fitting and removal followed by averaging of five spectra. Normalization of compositional data does not appear to be a necessary step.

References

Alvey, D.C., Morton, K., Harmon, R.S., Gottfried, J.L., Remus, J.J., Collins, L.M., Wise, M.A., 2010. Laser-induced breakdown spectroscopy-based geochemical fingerprinting for the rapid analysis and discrimination of minerals: the example of garnet. Applied Optics 49, C168–C180. doi:10.1364/AO.49.00C168.

Ashley, G.M., Driese, S.G., 2000. Paleopedology and paleohydrology of a volcaniclastic paleosol interval: implications for early Pleistocene stratigraphy and paleoclimate record, Olduvai Gorge, Tanzania. Journal of Sedimentary Research 70, 1065–1080.

Barefield, J.E., Clegg, S.M., La Montagne, S.A., Veal, K.D., Le, L., Lopez, L., 2010. Development of laser induced breakdown spectroscopy instrumentation for safeguards applications. IAEA Symposium on International Safeguards, Vienna, Austria, IAEA-CN-184/134.

Beleites, C., Sergo, V., 2011. hyperSpec: a package to handle hyperspectral data sets in R, J. Stat. Software, R package version 0.96. http://hyperspec.r-forge.r-project.org2011URL:.

Bence, A.E., Albee, A.L., 1968. Empirical correction factors for the electron microanalysis of silicate and oxides. Journal of Geology 76, 382–403.

Bhatia, M.R., Crook, A.W., 1986. Trace element characteristics of graywackes and tectonic setting discrimination of sedimentary basins. Contributions to Mineralogy and Petrology 92, 181–193.

Bibring, J.-P., Langevin, Y., Gendrin, A., Gondet, B., Poulet, F., Berthé, M., Soufflot, A., Arvidson, R., Mangold, N., Mustard, J., Drossart, P., 2005. Mars surface diversity as revealed by the OMEGA/Mars Express observations. Science 307, 1576–1581.

Bibring, J.-P., Langevin, Y., Mustard, J., Poulet, F., Arvidson, R., Gendrin, A., Gondet, B., Mangold, N., Pinet, P., Forget, F., 2006. Global mineralogical and aqueous Mars history derived from OMEGA/Mars Express data. Science 312, 400–404.

Brimhall, G.H., Lewis, C.J., Ford, C., Bratt, J., Tylor, G., Warin, O., 1991a. Quantitative geochemical approach to pedogenesis: importance of parent material reduction, volumetric expansion, and eolian influx in laterization. Geoderma 51, 51–91.

Brimhall, G.H., Chadwick, O.A., Lewis, C.J., Compston, W., Williams, I.S., Danti, K.J., Dietrich, W.E., Power, M., Hendricks, D., Bratt, J., 1991b. Deformational mass transfer and invasive processes in soil evolution. Science 255, 695–702.

Ciucci, A., Palleschi, V., Rastelli, S., Barbini, R., Colao, F., Fantoni, R., Palucci, A., Ribezzo, S., van der Steen, H.J.L., 1996. Trace pollutants analysis in soil by a time-resolved laser-induced breakdown spectroscopy technique. Applied Physics B: Lasers and Optics 63, 185–190. doi:10.1007/BF01095271.

Clegg, S.M., Sklute, E.C., Dyar, M.D., Barefield, J.E., Wiens, R.C., 2009. Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and related techniques. Spectrochimica Acta Part B: Atomic Spectroscopy 64, 79–88. doi:10.1016/j.sab.2008.10.045.

Colao, F., Fantoni, R., Lazic, V., Spizzichino, V., 2010. Laser-induced breakdown spectroscopy for semi-quantitative and quantitative analyses of artworks—application on multi-layered ceramics and copper based alloys. Spectrochimica Acta Part B 57, 1219–1234. doi:10.1016/S0584-8547(02)00054-X.

Cremers, D.A., Radzeimski, L.J., 2006. Handbook of Laser-induced Breakdown Spectroscopy. Wiley. 302 pp.

Ctvrtnickova, T.M., Mateo, P., Yanez, A., Nicolas, G., 2009. Characterization of coal fly ash components by laser-induced breakdown spectroscopy. Spectrochimica Acta Part B 64, 1093–1097. doi:10.1016/j.sab.2009.07.032.

Cunat, J., Palanco, S., Carrasco, F., Simon, M.D., Laserna, J.J., 2005. Portable instrument and analytical method using laser-induced breakdown spectrometry for in situ characterization of speleothems in karstic caves. Journal of Analytical Atomic Spectrometry 20, 295–300.

Dyar, M.D., Tucker, J.M., Humphries, S., Clegg, S.M., Wiens, R.C., Lane, M.D., 2011a. Strategies for Mars remote laser-induced breakdown spectroscopy analysis of sulfur in geological samples. Spectrochimica Acta 66, 39–56.

Dyar, M.D., Carmosino, M.L., Tucker, J.M., Speicher, E.A., Brown, E.B., Clegg, S.M., Wiens, R.C., Barefield, J.E., Delaney, J.S., Ashley, G.M., Driese, S.G., 2011b. Error analysis for remote laser-induced breakdown spectroscopy analysis using combinations of igneous, sedimentary, and phyllosilicate samples. 42nd Lunar and Planetary Science Conference. Lunar and Planetary Institute, Woodlands, Tex. Abstr. 1258.

Edgett, K.S., Malin, M.C., 2002. Martian sedimentary rock stratigraphy: outcrops and interbedded craters of northwest Sinus Meridiani and southwest Arabia Terra. Geophysical Research Letters 29, 2179. doi:10.1029/2002GL016515.

Fairen, A.G., Chevrier, V., Abramov, O., Marzo, G., Gavin, P., Davila, A.F., Tornabene, L.L., Bishop, J.L., Roush, T.L., Gross, C., Kneissl, T., Uceda, E.R., Dohm, J.M., Schulze-Makuch, D., Rodriguez, J.A.P., Amiols, R., McKay, C.P., 2010. Noachian and more recent phyllosilicates in impact craters on Mars. Proceedings of the National Academy of Sciences 107, 12095–12100.

Floyd, P.A., Leveridge, B.E., 1987. Tectonic environment of the Devonian Gramscatho Basin, South Cornwall - Framework mode and geochemical evidence from turbiditic sandstones. Journal of the Geological Society 144, 5331–5542.

Fortes, F.J., Laserna, J.J., 2010. Characteristics of solid aerosols produced by optical catapulting studied by laser-induced breakdown spectroscopy. Applied Surface Science 256, 5924–5928. doi:10.1016/j.apsusc.2010.03.077.

Francis, P., 1993. Volcanoes: A Planetary Perspective. Clarendon Press. 443 pp.

Frank, I.E., Friedman, J.H., 1993. A statistical view of some chemometrics regression tools. Technometrics 35, 109–135.

Gendrin, A., Mangold, N., Bibring, J.-P., Langevin, Y., Gondet, B., Poulet, F., Bonello, G., Quantin, C., Mustard, J., Arvidson, R., LeMouelic, S., 2005. Sulfates in the Martian layered terrains: the OMEGA/Mars Express view. Science 307, 1587. doi:10.1126/science.1109087.

Gooding, J.L., 1992. Soil mineralogy and chemistry on Mars: possible clues from salts and clays in SNC meteorites. Icarus 99, 28–41.

Gooding, J.L., Wentworth, S.J., Zolensky, M.E., 1991. Aqueous alteration of the Nakhla meteorite. Meteoritics 26, 135–143.

Gottfried, J.L., Harmon, R.S., De Lucia, F.C., Miziolek, A.W., 2009. Multivariate analysis of laser-induced breakdown spectroscopy chemical signatures for geomaterial classification. Spectrochimica Acta Part B: Atomic Spectroscopy 64, 1009–1019. doi:10.1016/j.sab.2009.07.005.

Griffith, L.L., Shock, E.L., 1997. Hydrothermal hydration of martian crust: illustration via geochemical model calculations. Journal of Geophysical Research 102, 9135–9143.

Harmon, R.S., De Lucia, F.C., Miziolek, A.W., McNesby, K.L., Walters, R.A., French, P.D., 2005. Laser-induced breakdown spectroscopy, LIBS—an emerging field-portable sensor technology for real-time, in-situ geochemical and environmental analysis. Geochemistry: Exploration, Environment, Analysis 5, 21–28. doi:10.1144/1467-7873/03-059.

Harmon, R.S., Remus, J., McMillan, N.J., McManus, C., Collins, L., Gottfried, J.L., DeLucia, F.C., Miziolek, A.W., 2009. LIBS analysis of geomaterials: geochemical fingerprinting for the rapid analysis and discrimination of minerals. Applied Geochemistry 24, 1125–1141. doi:10.1016/j.apgeochem.2009.02.009.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, 2nd Ed. Springer Science, New York. 745 pp.

Hay, R.L., Kyser, T.K., 2001. Chemical sedimentology and paleoenvironmental history of Lake Olduvai, a Pleistocene lake in northern Tanzania. Geological Society of America Bulletin 113, 1505–1521.

Hover, V.C., Ashley, G.M., 2003. Geochemical signatures of paleodepositional and diagenetic environments: a STEM/AEM study of authigenic clay minerals from an arid rift basin, Olduvai Gorge, Tanzania. Clays and Clay Minerals 51, 231–251.

Jolliffe, I.T., 2002. Principal Component Analysis, Series: Springer Series in Statistics2nd ed. Springer, NY. 487 pp.

Maurice, S., Wiens, R., Manhès, G., Cremers, D., Barraclough, B., Bernardin, J., Bouyé, M., Cros, A., Dubois, B., Durand, E., Hahn, S., Kouach, D., Lacour, J.-L., Landis, D., Moore, T., Parès, L., Platzer, J., Saccoccio, M., Sallé, B., Whitaker, R., 2005. ChemCam instrument for the Mars Science Laboratory, MSL. Rover. 36th Lunar and Planetary Science Conference. Lunar and Planetary Institute, Houston, Tex. Abstr. 1735.

McHenry, L.J., 2005. Phenocryst composition as a tool for correlating fresh and altered tephra, Bed I, Olduvai Gorge, Tanzania. Stratigraphy 2, 101–115.

McLennan, S.M., Bell, J.F., Calvin, W.M., Christensen, P.R., Clark, B.C., de Souza, P.A., Farmer, J., Farrand, W.H., Fike, D.A., Gellert, R., Ghosh, A., Glotch, T.D., Grotzinger, J.P., Hahn, B., Herkenhoff, K.E., Hurowitz, J.A., Johnson, J.R., Johnson, S.S., Jolliff, B., Klingelhöfer, G., Knoll, A.H., Learner, Z., Malin, M.C., McSween, H.Y., Pocock, J., Ruff, S.W., Soderblom, L.A., Squyres, S.W., Tosca, N.J., Watters, W.A., Wyatt, M.B., Yen, A., 2005. Provenance and diagenesis of the evaporite-bearing Burns formation, Meridiani Planum, Mars. Earth and Planetary Science Letters 240, 95–121.

McManus, C.E., McMillan, N.J., Harmon, R.S., Whitmore, R.C., DeLucia, F.C., Miziolek, A.W., 2008. The use of laser induced breakdown spectroscopy, LIBS, in the determination of gem provenance: beryls. Applied Optics 47, G72–G79.

McMillan, N.J., McManus, C.E., Harmon, R.S., DeLucia Jr., F.C., Miziolek, A.W., 2006. Laser-induced breakdown spectroscopy analysis of complex silicate minerals—beryl. Analytical and Bioanalytical Chemistry 385, 263–271.

McMillan, N.J., Harmon, R.S., De Lucia Jr., F.C., Miziolek, A.W., 2007. Laser-induced breakdown spectroscopy analysis of minerals—carbonates and silicates. Spectrochimica Acta Part B 62, 1528–1536. doi:10.1016/j.sab.2007.10.037.

McSween Jr., H.Y., Taylor, G.J., Wyatt, M.B., 2009. Elemental composition of the Martian crust. Science 324, 736–739. doi:10.1126/science.1165871.

Mevik, B.-H., Cederkvist, H.R., 2004. Mean squared error of prediction, MSEP, estimates for principal component regression (PCR) and partial least squares regression (PLSR). Journal of Chemometrics 18, 422–429.

Mevik, B.-H., Wehrens, R., 2007. The pls package: principal component and partial least squares regression in R. Journal of Statistical Software 18, 1–24.

Morháč, M., 2008. Peaks: Peaks. R package version 0.2. http://www.slac.stanford.edu/comp/unix/package/cernroot/22312/TSpectrum.html2008URL.

Morháč, M., 2009. An algorithm for determination of peak regions and baseline elimination in spectroscopic data. Nuclear Instruments and Methods in Physics Research Section A 600, 478–487.

Morris, R.V., Klingelhöfer, G., Schröder, C., Rodionov, D.S., Yen, A., Ming, D.W., deSouza, P.A., Fleischer, I., Wdowiak, T., Gellert, R., Bernhardt, B., Evlanov, E.N., Zubkov, B., Foh, J., Bonnes, U., Kankeleit, E., Gutlich, P., Renz, F., Squyres, S.W., Arvidson, R.E., 2006. Mössbauer mineralogy of rock, soil, and dust at Gusev crater, Mars: Spirit's journey through weakly altered olivine basalt on the plains and pervasively altered basalt in the Columbia Hills. Journal of Geophysical Research 111, E02S13.

Palanco, S., López-Moreno, C., Laserna, J.J., 2006. Design, construction, and assessment of a field-deployable laser-induced breakdown spectrometer for remote elemental sensing. Spectrochimica Acta Part B: Atomic Spectroscopy 61, 88–95. doi:10.1016/j.sab.2005.12.004.

Pasquini, C., Cortez, J., Silva, L.M.C., Gonzaga, F.B., 2007. Laser-induced breakdown spectroscopy. Journal of the Brazilian Chemical Society 18, 463–512. doi:10.1590/S0103-50532007000300002.

R Development Core Team, 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org2010ISBN 3-900051-07-0, URL.

Remus, J.J., Gottfried, J.L., Harmon, R.S., Draucker, A., Baron, D., Yohe, R., 2010. Archaeological applications of laser-induced breakdown spectroscopy: an example from the Coso Volcanic Field, California, using advanced statistical signal processing analysis. Applied Optics 49, C120–C131. doi:10.1364/AO.49.00C120.

Retallack, G.J., 2001. Soils of the Past: An Introduction to Paleopedology, 2nd ed. Blackwell Science, Ltd, Oxford, UK. 404 pp.

Roser, B.P., Korsch, R.J., 1988. Provenance signatures of sandstone mudstone suites determined using discriminant function-analysis of major element data. Chemical Geology 67, 119–139.

Singer, M.J., Janitzky, P., 1986. Field and laboratory procedures used in a soil chronosequence study. United States Geological Survey Bulletin 1648 49 pp..

Sirven, J.-B., Bousquet, B., Canioni, L., Sarger, L., 2006. Laser-induced breakdown spectroscopy of composite samples: comparison of advanced chemometrics methods. Analytical Chemistry 78, 1462–1469. doi:10.1021/ac051721p.

Sirven, J.-B., Sallé, B., Mauchien, P., Lacour, J.-L., Maurice, S., Manhes, G., 2007. Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods. Journal of Analytical Atomic Spectrometry 22, 1471–1480. doi:10.1039/b704868h.

Speicher, E.A., Dyar, M.D., Carmosino, M.L., Tucker, J.M., Clegg, S.M., Wiens, R.C., 2011. Single variable and multivariate analyses of remote laser-induced breakdown spectra for predictions of Rb, Sr, Cr, Ba, S, and V in igneous rocks. 42nd Lunar and Planetary Science Conference. Lunar and Planetary Institute, Woodlands, Tex. Abstr. 2385.

Squyres, S.W., Arvidson, R.E., Bell III, J.F., Brückner, J., Cabrol, N.A., Calvin, W.M., Carr, M.H., Christensen, P.R., Clark, B.C., Crumpler, L., Des Marais, D.J., d'Uston, C., Economou, T., Farmer, J., Farrand, W.H., Folkner, W., Golombek, M.P., Gorevan, S., Grant, J.A., Greeley, R., Grotzinger, J., Haskin, L.A., Herkenhoff, K.E., Hviid, S., Johnson, J., Klingelhöfer, G., Knoll, A., Landis, G., Lemmon, M., Li, R., Madsen, M.B., Malin, M.C., McLennan, S.M., McSween Jr., H.Y., Ming, D.W., Moersch, J., Morris, R.V., Parker, T.J., Rice Jr., J.W., Richter, L., Rieder, R., Sims, M., Smith, M., Smith, P., Soderblom, L.A., Sullivan, R., Wänke, H., Wdowiak, T.J., Wolff, M.J., Yen, A.S., 2004. The Opportunity rover's Athena science investigation at Meridiani Planum, Mars. Science 306, 1698–1703.

Squyres, S.W., Knoll, A.H., 2005. Sedimentary rocks at Meridiani Planum: origin, diagenesis, and implications for life on Mars. Earth and Planetary Science Letters 240, 1–10.

Tucker, J.M., Dyar, M.D., Clegg, S.M., Wiens, R.C., Barefield II, J.E., Schaefer, M.W., Bishop, J.L., 2008. Quantitative chemistry of phyllosilicate minerals using laser-induced breakdown spectroscopy. Workshop on Martian Phyllosilicates: Recorders of Aqueous Processes. Lunar and Planetary Institute, Houston, Tex. Abstr. 7028.

Tucker, J.M., Dyar, M.D., Clegg, S.M., Schaefer, M.W., Wiens, R.C., Barefield II, J.E., 2009. LIBS analysis of minor elements in geologic samples. 40th Lunar and Planetary Science Conference. Lunar and Planetary Institute, Woodlands, Tex. Abstr. 2024.

Tucker, J.M., Dyar, M.D., Schaefer, M.W., Clegg, S.M., Wiens, R.C., 2010. Optimization of laser-induced breakdown spectroscopy for rapid geochemical analysis. Chemical Geology. doi:10.1016/j.chemgeo.2010.07.016.

Wehrens, R., Mevik, B.-H., 2007. PLS: Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR). R package version 2.1-0. http://mevik.net/work/software/pls.html2007.

Wiens, R.C., Arvidson, R.E., Blacic, J.D., Chevrel, S., Cremers, D.A., Brennetor, R., Maurice, S., Newsom, H., 2002. Critical issues in martian geochemistry involving minor and trace elements, and the applicability of laser-induced breakdown spectroscopy (LIBS). 33rd Lunar and Planetary Science Conference. Lunar and Planetary Institute, Houston, Tex. Abstr. 1348.

Wiens, R., Maurice, S., Bridges, N., Clark, B., Cremers, D., Herkenhoff, K., Kirkland, L., Mangold, N., Manhes, G., Mauchien, P., McKay, C., Newsom, H., Poitrasson, F., Sautter, V., d'Uston, L., Vaniman, D., Shipp, S., 2005. ChemCam science objectives for the Mars Science Laboratory (MSL) rover. 36th Lunar and Planetary Science Conference. Lunar and Planetary Institute, Houston, Tex. Abstr. 1580.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laborary Systems 58, 109–130.