Advanced signal processing analysis of laser-induced breakdown spectroscopy data for the discrimination of obsidian sources

Obsidian is a natural glass of volcanic origin and a primary resource used by indigenous peoples across North America for making tools. Geochemical studies of obsidian enhance understanding of artifact production and procurement and remain a priority activity within the archaeological community. Laser-induced breakdown spectroscopy (LIBS) is an analytical technique being examined as a means for identifying obsidian from different sources on the basis of its 'geochemical fingerprint'. This study tested...
Advanced signal processing analysis of laser-induced breakdown spectroscopy data for the discrimination of obsidian sources

ABSTRACT

Obsidian is a natural glass of volcanic origin and a primary resource used by indigenous peoples across North America for making tools. Geochemical studies of obsidian enhance understanding of artifact production and procurement and remain a priority activity within the archaeological community. Laser-induced breakdown spectroscopy (LIBS) is an analytical technique being examined as a means for identifying obsidian from different sources on the basis of its ‘geochemical fingerprint’. This study tested whether two major California obsidian centers could be distinguished from other obsidian localities and the extent to which subsources could be recognized within each of these centers. LIBS data sets were collected in two different spectral bands using a Nd:YAG 1064 nm laser operated at approx. 23 mJ, a Czerny–Turner spectrograph with 0.2–0.3 nm spectral resolution and a high performance imaging charge couple device (ICCD) detector. Classification of the samples was performed using partial least-squares discriminant analysis (PLSDA), a common chemometric technique for performing statistical regression on high-dimensional data. Discrimination of samples from the Coso Volcanic Field, Bodie Hills, and other major obsidian areas in north-central California was possible with an accuracy of greater than 90% using either spectral band.
Continuation for Block 13

ARO Report Number 60674.2-EV
Advanced signal processing analysis of laser-ind ...

Block 13: Supplementary Note
© 2012. Published in Applied Optics, Vol. Ed. 0 51, (7) (2012), (, (7). DoD Components reserve a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so (DODGARS §32.36). The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

Approved for public release; distribution is unlimited.
Advanced signal processing analysis of laser-induced breakdown spectroscopy data for the discrimination of obsidian sources

Jeremiah J. Remus,1 Russell S. Harmon,2 Richard R. Hark,3,* Gregory Haverstock,4 Dirk Baron,5 Ian K. Potter,3 Samantha K. Bristol,3,6 and Lucille J. East7

1Department of Electrical and Computer Engineering, Clarkson University, Potsdam, New York 31699 USA
2Department of Marine, Earth & Atmospheric Sciences, North Carolina State University, Raleigh, North Carolina 27607 USA
3Department of Chemistry, Juniata College, Huntingdon, Pennsylvania 16652 USA
4Bureau of Land Management, Bishop Field Office, Bishop, California 93514 USA
5Department of Geology, California State University, Bakersfield, California 93311 USA
6Currently at the Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa 50011 USA
7Applied Spectra, Inc., 46661 Fremont Boulevard, Fremont California 94538 USA
*Corresponding author: hark@juniata.edu

Received 17 October 2011; accepted 9 December 2011; posted 10 January 2012 (Doc. ID 156647); published 9 February 2012

Obsidian is a natural glass of volcanic origin and a primary resource used by indigenous peoples across North America for making tools. Geochemical studies of obsidian enhance understanding of artifact production and procurement and remain a priority activity within the archaeological community. Laser-induced breakdown spectroscopy (LIBS) is an analytical technique being examined as a means for identifying obsidian from different sources on the basis of its ‘geochemical fingerprint’. This study tested whether two major California obsidian centers could be distinguished from other obsidian localities and the extent to which subsources could be recognized within each of these centers. LIBS data sets were collected in two different spectral bands (350 ± 130 nm and 690 ± 115 nm) using a Nd:YAG 1064 nm laser operated at ~23 mJ, a Czerny–Turner spectrograph with 0.2–0.3 nm spectral resolution and a high performance imaging charge couple device (ICCD) detector. Classification of the samples was performed using partial least-squares discriminant analysis (PLSDA), a common chemometric technique for performing statistical regression on high-dimensional data. Discrimination of samples from the Coso Volcanic Field, Bodie Hills, and other major obsidian areas in north-central California was possible with an accuracy of greater than 90% using either spectral band. © 2012 Optical Society of America

1. Introduction

Obsidian is a natural glass of volcanic origin that has been used a primary resource for making stone tools in many areas worldwide. Obsidian glass fractures conchoidally to produce sharp-edged tools and frequently was traded over great distances across North America by indigenous peoples. Obsidian sources tend to occur in relatively discrete locations that are geochemically distinct (e.g., see [1]), although geological erosion can produce fluvial deposits in which obsidian is present in clasts of sufficient size and quality to be worked for tools. Geochemical studies of obsidian can provide an important means of understanding artifact production and procurement [1–5].
Obsidian sources are widespread across the southwestern United States and many are present in eastern California and western Nevada [9–11]. Determining the provenance of obsidian artifacts remains a priority activity within the archaeological community. Variations in trace element composition have been used in this context as a basis for establishing a chemical signature that can be utilized as a means to distinguish obsidian from different sources [12,13], but the analytical techniques needed for such analysis (XRD, INAA, & ICP-MS) are time consuming and require expensive laboratory instrumentation. Laser-induced breakdown spectroscopy (LIBS) is a comparatively low-cost technique that can simultaneously detect major and trace elements and has the potential to be used in the field for in situ analysis of artifacts in real time. This study examined whether two major California obsidian centers, the Bodie Hills and Coso Volcanic Field (CVF), could be distinguished from other obsidian localities and the extent to which subsources could be recognized within each of these centers.

2. Obsidian in North-Central California

Obsidian is found extensively across north-central California (Fig. 1). Ten obsidian localities have been recorded on the east side of the Sierra Nevada mountain range in Inyo and Mono Counties. Of these, we have sampled Coso, Saline Valley, Mono-Inyo, and Bodie Hills localities. Eight localities are located in the Coast Ranges north of San Francisco in Marin, Sonoma, and Napa Counties. Of these, we have analyzed samples from the Annadel, Napa Glass Mountain, and Franz Valley locations. Twenty-four localities are known to the far north in Butte, Tehama, Plumas, Shasta, Lassen, Siskiyou, and Modoc Counties. Of these, we have samples from five sites within the Medicine Lake Volcanic Center.

The Bodie Hills are located on the western margin of the Great Basin physiographic region approximately 30 kilometers north of Mono Lake in Mono County, California. This area contains one of the most archaeologically significant obsidian sources of California prehistory, with obsidian artifacts of Bodie Hills type identified in archaeological deposits throughout Northern and Central California and as far west as the Pacific coast [15,16]. In addition to its substantial areal distribution, the Bodie Hills obsidian source has a well-documented period of utilization that began during terminal-Pleistocene/early-Holocene time and continued through the contact period [17].

The underlying geologic structure of the Bodie Hills is comprised of Tertiary volcanics intruding onto a Paleozoic and Mesozoic basement. The extant topography reflects a complex geologic landscape shaped primarily by volcanism and episodic erosion during the Oligocene, Miocene, and Pliocene epochs. Early Tertiary erosion resulted in significant unconformity between pre-Tertiary deposits and those that followed, while the majority of formative events occurred during Pliocene time [18–20]. The abundance of obsidian deposits in the Bodie Hills region was noted in a California State Mining and Mineral report as early as 1888 [21] and again by Meighan in 1955 [22], although most of the contemporary attention directed towards the area has focused on its economic geology. Natural obsidian occurrences in the Bodie Hills area take two forms, either as discrete terrace outcrops eroding from step hillsides or as fluvially/alluvially deposited lag flows [17,23].

The Bodie Hills obsidian source was first described geochemically by Jack and Carmichael in 1969 [24] and archaeologically by Singer and Ericson in 1977 [16]. Singer and Ericson identified the ‘minimal’ spatial extent of the geological obsidian deposit, noted the variation in obsidian macroattributes, and proposed a utilization curve based on an obsidian hydration analysis of what they described as the main quarry area. That study identified the main source as containing eight square kilometers (1462 acres) of culturally modified material derived from three primary outcrops. Subsequent research by Halford [23] identified 11 additional primary outcrops in two loci, termed Bodie Hills North and Bodie Hills West, as well as a substantial cobble flow trailing from them. Field surveys during those studies identified 2215 acres (8.96 km²) of previously unreported obsidian deposit. In total, 3677 acres (14 km²) of flake-stone-viable obsidian deposits from 14 primary outcrops have been identified within the Bodie Hills.

Singer and Ericson [16] also proposed a bell-shaped utilization curve for the Bodie Hills quarry with a zenith occurring during the early- and late-Holocene time (~2500 YBP). Their hydration sample was selected entirely from within the main quarry area. In contrast to Singer and Ericson’s proposed normal distribution use curve, obsidian hydration values from a random sample (n = 131) of archaeological specimens selected from both the lag deposit and discrete outcrops produced a bimodal and negatively skewed utilization curve with peaks in the early- and late-Holocene [23]. This bimodal utilization curve

![Fig. 1. The major obsidian fields of north-central California [14] sampled in this study.](image-url)
is supported by other studies which were conducted along the far western edge of the cobble flow; those research efforts also identified a substantial early-Holocene use of the obsidian cobble flow [25].

Singer and Ericson [16] did not attempt to segregate obsidian subsources within the Bodie Hills deposit, although they did note differences in obsidian macroattributes and reasoned that these were likely reflective of variation in trace chemical composition. To date there have been no attempts to identify geochimically discrete subsources within the Bodie Hills obsidian source. This current study includes geologic samples from seven of the 14 culturally targeted primary outcrops identified to date in an attempt to distinguish the trace chemical variation within the deposit.

The CVF of California lies at the west edge of the Basin-and-Range physiographic province, some 280 km southeast of Mono Lake. The CVF contains at least 38 high-silica rhyolite volcanic extrusions of late Pleistocene age [26] that occur most commonly as steep-sided domes and less frequently as lava flows erupted onto pre-Cenozoic basement rocks. K-Ar age dates and geochemical profiles of the rhyolite have been obtained on the domes and flows [26,27]. These studies have identified seven chemical groups of rhyolite that erupted at 1047 ± 20 Ka, 587 ± 18 Ka, 235 ± 25 Ka, 170 ± 11 Ka, 160 ± 30 Ka, 89 ± 10 Ka, and 63 ± 9 Ka (1 Ka = 10^2 years BP). Across the CVF, rhyolite is of extremely uniform appearance, so that it is not possible to attribute a specimen to a source by visual inspection. Occasionally, lava domes amalgamate to form compound structures, of which the Sugarloaf Mountain complex is the largest. Many of the Coso rhyolite localities and pyroclastic deposits contain workable obsidian that has been quarried for tools by the indigenous population for more than 12000 years [28].

Several studies have attempted to delineate Coso obsidian subsources on the basis of geochemistry, each generating a different classification scheme depending on the approach taken. The initial and still definitive study of Bacon et al. [27] divided the volcanic features of the CVF into seven groups on the basis of both K-Ar geochronology and geochemical character. Importantly, these authors note that the dome and flow surface morphology, geological field relationships, and age dating results indicate that each rhyolite group consists of essentially coeval extrusions that occurred in times spans that were very short compared to the overall life of the Coso magmatic system. Hughes [12] used X-ray fluorescence analyses of the incompatible elements Rb and Zr to define four geochemical subsources: Joshua Ridge, West Cactus Peak, West Sugarloaf, and Sugarloaf Mountain. Bouey [29] later suggested that the Sugarloaf Mountain and West Sugarloaf subsources might not be as readily distinguishable as proposed by Hughes [12]. Ericson and Glascock [13] used instrumental neutron activation analysis data for 14 trace elements to confirm the four subsources identified by Hughes [12] and postulated two additional subsources, a conclusion more in line with the initial results of Bacon et al. [27] based on K-Ar dating, field relationships, and whole-rock geochemical analysis. Eerkens and Rosenthal [30] question the idea of archaeological subsources within obsidian fields, concluding that the concept has merit when large datasets encompassing spatially expansive areas can be assembled and statistically analyzed. Such was the approach taken by Draucker [31], who used laser ablation ICP/MS analysis data for 16 major and trace elements and stepwise multielement discriminant analysis to identify Coso subsources. Four distinct obsidian types were recognized and confirmed the Joshua Ridge and East Sugarloaf groups from Hughes [12]. However, the other two groups identified are a West Cactus group and a West Sugarloaf group. The West Cactus group includes Hughes’ West Cactus locations and the newly identified Steward Quarry site that had not been sampled for previous studies. The West Sugarloaf group includes samples from the South Sugarloaf and Southeast Sugarloaf sites on the older South Sugarloaf Mountain, which is part of Hughes [12] West Sugarloaf group but includes an additional West Sugarloaf site. Remus et al. [32] analyzed the sample set used by Draucker [31] by LIBS and confirmed her classifications.

In this study, a total of 170 obsidian samples were analyzed as follows: 37 samples from eight Coso sites, 86 samples from seven Bodie sites, and 47 samples from four other California obsidian areas (Mono-Inyo Volcanic Center, the Saline Valley region, the North Coast Range, and the Medicine Lake Volcanic Center). Given the focus of the effort to assess if subsources could be recognized, more samples were analyzed from the Coso and Bodie localities than for any other single obsidian source.

3. Analytical Methodology, Signal Processing, and Statistical Analysis

Fresh obsidian chips of samples from six different California obsidian areas (Table 1) were analyzed using an RT100-HP commercial LIBS instrument (Applied Spectra, Inc.) equipped with a 50 mJ Nd:YAG 1064 nm laser, a Czerny–Turner spectrograph with a 600 g/mm grating providing 0.2–0.3 nm spectral resolution, and a high performance ICCD detector. The experimental parameters included a gate delay of 1 μs, gate width of 3 μs, ~23 mJ laser power and a repetition rate of 3 Hz. For each sample, a single location was analyzed using 50 cleaning shots followed by collection of 50 single-shot LIBS spectra at each of the two wavelength regions: (350 ± 130 nm and 690 ± 115 nm).

Three distinct classification tasks were performed using the available LIBS spectral data. In the first task, the 170 samples were categorized either as a sample from (i) Coso, (ii) Bodie Hills, or (iii) Another Location. The second task considered just the samples from the Coso sites, which were categorized according to one of the eight possible subsources.
of 100%. The two data sets were fused by taking the highest-confidence classification from the two data sets (i.e., decision-level fusion) to produce estimates of the class label for each sample.

4. Results and Discussion

The idea being tested in this and other recent papers [32,35–38] is that of LIBS ‘geochemical fingerprinting’. The concept is that the full LIBS broadband spectrum, or a sufficiently large portion thereof, contains sufficient compositional information to provide a unique chemical description of any particular sample. Thus, if advanced statistical signal processing and classification techniques are applied to a sufficiently robust spectral data set, it should be possible to distinguish samples of the same kind originating from one place from those originating in another. The geological basis for this hypothesis is that the Earth is composed of rocks of different composition and is compositionally heterogeneous, both horizontally and vertically, so that minerals and rocks originating in the crust will directly inherit the chemical signature of that crust. Similarly, bodies of water and hydrothermal fluids will differ in composition from one place to another, so that minerals precipitated from solution at one place under a particular set of geochemical boundary conditions should be readily distinguishable from those formed elsewhere on the basis of their LIBS spectral signature. We demonstrated this concept for different kinds of minerals in previous studies [32,34–37] and extended the idea to obsidian from the CVF [31]. Obsidian is a particularly challenging material for geochemical fingerprinting because obsidian is a high-Si, rhyolitic glass that tends to have similar bulk composition wherever it is found. It is only on the basis of trace element compositions that obsidian of different provenance can be distinguished. Thus the problem of distinguishing obsidian sources and subsources should be particularly challenging for LIBS, given that elemental detection limits for LIBS are generally in the low parts per million range.

The results of the first classification task, which required discriminating Coso versus Bodie Hills versus Other, are shown in Fig. 2. In each subplot, the classification results for the 350 nm and 690 nm data sets are compared as a function of the number of components used in the partial least-squares decomposition.
(ranging from 5 to 35). Results are shown for PLSDA runs on both the raw, individual spectra (left) and the sets of averaged spectra (right). Classification performance plateaus near 90% correct, which is well above chance-level performance for this classification task. For larger numbers of PLSDA components (i.e. greater than 20), there appears to be little difference between the 350 nm and 690 nm data sets, with slightly higher classification scores (approximately 1%) using the averaged spectra.

In Fig. 3, the results from Fig. 2 are reproduced in four separate subplots with the results for the stitched spectra overlaid for comparison. The top row of subplots shows results using the individual spectra whereas results in the bottom row of subplots use the averaged spectra. Results are again shown as the number of PLSDA components is varied from 5 to 35. For this classification task, the stitched spectra fail to provide a noticeable improvement over either of the separate spectral bandwidths.
Figures 4 and 5 show results for the within-Coso subsource classification task presented in a format similar to Figs. 2 and 3. From Fig. 4, it can be seen that classification accuracy greater than 60% correct is possible using the 350 nm data set with averaged spectra and optimal parameterization of PLSDA. In Fig. 5, the comparison of results for the stitched spectra and separate spectral bands does not indicate a benefit to using stitched spectra on this data set.

Figure 6 illustrates results for the within-Bodie Hills location classification. There are seven different sampling locations for which the labels were estimated (chance-level performance = 16.7%) and the results shown in Fig. 6 indicate that this is the most...
difficult of the three classification tasks. The four variations of data processed using PLSDA suggest a peak performance level of approximately 30% correct classification with no consistent improvement through the use of amalgamated spectra.

As an alternative to the combined spectra for fusion of the 350 nm and 690 nm data sets at the feature-level, fusion at the decision level was also explored. In the decision-level fusion, each of the two data sets provides an estimate of the label for each sample, with a confidence based on the proportion of spectra collected from that sample that were in agreement with the assigned label. The final classification for the sample is based on the most confident classification from the 350 nm and 690 nm data sets. Table 2 shows the percent correct classification for the separate data sets, the stitched spectra, as well as decision-level fusion results, for the three classification tasks using both individual shots and averaged shots. For each of the experiments, the decision-level fusion consistently provides a measurable increase in classification performance over the separate spectral bands and outperforms the stitched spectra in all but one experiment. Thus, there may be some benefit to providing additional information in the form of LIBS measurements in multiple spectral bands using decision-level fusion.

These results did not appear to be sensitive to the presence of peaks associated with the major elements in obsidian. The values for percent correct classification in Table 2 changed by only ±1% in most cases (although a few of the results were significantly worse) after removing seven wavelength ranges (248–255, 262–264, 285–290, 392–400, 404–405, 585–592, and 764–772 nm) associated with prominent Na, Al, Si, and K lines.

5. Summary and Conclusion

In this study, the feasibility of using LIBS for discriminating between obsidian samples from two sources in California, as well as discriminating between locations within those sources, was examined using data sets collected in two different spectral bands (350 ± 130 nm and 690 ± 115 nm). Discrimination between samples from the CVF, Bodie Hills, and other major obsidian areas in north-central California was possible with a high degree of accuracy (greater than 90%) using data collected in either spectral band. However, the separation of samples from the sublocalities within each source proved much more challenging. In particular, the poor discrimination between samples from Bodie Hills may indicate the labeled locations cannot be isolated based on the chemical composition of the samples, suggesting that the obsidian material at Bodie Hills is derived from a single obsidian source.

A primary interest in this study was to determine whether the classification performance would benefit from the inclusion of information from both the spectral bands centered at 350 nm and 690 nm, respectively, or if a single spectral band provided the maximum achievable classification performance.
As part of the investigation, two methods for fusion of the information from the two spectral bands were considered: a feature-level fusion where spectra were ‘stitched’ together to form an amalgamated spectrum and the PLSDA models were generated using the resulting composite spectra and a decision-level fusion approach where the data from the two spectral bands was processed separately with different PLSDA models, labels for each sample were estimated, and a final label estimate was generated from the decisions made by the PLSDA models operating on the two separate data sets. The results presented in this study indicate slightly higher performance from fusion of the two datasets, with decision-level fusion outperforming feature-level fusion in most experiments. One reason why the feature-level fusion may not have provided a greater level of benefit is that the order for pairing spectra from the 350 nm and 690 nm data sets is arbitrary. The 350-nm- and 690-nm-centered spectra were not collected simultaneously, but rather sequentially from separate firings of the laser, thus corresponding to different sets of plasma events. The feature-level fusion would be able to take advantage of dependence between wavelengths in the 350-nm-centered and 690-nm-centered data; however, that is only expected in a scenario with simultaneous collection of the spectra such that they are observing the same plasma event, as in a multichannel, broadband LIBS analytical approach.

The authors wish to thank Applied Spectra, Inc. for technical support and the loan of instrumentation, the II-VI Foundation (RRH) and ARO Grant 60674-EV (IP) for financial support, and Peter Torrione and Kenneth Morton of New Folder Consulting, Durham, NC for chemometric software. Partial financial support for this research was provided by an ARL Fellow stipend to RSH.

References


