

# Digitization of X-ray Fluorescence Data from the Lawrence Berkeley National Laboratory *Nuclear Archaeology* Program 1: South American Obsidian

Matthew T. Boulanger\*and Nicholas Tripcevich†

February 27, 2021

In the mid-2020 Richard Burger provided a box of paper records from his collaborative research with the Lawrence Berkeley National Laboratory's *Nuclear Archaeology* Program (LBNL) led by Frank Asaro and Helen Michel. These records consisted primarily of data printouts of from analyses of Peruvian obsidian artifacts and source samples by neutron activation analysis (NAA) and by X-ray fluorescence (XRF). Previously, the senior author devoted significant efforts to digitize similar paper records from the LBNL. Nearly all of the records involved in that effort were from NAA work—including Burger's obsidian analyses—making the NAA printouts in this newly provided set of records redundant. These NAA records will not be discussed further.

The data contained in these printouts resulted from a collaborative during the mid-late 1970s. As recounted in Burger and Asaro (1978), Burger had previously generated XRF data for Peruvian obsidian using an XRF spectrometer in the Department of Geology at UC-Berkeley. However, the data produced using a nondestructive methodology with this instrument could not distinguish between major Peruvian obsidian sources. Burger then began collaborating with Asaro at the LBNL to analyze obsidian using NAA. Using insights gained regarding the chemistries of the various obsidian sources as determined by NAA, Burger and Asaro developed an XRF methodology at LBNL that was “rapid, non-destructive, and inexpensive.” Details of the methodology are published in Burger and Asaro (1979). Of significant importance to the preservation of data from Asaro's and Michel's work, however, is that nearly all of the records describing output from analyses by XRF had not previously been encountered. Moreover, few if any of the individual elemental abundances had ever been published except in aggregated form (i.e., means and standard deviations for source groupings). Given the condition of these paper records after more than 40 years, we set out to produce archival quality scans of all the documents, and to tran-

---

\*Department of Anthropology, Southern Methodist University

†University of California–Berkeley

scribe the elemental abundances recorded on them to preserve these data for future use.

## 1 Description of the Records

Records provided by Burger to Tripcevich consist of a series of  $11 \times 15"$  fan-fold dot-matrix printouts (Figure 1). The laboratory would typically produce two sets of these printouts for each XRF Run, or series of assays: One reporting estimated concentrations of mid-Z elements (Fe, Cr, Mn, Ti, Ca, V, Zn, Cu, Pb, Rb, Sr, Y, Zr, Nb, Mo, Ni, and K), and one reporting estimated concentrations of high-Z elements (Sn, Ba, La, and Ce). The Burger records contain at least one of these two printouts for the following sequentially numbered XRF Runs: 8031, 8033, 8057, 8058, 8059, 8060, 8061, 8062, 8063, 8064, 8065, 8068, and 8074. Some of these printouts are exact duplicates—meaning that, for example, two printouts of the same mid-Z Run are included. A set of photocopies of the original data printouts, reduced to  $8.5 \times 11"$ , is also present in the archive. All of these photocopies are duplicates of the  $11 \times 15"$  printouts, except for Run 8068 for which no  $11 \times 15"$  printouts have been located. Only the mid-Z printout is present for Runs 8058 and 8068. Only the high-Z printout is present for Run 8057.

The data printouts are arranged such that each Run begins with a summary of the specimens assayed. This summary includes the date on which the Run was assayed, the Run number (identified as a “Bomb” on the printouts), and various other information including a listing of the specimens included in the Run. Following this summary, elemental abundances calculated for each specimen are listed, one specimen per page. Most Run printouts in the archive include a background measurement, a measurement for a blank (the plastic used to hold specimens), and some standard reference material. In the Runs included in this archive the Perlman/Asaro Standard Pottery and a specimen from the El Chayal obsidian source are used. Both of these reference materials were well characterized by NAA at Berkeley and other reactors. Their inclusion should help to produce intra- and interlaboratory correction factors to facilitate comparisons with other compositional datasets.

Data on the pages are arranged in a tabular format, with each column representing a particular measured or derived variable, and each row representing a particular element or compound of interest. Most of these columns deal with technical aspects of determining elemental abundances from spectral data: Total photon counts, background counts, approximate peak location, real peak location, etc. While these are useful to have access to for historical purposes, we do not consider them further. Instead, we focus on the final four columns of data on each sheet (Figure 2). These include the calculated elemental abundance, the estimated uncertainty in that calculation, and the scaling factor (given as an exponent of 1,000,000). We also record the LBNL sample identification number, which consists of a four-digit sequential number representing each Run, along with a single character, number, or symbol representing the sequential position

8031 H BUR-16Q CS5 CHUPAS, YAUCUCHO, EIP  
 GAMMA SPECTRUM-B 5730+4

THE IN {23.11KEV} PEAK HAS A HALFWIDTH OF 8.17CHANNELS  
 STD NUMBER 1 -573030 A SAMPLE WEIGHT = -0. MG DEAD TIME = .68 O/O EOB = 0+ MJD  
 IRRADIATION TIME = J. MIN DECAY TIME = #325.29610 DAYS COUNT TIME = 79.993 MIN C/SEC BEG. = 0 C/SEC END = 0  
 START TIME = 43325.29610 MJD PILL THICKNESS = -0. MILS SPECTRUM BEGAN 6/30/1977 PST

NUCLIDE	COUNTS EL	COUNTS REMOVED FROM PEAK	MULT	GROSS COUNTS	BKGD	APPR REAL	I	FLUX(N/MIN-CM2)	CPM	ELEMENT	ELEMENT	
INCHE				7236	127417	104	387	5.137 +/- .016E 5	558245.0	.109 +/- .3249E 1		
CCH				129655	43708#0412	415		1.826 +/- .012E 3	1539.6	1.009 +/- .011E 0	01	
FE				12243	1672#0105	104		4.011 +/- .031E 4	189.4	4.721 +/- .066E -3	FE	
CR				958	1672#0105	84		1.965 +/- .015E 4	4.4	-.217 +/- .3905E -5	CR	
MN	12	FE	100120	0 CR	.11000	1644	1177#5878	.94	2.599 +/- .020E 4	8.1	3.131 +/- .408E -6	MN
FE PEAK IS AT CHANNEL 103.78 WITH HALFWIDTH OF 4.55. CR PEAK IS SUMMED STARTING -1.20 CHANNELS HIGHER.												
TI				1409	1307#9604	67		3.610 +/- .028E 3	1.8	5.061 +/- .3866E -6	TI	
CA				1080	1522#4343	50		1.304 +/- .010E 3	6.2	4.768 +/- .801E -3	CA	
V				978	1063#9704	76		8.824 +/- .067E 3	-1.5	-1.725 +/- .877E -4	V	
ZN				1045	1420#3930	148		1.404 +/- .012E 5	8.0	4.968 +/- .1073E -5	ZN	
CU				1411	1365#0284	136		1.163 +/- .009E 5	.8	.708 +/- .1342E -5	CU	
PB				2680	1260#0284	28		1.304 +/- .020E 4	1.420	+.321E -6	PB	
RB				8175	2481#0230	242		7.681 +/- .333E 5	102.0	1.328 +/- .0166E -6	RB	
SR				6388	1260#0284	28		1.304 +/- .020E 4	5.317	+.263E -5	SR	
Y	940 R9	.16500	0	4211	2608#0533	273		1.373 +/- .028E 6	11.9	8.656 +/- .1.819E -6	Y	
ZR	480 SR	.16500	0 PB	11934	11934#0683	190		1.304 +/- .020E 4	.393	+.312E -5	ZR	
NB	101 Y	.15200	0	5820	3696#0865	306		1.940 +/- .028E 6	36.2	1.888 +/- .153E -5	NB	
MO	1032 ZT	.15900	0	7234	613#0052	326		2.472 +/- .020E 6	.043	+.1.412E -6	MO	
NI	PEAK IS AT CHANNEL 103.78 WITH HALFWIDTH OF 4.55.	NI PEAK IS SUMMED STARTING 20.00 CHANNELS HIGHER.		564	686#0200	125		5.411 +/- .041E 4	-2.1	-3.973 +/- .1.902E -5	NI	
K	FE PEAK IS AT CHANNEL 103.78 WITH HALFWIDTH OF 4.55.	K PEAK IS SUMMED STARTING -63.50 CHANNELS HIGHER.		2657	1747#0365	43		4.011 +/- .031E 4	16.3	4.064 +/- .452E -4	K	
	FE PEAK IS AT CHANNEL 103.78 WITH HALFWIDTH OF 4.55.	K PEAK IS SUMMED STARTING -63.50 CHANNELS HIGHER.										
	COUNT RATE CORRECTION FOR LAST ELEMENT =	I										

Figure 1: Example of the LBNL data output for analysis of obsidian by non-destructive X-ray fluorescence.

A		B		C		D		E		F	
8031 H	BUR-16J CS5 CHUPAS, AYACUCHO, EIP										
8031 H	BUR-16J CS5 CHUPAS, AYACUCHO, EIP										
THE IN (23.1KEV) PEAK HAS A HALFWIDTH OF 8.17CHANNELS											
STD NUMBER 1	-573030 A	SAMPLE WEIGHT =	-0.	MG DEAD TIME =	.68 0/0	EDB =	0.	KID			
IRRADIATION TIME = .0.	MIN DECAY TIME = #325.29610 DAYS	COUNT TIME = #	79.993 MIN	C/SEC BEG. =	0.	C/SEC END =	0				
START TIME = 43325.29610 MJD	PILL THICKNESS = -0.	MILS	SPECTRUM BEGAN	6/30/1977 PST							
NUCLIDE	COUNTS REMOVED FROM PEAK	GROSS COUNTS EL	APPR PEAK COUNTS EL	I MULT	FLUXIN/MIN-CN2)	CPM	ELEMENT	ELEMENT			
INCHE	127417 388	5.137 +/- .016E 5	5 558245.0			1.109 +/- .010E 5	3.249E -5				
CCM	129655 43708#0412 415	1.526 +/- .012E 3	3 1539.6			1.009 +/- .009E 3	.011E 0				
FE	12243 1672#0105 104	4.011 +/- .031E 4	4 189.2			5.721 +/- .064E -5	.064E -5	FE			
CR	699 699#0000 0	1.000 +/- .000E 0	0 0			4.917 +/- .030E -5	.030E -5	CR			
FE PEAK IS AT CHANNEL	103.78 WITH HALFWIDTH OF	4.455.	CR PEAK IS SUMMED STARTING	-21.60	CHANNELS HIGHER.						
MN	13 FE .01020 0 CR .11000	1644. 117#0974 94	2.599 +/- .020E 4	8.1	3.131 +/- .408E -6	MN					
FE PEAK IS AT CHANNEL	103.78 WITH HALFWIDTH OF	4.455.	MN PEAK IS SUMMED STARTING	-12.60	CHANNELS HIGHER.						
TI	101 TI .01020 0 CR .11000	1404. 117#0974 97	2.599 +/- .020E 4	8.1	3.131 +/- .408E -6	TI					
FE PEAK IS AT CHANNEL	103.78 WITH HALFWIDTH OF	4.455.	TI PEAK IS SUMMED STARTING	-39.60	CHANNELS HIGHER.						
CA	1869 152#09434 50	1.304 +/- .010E 3	6.2 4.768 +/- .080E -3	CA							
FE PEAK IS AT CHANNEL	103.78 WITH HALFWIDTH OF	4.455.	CA PEAK IS SUMMED STARTING	-56.60	CHANNELS HIGHER.						
V	9734 1672#07050 148	1.373 +/- .028E 6	6 11.9 8.656 +/- .1.819E -6	V							
ZN	1901 1456#03939 148	1.604 +/- .012E 5	8.0 4.968 +/- .1.073E -5	ZN							
FE PEAK IS AT CHANNEL	103.78 WITH HALFWIDTH OF	4.455.	ZN PEAK IS SUMMED STARTING	39.40	CHANNELS HIGHER.						
CU	1411 137#02828 0	1404. 117#0974 95	2.599 +/- .020E 4	8.1	3.131 +/- .408E -6	CU					
FE PEAK IS AT CHANNEL	103.78 WITH HALFWIDTH OF	4.455.	CU PEAK IS SUMMED STARTING	-40.40	CHANNELS HIGHER.						
PB	2642 327#08015 229	8.022 +/- .061E 4	-11.4 -1.420 +/- .321E -6	PB							
PEAK IS AT CHANNEL	414.59 WITH HALFWIDTH OF	6.688.	PB PEAK IS SUMMED STARTING	#0.100	CHANNELS HIGHER.						
RB	8175 2481#0230 242	7.681 +/- .333E -5	102.4 -1.328 +/- .066E -6	RB							
PEAK IS AT CHANNEL	414.59 WITH HALFWIDTH OF	6.688.	RB PEAK IS SUMMED STARTING	-40.40	CHANNELS HIGHER.						
SR	9124 2213#03930 259	9.808 +/- .1.198E -5	92.1 5.317 +/- .263E -5	SR							
Y	940 Rb .1.6500 0	4211. 2608#0535 273	1.373 +/- .028E 6	6 11.9 8.656 +/- .1.819E -6	Y						
PEAK IS AT CHANNEL	414.59 WITH HALFWIDTH OF	6.688.	Y PEAK IS SUMMED STARTING	#0.100	CHANNELS HIGHER.						
ZR	101 ZR .1.6500 0 PB .05500	11931. 4559#0685 290	1.384 +/- .020E 6	116.3 8.393 +/- .312E -5	ZR						
NB	PEAK IS AT CHANNEL	414.59 WITH HALFWIDTH OF	6.688.	ZR PEAK IS SUMMED STARTING	#0.150	CHANNELS HIGHER.					
101 Y .1.5200 0	5820. 3696#0865 306	1.940 +/- .028E 6	36.2 1.868 +/- .1.53E -5	NB							
PEAK IS AT CHANNEL	414.59 WITH HALFWIDTH OF	6.688.	NB PEAK IS SUMMED STARTING	-13.60	CHANNELS HIGHER.						
MO	1032 ZN .05900 0	7234. 617#0050 326	1.772 +/- .039E 6	1	0.043 +/- .1.412E -6	MO					
PEAK IS AT CHANNEL	414.59 WITH HALFWIDTH OF	6.688.	MO PEAK IS SUMMED STARTING	-95.00	CHANNELS HIGHER.						
NI	564. 684#0200 125	5.411 +/- .041E 4	-2.1 -3.973 +/- .1.902E -5	NI							
K	FE PEAK IS AT CHANNEL	103.78 WITH HALFWIDTH OF	4.455.	K PEAK IS SUMMED STARTING	20.40	CHANNELS HIGHER.					
FE PEAK IS AT CHANNEL	103.78 WITH HALFWIDTH OF	4.455.	K PEAK IS SUMMED STARTING	-65.50	CHANNELS HIGHER.						
COUNT RATE CORRECTION FOR LAST ELEMENT = I											

Figure 2: Example LBNL XRF output, showing the information transcribed from each sheet. A: LBNL specimen identifier; B: specimen identifier and descriptive information provided to the laboratory; C: calculated abundance of element; D: estimated uncertainty in abundance determination; E: scale (base-1,000,000) of the measurement; and, F: element of interest.

of the specimen within each Run. The second component of this unique identifier begins each Run with the letters A through Z (with positions B and C excluded, as they were used for standard reference materials), then moves on to numbers 1 through 9, and then the following series of characters: + - \* / ( \$ . ] ≠ → ^ ↑ ; . Thus, “8062 A” is the first specimen measured in Run 8062, “8062 1” is the twenty-fifth specimen measured, and “8062 ;” is the forty-sixth and final specimen measured in the Run.

## 2 Digitization and Transcription of the Records

When necessary, pages were separated along the fan-fold perforations to facilitate feeding through a sheet-feed scanner. Each unique Run printout was scanned at 600 dpi in full color to produce an archival-quality portable document format (PDF) file. Duplicate printouts and photocopies of printouts already present in the archive were not scanned. Each PDF was named using the Run number, an

underscore, and an indication of what elements are reported in the file. Thus file “8031\_mid\_Z.pdf” is a scan of the printouts for Run 8031 reporting on the mid-Z elements.

Following scanning, the calculated elemental abundances, the analytical uncertainties, and the scaling factors were transcribed by hand into a spreadsheet with each column representing a unique specimen, and each row representing these three variables. A series of formulae were used to apply the scaling factors to each reported concentration to scale them all to parts per million (PPM), and the data were transposed to arrange them into a matrix where each column represents an element of interest and the uncertainty of its determination, and each row represents the calculated value for each specimen. Specimen identifiers and any descriptive information for the specimen were appended to this matrix.

As of this writing, there has not been an independent quality-control review of the accuracy of the transcriptions. However, the first author calculated the maximum, minimum, and standard deviation of each column (elemental abundance) to identify any clear outliers caused by errors in scaling factors or in misplaced decimal points. Additionally, a series of bivariate scatterplots were produced to identify any potential errors. The lead author also spot-checked entries in the database against published values, when they could be found, to ensure that they comported with the published values. Although we are reasonably confident in the transcription accuracy of these data, other users of these data should be alert for potential transcription errors.

Data for the mid-Z and high-Z elemental calculations (and associated analytical uncertainties) for each Run were concatenated into a single spreadsheet for each Run. These files are named simply by their Run number and are saved in Microsoft Excel and comma-separated-variable (csv) formats. These individual files were also appended to each other to create a single Excel and csv file containing all of the transcribed data.

### 3 Evaluation of the Data

There are 58 specimens in the transcribed data for which neutron activation data are also present in the LBNL database previously digitized by Boulanger. We extracted the NAA data for these specimens and compared them to the XRF data on an element-by-element basis using a linear model (Table 1). The fit of the NAA data to the XRF data is reasonable for Fe, Ba, and Ce; however, it is not particularly good for Mn, Rb, and Sr. Plots comparing these elemental determinations (Figures 3 – 8) show that the relationship between data from the techniques is linear, but that for some elements there is a difference in determined concentrations exceeding 100 ppm. For most of these elements, the poor fit appears to be driven by a handful of specimens. Indeed, with the exception of Mn, the average differences between the XRF and NAA determinations are within 5% of the NAA-determined concentration of the element (Table 2). We note that Burger and Asaro previously noted the relatively low precision of some elemental determinations. In most published works they report only

Table 1: R<sup>2</sup> values comparing the LBNL XRF and NAA data of those specimens for which common elements were determined using both techniques.

Element	R <sup>2</sup>	n
Fe	0.6476	36
Mn	0.3194	58
Rb	0.3843	36
Sr	0.4941	27
Ba	0.7623	36
Ce	0.7856	14

Table 2: Average (mean) XRF:NAA ratios for specimens in the Burger archives.

Element	XRF:NAA mean
Fe	0.9671
Mn	0.8010
Rb	1.0362
Sr	0.9692
Ba	0.9925
Ce	0.9726

data for Ba, Sr, Zr, and Rb, and even then frequently used ratios of abundances (e.g., Sr/Zr, Rb/Zr) when reporting their data. Thus we the values reported here for Fe, Mn, and Ce, regardless of goodness of fit, should be considered provisional/information-only values.

Some of the XRF runs contained specimens of known Guatemalan obsidian sources (El Chayal, Ixtepeque, and San Martin Jilotepeque), presumably for use as check standards against known values generated by NAA. Data for these three obsidian sources can be useful for adjusting the LBNL XRF data for inter-laboratory comparisons. We note that some runs also contained samples of the Perlman/Asaro Standard Pottery; however, given the differences in chemistry between pottery and obsidian—and the fact that Standard Pottery is no longer available—we do not report further on these specimens. The XRF data for Guatemalan obsidian-source check standards is given in Table 3, alongside data for the same sources produced by NAA at LBNL and at MURR. These data demonstrate that the XRF values for K are clearly below the detection limits of the method used. These data also show that the LBNL XRF data are, though less precise than the NAA data, within the one-sigma boundaries in most elements. Notable exceptions are Mn values (El Chayal and San Martin) and the Zn value for the El Chayal samples. But, the key elements used by LBNL for obsidian sourcing (Ba, Sr, Zr, and Rb) are all within the ranges observed for source and artifact samples analyzed by NAA at LBNL and MURR.

[STILL NEED AN EVALUATION OF Y, ZR, AND NB]

Based on these comparisons, the LBNL XRF data should be broadly com-

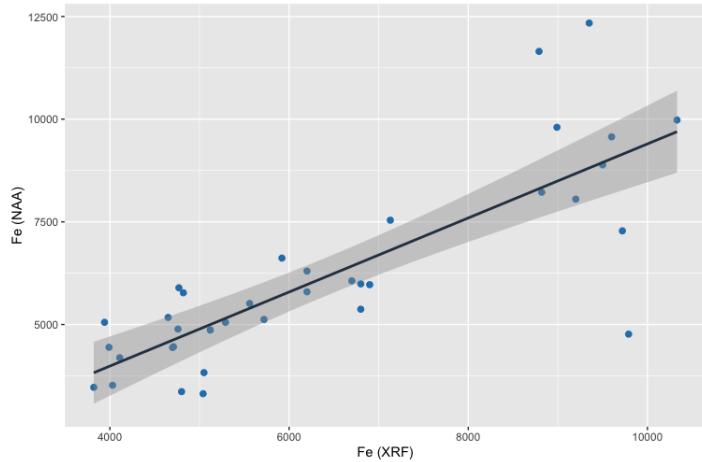


Figure 3: Linear regression plot of Fe concentrations determined by XRF and NAA at LBNL.

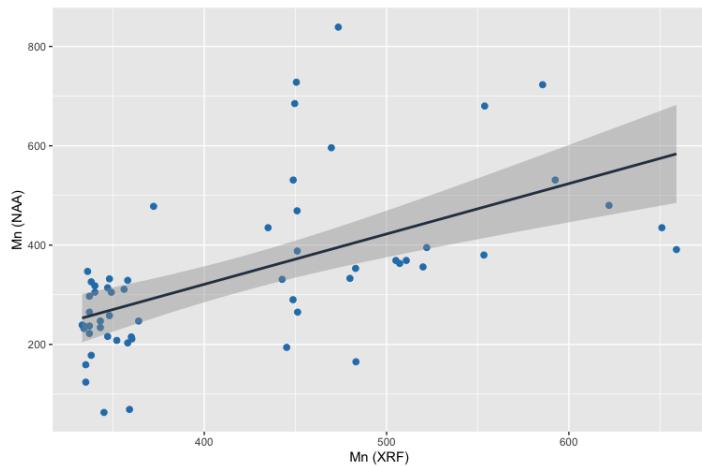


Figure 4: Linear regression plot of Mn concentrations determined by XRF and NAA at LBNL.

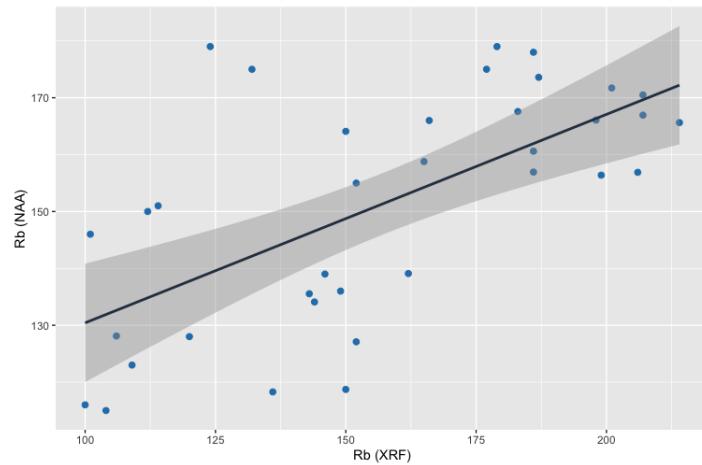


Figure 5: Linear regression plot of Rb concentrations determined by XRF and NAA at LBNL.

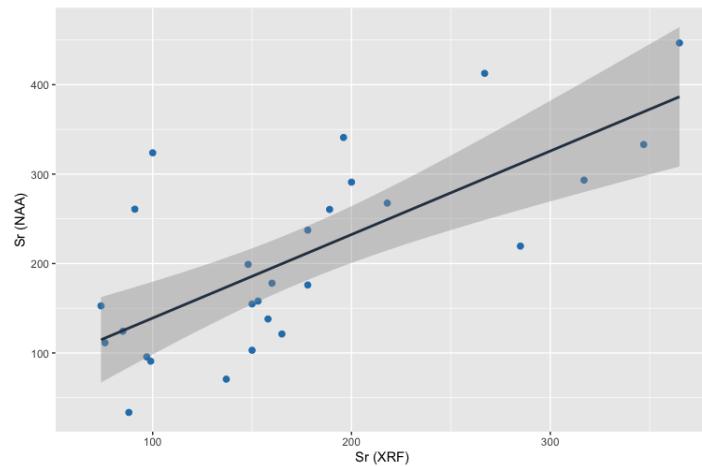


Figure 6: Linear regression plot of Sr concentrations determined by XRF and NAA at LBNL.

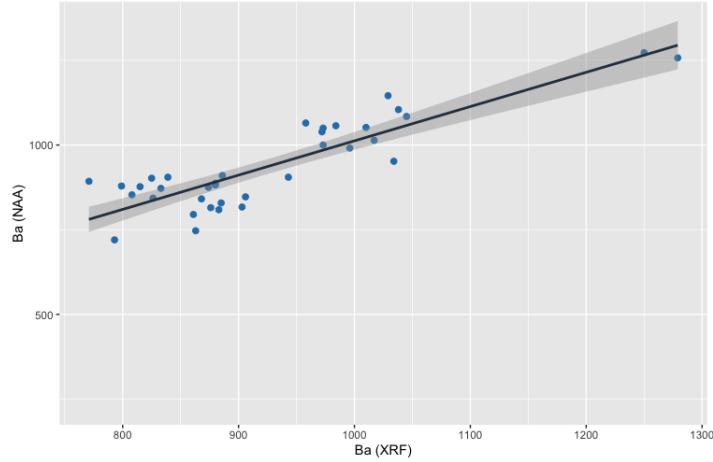


Figure 7: Linear regression plot of Ba concentrations determined by XRF and NAA at LBNL.

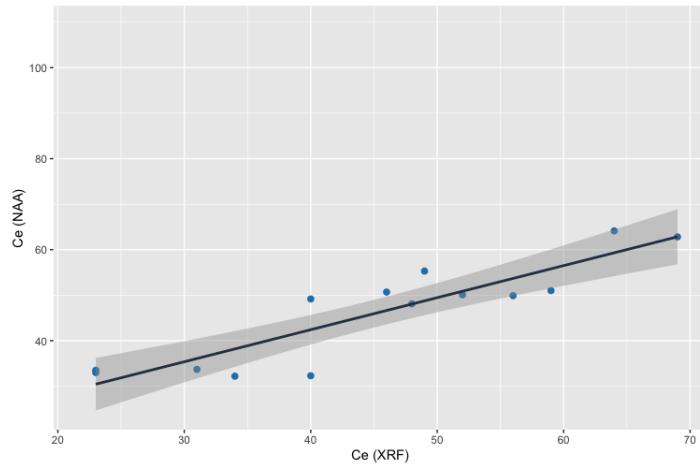


Figure 8: Linear regression plot of Ce concentrations determined by XRF and NAA at LBNL.

Table 3: Comparison of data for LBNL XRF obsidian check-standard samples from El Chayal, Ixtepeque, and San Martin (Guatemala) with LBNL and MURR NAA data for these same sources. Note that values for K appear to be below detection limits for the LBNL XRF. Also note the extremely large standard deviation for La in the El Chayal sample, indicating poor detection of this element via XRF.

		<b>Fe%</b>	<b>Mn</b>	<b>Zn</b>	<b>Rb</b>	<b>Sr</b>	<b>Zr</b>	<b>K%</b>	<b>Ba</b>	<b>La</b>	<b>Ce</b>
<b>El Chayal</b>											
	LBNL XRF	.619 ± .013	475 ± 68	55 ± 18	149 ± 6	153 ± 4	117 ± 3	.027 ± .001	943 ± 11	25 ± 155	48 ± 9
	LBNL NAA	.627 ± .027	649 ± 13	149 ± 8				3.45 ± .26	915 ± 24	25 ± 1	47 ± 1
	MURR NAA	.595 ± .015	649 ± 14	35 ± 12	140 ± 4	168 ± 19	124 ± 19	3.35 ± .22	891 ± 78	24 ± 1	45 ± 1
<b>Ixtepeque</b>											
	LBNL XRF	.847 ± .013	412 ± 57	23 ± 14	102 ± 6	155 ± 4	175 ± 3	.026 ± .001			
	LBNL NAA	.923 ± .019	449 ± 9	103 ± 6				3.61 ± .28	1030 ± 27	24 ± 1	43 ± 1
	MURR NAA	.898 ± .02	458 ± 10	30 ± 3	95 ± 2	173 ± 22	159 ± 9	3.54 ± .22	1017 ± 80	23 ± 1	41 ± 1
<b>San Martin</b>											
	LBNL XRF	.603 ± .01	376 ± 57	36 ± 15	114 ± 7	198 ± 5	121 ± 3	.021 ± .001			
	LBNL NAA	.653 ± .03	522 ± 17	38 ± 3	123 ± 6			3.52 ± .24	1079 ± 56	26 ± 1	48 ± 1
	MURR NAA	.628 ± .016	530 ± 13	32 ± 2	108 ± 2	197 ± 24	117 ± 8	3.38 ± .23	1057 ± 89	25 ± 1	46 ± 1

parable to modern data, at least when those data are produced in reference to NAA data from either the LBNL or MURR laboratories. We recommend that the values for Cr, Mn, Ti, Ca, V, Cu, Pb, Mo, Ni, K, Sn, and La be disregarded from these data, or used with extreme caution, as they neither accurate nor precise, or their correctness cannot be evaluated. Values for these elements are included in the digital file accompanying this report for posterity only. For the remaining elements, users should be aware that values for any individual element within a given specimen may appear anomalous or outside the anticipated range of variation for a given source. Thus, users should either use ratios, as done by Burger and Asaro, or consider a multivariate goodness-of-fit measure such as Mahalanobis distance when constructing compositional groups.

## 4 Disposition of the Paper and Digital Records

<<Where are we storing the paper records, and where are we going to distribute the digital records?>>