CHAPTER 15

Reproducibility and Validity of Portable ED-XRF Instruments: A Comparison of Spectral and Quantitative Results from Belizean and Ethiopian Obsidian

LUCAS R. MARTINDALE JOHNSON,*[a] B. LEE DRAKE,[b] STEVEN A. BRANDT,[c] KEITH PRUFER[b] AND ARLEN F. CHASE[d]

[a] Far Western XRF Lab, Far Western Anthropological Research Group, Inc., Davis, CA, USA; [b] University of New Mexico, Department of Anthropology, Albuquerque, NM, USA; [c] University of Florida, Department of Anthropology, Gainesville, FL, USA; [d] Pomona College, Department of Anthropology, Claremont, CA, USA
*Email: lucas@farwestern.com

15.1 Introduction

Elemental analysis of obsidian by energy-dispersive X-ray fluorescence (ED-XRF) has been an important tool for investigating ancient societies since the landmark study by Cann and Renfrew.[1] More recently, the development of handheld or portable ED-XRF (handheld pXRF hereafter) instruments has vastly increased access to elemental analysis for the study of archaeological materials, with new applications in ceramic sourcing, pigment analysis, and chemostratigraphy.[2] At the same time, however, the increasing application

Advances in Portable X-ray Fluorescence Spectrometry: Instrumentation, Application and Interpretation
Edited by B. Lee Drake and Brandi L. MacDonald

Published by the Royal Society of Chemistry, www.rsc.org

of handheld pXRF has raised important questions concerning scientific reliability, precision, accuracy, and validity.[3–7] Precision refers to low variation as expressed by a low standard deviation and coefficient of variation (CoV) between repeated measurements from the same instrument. Similarly, accuracy is how well the results from one instrument accords with other analyses such as Instrumental Neutron Activation Analysis (INAA). Simple regression statistics can confirm an accurate result by comparing the expected value to the predicted value (*i.e.*, r^2 and slope close to 1). If analyses are both precise and accurate then the measurement is reliable.[3,5] Furthermore, when analyses are reliable, two levels of validity are evaluated. These include the ability to distinguish between sources of obsidian and whether the data are robust enough to answer archaeological inquiry. In other words, does the study possess enough source information to assign an unknown specimen to a geochemical source? These issues are derived from the need to ensure the variation between instrumentation or repeated measurements is less than the variation between two or more sources of obsidian. Because of these issues we examine internal instrument consistency and the consistency between instruments.[7] If instrument variation is too great, then sources are indistinguishable. This issue is largely in the past with the advent of higher resolution detectors and shared calibration sets.

These concerns stem from the central problem with XRF technology: the primary analytic unit is the quantity of gross photon counts for each element (more strictly speaking, the number of photons for a specific energy in the spectrum), which contrasts with conventional reporting of elemental concentrations in weight % or parts per million (ppm). In this chapter, we report both, and use each type of unit to explore variations between similar models of handheld pXRF instruments manufactured by one company. While it is possible for any handheld to be sufficiently consistent to answer archaeological research questions,[5] that cannot be an argument against appropriate empirical calibration of systems and reporting of accurate geochemical results.[7] This is because empirical calibrations are needed not only to report results as relative weights, but also to account for inter-instrument and laboratory variability. ED-XRF is an optical technique and as such has physical variation from instrument components, *e.g.*, X-ray tubes or bulbs, detectors, and their geometric alignment. These variables are idiosyncratic to each instrument and will manifest themselves in analytical results if not appropriately calibrated. Recent studies have demonstrated that employing calibrations can resolve these issues and make XRF a viable and reproducible quantitative technique.[5,7]

In this study two regional obsidian assemblages were selected to compare inter-instrument variability of two different types of Bruker Elemental handheld portable ED-XRF instruments. These two types entailed nine handheld portable ED-XRF instruments used in this study; five instruments manufactured before 2011 contain pin-diode detectors and four instruments manufactured in 2011 and 2013 contain silicon drift detectors (Table 15.1). The configuration of these instruments means that they are capable of

Table 15.1 Inventory of Bruker Elemental handheld ED-XRF instruments used in the study.

Model	Serial #	Manufactured	Owner	Detector type	keV	μA
Tracer IIIV +	K0465	4/2/2007	UC Merced	Pin-diode (mfg 1)	40	13
Tracer IIIV +	K0467	5/2/2007	Emery U.	Pin-diode (mfg 1)	40	14
Tracer IIIV +	K0618	11/24/2008	Continuum Energy	Pin-diode (mfg 1)	40	10
Tracer IIIV +	K0709	9/24/2009	Library of Congress	Pin-diode (mfg 1)	40	27
Tracer IIIV +	K0722	1/11/2010	Simon Fraser U.	Pin-diode (mfg 1)	40	1.7
Tracer IIISD	T3S1241	1/21/2011	U. Mass Boston	Silicon drift	40	15.3
Tracer IIISD	T3S1331	3/7/2011	Continuum Energy	Silicon drift	40	9.8
Tracer IIISD	T3S2429	1/16/2013	Bruker demo instrument	Silicon drift	40	30
Tracer IIISD	T3S2430	1/16/2013	Bruker demo instrument	Silicon drift	40	30

evaluating the effects of differing components while keeping software and spectral analysis methods consistent. Through the analysis of 18 artifacts and 13 quarry samples we show that inter-instrument variation is statistically insignificant in this study. We demonstrate this by accounting for four key aspects: (1) critical operating instrument parameters (*e.g.*, kV, current, use of filters, and analysis duration), (2) the effects of artifact thickness and variation in surface geometry, (3) the influence of portable ED-XRF manufacturing parameters (*e.g.*, tube and detector types), and (4) how spectral data transformation methods (*e.g.*, calibration) quantify the raw data. Having accounted for these variables we conclude that within our sample inter-instrument variability is insignificant. However, it is advisable that should multiple instruments be used on shared archaeological materials subject to geochemical sourcing, that these variables be evaluated further within a localized research area (*e.g.*, Mesoamerica, the Southwestern US, the Great Basin, or the Mediterranean) to ensure the variation between instruments is less than the variation between sources of obsidian.

15.2 Brief Overview of XRF Theory

X-ray fluorescence functions similarly to color fluorescence. As a photon is emitted from a source, it interacts with outer-orbital electrons in molecules (color fluorescence) or inner-orbital electrons in atoms (X-ray fluorescence) in a material, and emits a photon characteristic of the material following electron excitation. In color fluorescence, this is perceived by most humans as color from red (1.91 eV) to violet (3.1 eV). X-ray fluorescence is the same process, but the range is typically larger including sodium (1.1 keV) to barium (32.2 keV) for K-alpha fluorescence and ranging up to uranium (13.61 keV) for L-alpha fluorescence. While photons are typically reported as

wavelengths of light (650 nm for red and 400 nm for violet), they can be converted to energies using the following expression:

$$E = hc/\lambda \tag{15.1}$$

where E is energy in joules, h is Plank's constant, c is the speed of light, and λ is the wavelength. To convert joules into electron volts, the energy in joules must be multiplied by 1.602×10^{-19}. In this sense, the spectra that result from ED-XRF analysis are a continuation of the color spectrum. As each photon is emitted as one K or L shell fluorescence, the spectral peaks resulting from such fluorescence can be thought of as atom counts of sorts – though the quantity of photons is determined by both the voltage of the tube and the depth of beam penetration. The depth of penetration can be calculated using Beer's Law:

$$I/I_0 = e^{[-(\mu/\rho)x]} \tag{15.2}$$

where I is the quantity of photons returning from the sample, I_0 is the quantity of photons entering the sample, μ/ρ represents the mass attenuation coefficient of a given element for a particular matrix, and x represents the density of the object. Assuming a limit of 1% returning photons from a silicate matrix, the depths of analysis of key elements are illustrated in Figure 15.1.

Elements that fluoresce with low energy, such as silica at 1.7 keV, will only return photons from 20 μm deep into a SiO_2 matrix, while elements that fluoresce at higher energies, such as zirconium at 15.77 keV, will return photons from as deep as 3.4 mm. Consequently, elements with higher energies of fluorescence will be more easily identified in smaller concentrations. As energy in X-ray tubes is not evenly distributed,[8] there is further discrimination against the fluorescence of elements on the extreme ends of

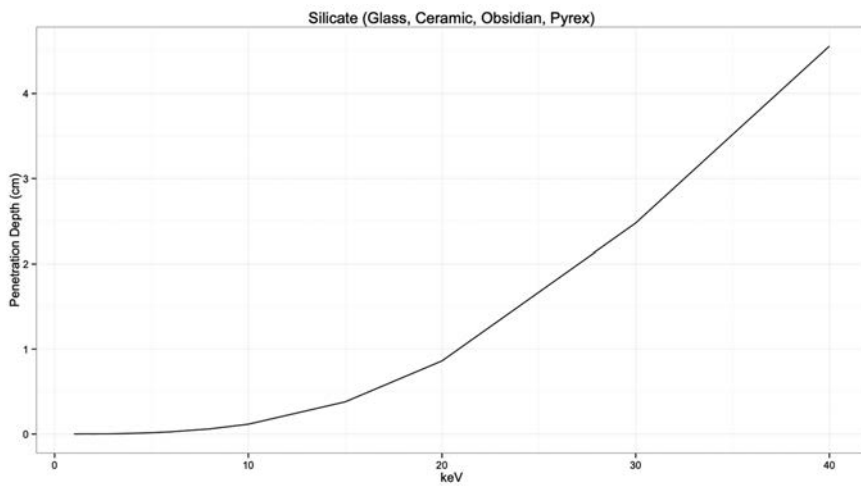

Figure 15.1 Depth of measurement in a pure silicate by keV.

an energy spectrum (near 1 and 40 keV). Furthermore, the fluorescence of one element can influence that of another (*i.e.*, secondary fluorescence, see Figure 4.7, Chapter 4). A high concentration of zinc, with a K-alpha fluorescence of 8.78 keV, sits on the absorption edge of copper, with a K-alpha fluorescence of 8.01 keV. A high concentration of zinc will distort the quantity of copper present in the spectra. In addition to these factors, fluorescence efficiencies vary from element to element.

15.2.1 Calibration Overview and Instrument Components

XRF is a technique that obtains data on the energy of mass-less photons which are then converted to units of relative mass of a sample *via* mathematical models and assumptions. This is because units such as ppm are more easily understood concentration values and used to show the relative difference between concentrations of elements. Mathematical transformations can be relatively simple and straightforward[9] or involve more complex processing using physical parameters.[10] Nonetheless, the analytic unit, photon count rates, and the synthetic unit (weight % or ppm) are distinct by source group. As defined by Ramenofsky and Steffan,[11] a synthetic unit is the combination of analytic units derived from instrumental data combined with outside assumptions. Any quantification of XRF spectral data beyond raw photon counts is a synthetic unit, placing profound emphasis on the assumptions used to acquire that synthetic unit, and the need for standardization for replicability.

Based on their energetic properties, the spectra in X-ray fluorescence cannot be considered a pure atomic measurement of elements in a sample. The combination of variability of depth measurement, uneven voltage distribution, and overlapping elemental peaks precludes any simple analysis of X-ray fluorescence data. Therefore, empirical calibrations were developed and are used.[9] Empirical calibrations typically employ a variant of the Lucas-Tooth empirical calibration equation:

$$\hat{C}_i = r_0 + I_i\left(r_i + \sum\left(r_{in}I_n\right) + e_{\mathrm{q}} + e_{\mathrm{c}} + e_{\mathrm{h}} + e_{\mathrm{i}}\right) \tag{15.3}$$

where $\hat{C}_i$ represents the concentration of the element or elements. r_0 is the intercept/empirical constant for element i, r_i is the slope/empirical coefficient for intensity of element i, r_n is the slope/empirical constant for effect of element n on element i, I_i is the net intensity of element I, and I_n is the net intensity of element n. In addition, four uncertainty terms are included in this equation, e_{q}, which is the uncertainty originating in the quantification procedure, e_{c}, uncertainty originating from the spectra acquisition, e_{h}, uncertainty originating from heterogeneity in the sample, and e_{i}, outside uncertainty from other sources not anticipated. A full discussion of these uncertainty terms can be found in Chapter 16. While e_{q} can be approximated by using root mean square error (RMSE), mean absolute error (MAE), or $1 - r^2$ from the quantification model used, it is dependent on both the

quality of the samples and the application of inter-elemental corrections. By contrast, e_c can be reduced by simply extending the measurement time. For example, extending the analysis time for pin-diode detectors reduces spectra counting uncertainty (see Figures 16.5–16.7 in Chapter 16). For ED-XRF e_c is simply the noise in count rates near the element in question and can be estimated using the index of dispersion/fano factor (Chapter 16). Heterogeneity in the sample, e_h, can be addressed by multiple observations on the sample, though, for obsidian samples used in this study it should be minimal. Finally, e_i is the result of various causes, varying from sample orientation to the beam, instrumental quality, and other variables which cannot be quantified in advance. However, the design of a study should reduce exposure to e_i, as it will regress to 0 with sufficient replication of measurements. Our study design employed multiple users with multiple devices on the same standards to attempt to reduce e_i.

The above equation assumes knowledge of the variation of other elements (r_n and I_n in this equation); this is because some elements influence the fluorescence of others. In obsidian sourcing, the k-alpha peak for zirconium overlaps with the k-beta peak for strontium. If a quantification procedure does not factor this overlap into its algorithms, then the reported quantity of zirconium will be influenced by the quantity of strontium. A key strength of the Lucas-Tooth and Price[9] algorithm is that it corrects for these effects by producing linear models for the quantification of each element. Empirical calibrations following this algorithm will be accurate within the confines of the regression line (*i.e.*, minimum and maximum point). Thus, the accuracy of the algorithms is contingent upon the elemental variation captured by the empirical reference set and its appropriateness to the material being studied.

For an empirical calibration to work, four key assumptions must be met. First, the analyzed material must be homogenous or, at the very least, sufficiently well mixed to be practically homogenous relative to the beam (e_h). Second, the reference set must consist of the same type of material (*e.g.*, rhyolitic glass in the same form as the analyte) for the calibration to be generalizable. For example, the obsidian calibration cited herein is of unmodified, not powdered, rock in the same form as archaeological materials. Third, the reference set should encompass elemental variation in the samples for all elements quantified. Fourth, the reference set must encapsulate the expected minimum and maximum abundance of every element. In addition to these four criteria, the analytical data must be taken with the same parameters on both the standards and the unknowns. These include the same energy, current, filter, and atmosphere (dry air, vacuum, He, *etc.*) within one instrument, but may vary between instruments as is shown below.

A calibration is necessary to translate luminescence data (*e.g.*, photon quantities/intensities) to quantitative units. A calibration has the secondary effect of reducing instrument-to-instrument variation by adjusting coefficients to match the composition of empirical standards. Every instrument, even those which contain the same components, have slight variations in photon excitation and detection due to manufacturing variation in bulbs

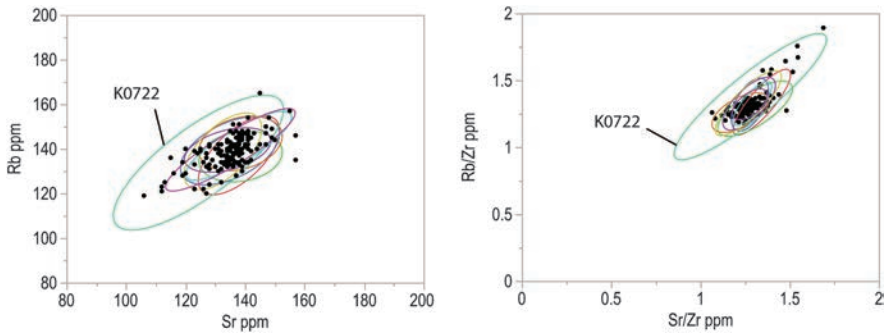

Figure 15.2 Biplot of El Chayal sourced obsidian using ppm grouped by pXRF instrument. Left: Rb by Sr. Right: Rb/Zr by Sr/Zr with 95% confidence ellipses.

and very slight variance in internal geometries[12] (Figure 15.2). These differences manifest themselves in the elemental spectra.

The elements identified by XRF are determined by the energy range of photons emitted from the sample and the reolution of the detector. Systems with a pin-diode detector and a 200 eV resolution at Full Width Height Maximum (FWHM) can see K-alpha peaks for elements ranging from magnesium to barium and L-alpha peaks from barium to uranium when a range of 40 keV photons is detected in the sample. Systems with a silicon drift detector can include both sodium and neon, as those detectors have a resolution of 141 eV FWHM. There is, however, no ED-XRF system that can detect oxygen directly *via* a K-alpha fluorescence peak, the chief reason being that the analytical depth for oxygen would only be 10 nanometers. As such, only elements in the periodic chart from sodium to uranium are quantifiable using modern portable ED-XRF systems. This is important, as the results of XRF analysis are often erroneously reported in oxide form. For obsidian, all elements detected *via* XRF are bound to oxygen, this is not true for all sedimentary rocks. Many elements have multiple oxidization states. For example, FeO Fe_3O_4, and Fe_2O_3 can be present in a sample. Or, hydroxide could be a possibility, $Fe(OH)_2$. And it is not the case that one state is predominant enough to justify the assumption that it exists exclusively in a material. FeO forms about 3.8% of oxides in continental composition, while Fe_2O_3 forms 2.5%. An XRF system that interprets a Fe K-alpha peak only as Fe_2O_3 regardless of the sample is almost certainly inaccurate. Similar problems exist for other geologically important elements, such as CaO and $CaCO_3$. ED-XRF cannot determine the molecular structure of a substance, it can only identify elements and their quantities.

In some cases, an often standardless calibration approach, known as Fundamental Parameters, can be employed.[13] This approach typically relies upon using physical variables to calculate weight % units from the spectra. In the case of modern metal alloys or other industrial materials this approach can be helpful because of the predictable chemical composition. For archaeological

materials that can be much more chemically diverse and variable, this approach requires greater examination. As artifacts like stone tools and ceramics are primarily composed of oxygen (SiO_2), the possibility of different oxidization states for elements precludes accurate estimation using the spectra as the only independent variable. For obsidian, in particular, the source assignment of an archaeological material can depend upon a difference of a few parts per million of a trace element like strontium or rubidium. This low-level sensitivity is why empirical calibrations are necessary as it can be evaluated externally to determine if the method is appropriate.

While empirical calibrations can provide highly accurate results, a calibration itself adds no new information that was not already present in the X-ray fluorescence spectra – the calibration only translates the variation present in the X-ray spectra into chemical weight-percent. As such, semi-quantitative analysis should converge on the same conclusions. One frequently used route is Bayesian Deconvolution, in which known inter-elemental effects are used in multiple re-simulations of the data to produce a net photon count for each element. Other, simpler methods can be used, such as normalizing gross photon counts to either the Compton curve or valid-count rate.

Filtration of the beam is a critical component of heightening sensitivity to trace elements useful in geochemical sourcing. A filter, among other potential functions, reduces the Bremsstrahlung radiation for a given portion of the spectrum, depending on its density. The filter used in this study (informally known as the ''green filter'') was a 150 μm Cu/25 μm Ti/300 μm Al filter designed to eliminate Bremsstrahlung radiation from 15 keV and below. This allows for a maximum signal–noise ratio for mid-Z elements like Th L-alpha, Rb K-alpha, Sr K-alpha, Y K-alpha, Zr K-alpha, and Nb K-alpha emission lines that are critical to sourcing obsidian.[1,3] A secondary impact of the green filter is that it reduces the variation in fluorescence from tube to tube, making it possible to transfer one instruments' calibration to another, though this procedure is not recommended as it can introduce systematic error in the range of ∼5 ppm to quantitative results, enough to influence geochemical source determinations in some cases.

15.2.2 Analytical Outputs of Bruker Handheld pXRF Instruments

Raw data is recorded as a spectral file with metadata which captures the resolution of the spectrum measured in the full width height maximum of the Mn K-alpha line (FWHM), time of acquisition, count rate, and several sensor variables capturing ambient detector temperature, air pressure, and other potentially influencing factors. While Bruker software uses the proprietary .pdz file format, data should be archived as open source .csv files. Quantification of the data uses normalization between 19.4 and 22.1 keV, capturing the Rayleigh scattering of the Rh tube, high-energy Compton scattering, and fluorescence of the Pd detector collimator as an internal control. From here, inter-elemental corrections to the elements follow the recommendations of

Glasscock and Ferguson[14] and Speakman.[15] These spectral files are then converted to ppm using the above calibration methods.

15.3 Materials and Methods

15.3.1 Handheld pXRF Instrument Configuration and Analysis of Variation

A sample of 31 obsidian specimens chosen from existing and convenient research collections (*i.e.*, 18 artifacts from Caracol, Belize and 13 quarry samples from Baantu, Ethiopia) were analyzed on nine portable ED-XRF instruments with different rhodium tubes with either a pin-diode or silicon drift detector. As such, each instrument had differences in microamps (μA) and was unchanged throughout the study. As the artifacts' surfaces were not uniform, some variation was inevitably added due to differences in spacing between the artifact and the detectors and the location scanned on each specimen. The project had four fundamental constraints. First, each machine was manufactured by the same company: Bruker Elemental. Second, each machine was operating within parameters that best characterized obsidian (*e.g.*, filter containing 150 μm copper, 25 μm titanium, 300 μm Al, 40 kV, and 40 μA, though note that some older tubes were not capable of producing this current and had to be run closer to 10 μA). These were the same settings used to characterize the calibration of the University of Missouri Research Reactor (MURR) calibration set, and any change would affect the predicted ppm due to changes in counting statistics. Third, each object was assayed for 60 seconds without the aid of a vacuum. Lastly, a standard calibration method was used to calculate parts per million (ppm) values from net photon counts normalized to the Compton peak of rhodium using the MURR 2 calibration.[14] The instruments, pedigrees, and their relevant specifications as of the 2013 study are detailed in Table 15.1.

All data was calibrated using the MURR 2 reference set[15] in Bruker S1CalProcess software (2.3.8.1). While normalizing to the elastic scatter of rhodium (18.5–19.5 keV) is generally recommended, this study included the analyses of niobium, which has a kβ peak located at 18.6 keV. To avoid problems of covariance, data were normalized to photons from 19.5 to 22 keV. Additionally, all data were analyzed using Bayesian Deconvolution in Bruker Artax software (version 7.4) using 10 stripping cycles between 1 and 40 keV to correct for Bremsstrahlung radiation (*e.g.*, backscatter) and inter-elemental overlaps in the spectra to produce net photon counts (*e.g.*, spectral intensity attributable to atomic abundance of the element in the analyte) for each element present in the sample. This was done to independently assess the appropriateness of the calibration reference set used.

Inter-machine variation or difference was assessed using the coefficient of variation statistic for each artifact using both parts per million data and net photon counts [$\mathrm{CoV} = (\sigma/\mu) \times 100$ or standard deviation divided by a given mean multiplied by 100 to express variation as a percent]. Artifacts were then arranged by source and the CoV was averaged to provide a general statement

of inter-instrument variation by element. Finally, a two-way analysis of variation (ANOVA) test was run to determine the proportion of variation between instruments and artifacts. The objective of this study was to determine how much an empirical calibration corrects for use-cases in obsidian studies.

15.3.2 Obsidian Specimens

For the 2013 study we used 31 obsidian specimens. Eighteen specimens were from the Classic Maya site of Caracol, Belize. Thirteen were from the obsidian quarry of Baantu in southwestern Ethiopia.

15.3.2.1 Maya Samples

Caracol, Belize is located on the Vacca Plateau in west central Belize and is a prominent Maya site spread over 200 sq km of terrain.[16–21] During the Classic period (AD 250–900) Caracol was a major consumer of obsidian from quarries located in the highlands of current day Guatemala.[22] Prior to this study it was determined that 16 of our 18 artifacts align with geochemistry from the El Chayal source and two share element concentrations with the Ixtepeque source.[22]

The Caracol specimens were selected at random and included thinner ($\leq$4 mm) artifacts, like prismatic blades and thicker ($\geq$4 mm) artifacts, like macro core-shaping debitage, platform preparation debitage, and both complete and fragmentary polyhedral blade-cores. These differing types generally represent the size and morphological diversity in a given archaeological assemblage from Classic Period Maya sites. One important feature of these artifacts is that blades, platform preparation flakes, and core-shaping flakes have thicker medial portions and pronounced bulbs of percussion/pressure that provide thicker but narrower areas to place over a portable XRF detection window. Blade-cores have wider flat surfaces that can cover the entire detection window. The varied thickness and degree of coverage can influence the beam intensity received by the detector (*i.e.*, lower count rates); however, the elemental ratios and peak percentages do stay consistent. In other words, thicker specimens return higher net photon counts in contrast to thinner specimens (2–3 mm), but the ratios and relative percentages of elements remain consistent.[23] However, when thinner specimens are analyzed and photon counts are converted to ppm, some elemental variation will exist (*e.g.*, Zr). Notwithstanding, these slight elemental intensity variations do not significantly affect the ability to assign artifacts to obsidian source groups in the Maya area because the sources are highly distinct.[5,22]

15.3.2.2 Ethiopian Specimens

The 18 Ethiopian samples are from the Baantu obsidian quarry, an extensive, eroded area in the Ethiopian Main Rift $\sim$ 20 km SE of Mochena Borago Rockshelter in SW Ethiopia.[24,25] The samples represent a smaller sample of robust spalls and chunks. These included samples that were purposefully struck off

via direct freehand percussion from a larger random collection of natural, cortex covered nodules obtained from the surface obsidian flows at Baantu. They were initially used as geological samples for a pXRF study to determine the source(s) of obsidian artifacts from Mochena Borago Rockshelter.[26]

15.4 Results and Use of Instrument Outputs

15.4.1 Archaeological Obsidian recovered from Caracol, Belize (El Chayal and Ixtepeque Source-Groups)

Our analysis of obsidian artifacts from the site of Caracol showed that the variation between instruments is not significant enough to confuse two widely utilized obsidian source-groups in the Mesoamerican area. The two Mesoamerican source groups exhibit low geochemical variability, ranging from 119–165 ppm Rb, 106–157 ppm Sr, 9–28 ppm Y, 74–120 ppm Zr, and 3–14 ppm Nb when compared to the Ethiopian samples. Additionally, the Guatemalan obsidian source-groups overlap with the Baantu Quarry site samples from Ethiopia (see Table 15.2). For obsidian artifacts that were sourced to the El Chayal source-group, yttrium and niobium showed the most variation (CoV of 14 and 16% respectively), while other elements showed less of a range in their respective coefficient of variation inclusive of all source specific specimens analyzed by all instruments. The same is true for artifacts sourced to the Ixtepeque obsidian source. The variation in yttrium and niobium is likely due to the lower abundance and associated lower photon counts for these elements. The other elements, however, have lower coefficients because of the higher ppm values and greater photon counts. Important here is that the CoV values are low (<17%), artifact ppm ranges are minimal, and photon counts across instruments are consistent, respective of instrument kV and μA current settings.

Our study included many specimens that are greater than 3 mm and, in most cases, greater than 4 mm thick, *i.e.*, they are ''infinitely thick.''[27] Small and/or thin artifacts can influence ppm and photon values,[22,27,28] but size and geometry do not seem to have affected the ppm range or the ability of each instrument to precisely produce ppm values in our samples, even when photon counts are low. The major reason for this is that each instrument was operating within the same parameters (kV and μA current) as the calibration reference set, so the spread in ppm values is likely a normal variation in source-group geochemistry as well as standard variation with regard to instrument variations. Although both obsidian test cases exhibit some variation in ppm values, this is not significant enough to alter source-group assignments (Figures 15.2 and 15.3). In the case of El Chayal and Ixtepeque obsidian, the introduction of a third element helps to further separate the sources and mitigate any possible overlap due to artifact surface morphology or geometry and inter-instrument variation (Figure 15.4 right). Hughes[23] advocates that when specimen size is an important factor of consideration, analysts may wish to focus their attention away from ppm values by utilizing

Table 15.2 Summary CoV statistics expressed as a percent (CoV x 100) of variation for Guatemalan obsidian samples analyzed on all instruments from the site of Caracol, Belize.

Cat. number		Parts per million (ppm)					Ratio of net photon count[a]			
Artifact type		Rb	Sr	Y	Zr	Nb	Rb%	Sr%	Zr%	Source
C12A/49-1a macro-flake	μ	138.6	136.6	18.9	107.0	9.9	0.292	0.343	0.366	El Chayal
	σ	5.0	3.2	2.0	4.4	1.5	0.009	0.004	0.009	
	CoV	3.6	2.4	10.7	4.2	14.7	3.2	1.2	2.4	
C12A/49-1b blade	μ	142.7	139.2	19.2	106.7	8.8	0.295	0.344	0.361	El Chayal
	σ	9.9	3.9	2.4	6.5	2.4	0.012	0.006	0.016	
	CoV	6.9	2.8	12.7	6.1	27.8	4.2	1.9	4.4	
C12A/49-1c rejuvenation flake	μ	139.1	142.1	18.8	106.9	10.0	0.290	0.346	0.364	El Chayal
	σ	5.7	13.1	1.9	3.9	0.9	0.010	0.014	0.016	
	CoV	4.1	9.2	10.2	3.6	8.7	3.5	3.9	4.4	
C12A/49-1d Platform flake	μ	137.3	133.3	19.1	106.4	9.9	0.291	0.341	0.367	El Chayal
	σ	6.2	3.6	2.8	7.0	1.1	0.007	0.009	0.007	
	CoV	4.5	2.7	14.4	6.5	10.7	2.2	2.6	1.9	
C12A/49-1e blade	μ	141.3	136.2	19.1	108.9	10.4	0.294	0.337	0.369	El Chayal
	σ	5.3	9.1	2.6	3.5	1.0	0.008	0.012	0.016	
	CoV	3.8	6.7	13.5	3.2	9.7	2.9	3.7	4.3	
C12A/49-1f platform flake	μ	139.3	134.4	20.6	109.4	10.0	0.290	0.338	0.372	El Chayal
	σ	8.2	10.1	1.5	6.4	1.5	0.006	0.008	0.013	
	CoV	5.9	7.5	7.3	5.9	15.0	2.2	2.4	3.5	
C12A/49-1g blade	μ	142.6	135.9	18.6	107.4	9.1	0.296	0.339	0.365	El Chayal
	σ	7.6	9.6	2.8	5.8	2.1	0.007	0.009	0.013	
	CoV	5.3	7.0	15.3	5.4	22.9	2.4	2.8	3.5	
C39B/7e core fragment	$\bar{x}$	135.3	133.9	18.9	105.8	9.3	0.290	0.342	0.368	El Chayal
	μ	4.2	7.2	4.3	4.2	1.0	0.008	0.008	0.012	
	CoV	3.1	5.4	22.5	3.9	10.7	2.9	2.4	3.2	
C39B/7f macro-blade	μ	145.8	139.9	20.3	110.0	9.5	0.294	0.341	0.365	El Chayal
	σ	7.6	6.5	3.7	3.9	1.8	0.009	0.005	0.008	
	CoV	5.2	4.6	18.0	3.5	18.7	3.0	1.4	2.3	

Table 15.2 (*Continued*)

Cat. number		Parts per million (ppm)					Ratio of net photon count[a]			
Artifact type		Rb	Sr	Y	Zr	Nb	Rb%	Sr%	Zr%	Source
C39B/7g core fragment	μ	137.7	132.8	18.1	104.1	10.2	0.297	0.338	0.365	El Chayal
	σ	7.4	6.4	2.0	4.9	2.0	0.009	0.005	0.010	
	CoV	5.4	4.8	11.2	4.7	20.0	3.0	1.5	2.8	
C39B/7h core fragment	μ	137.9	133.4	17.7	105.9	9.3	0.294	0.340	0.367	El Chayal
	σ	4.8	6.8	2.8	5.7	2.1	0.006	0.005	0.009	
	CoV	3.4	5.1	15.8	5.4	22.7	2.2	1.5	2.4	
C39B/7i rejuv. flake	μ	139.2	137.0	18.3	107.9	10.3	0.291	0.342	0.367	El Chayal
	σ	8.5	6.3	2.9	6.1	1.2	0.007	0.008	0.007	
	CoV	6.1	4.6	15.7	5.7	11.9	2.4	2.4	1.9	
C3C/12-2 core fragment	μ	133.3	132.4	18.1	105.3	10.1	0.288	0.342	0.370	El Chayal
	σ	7.6	8.5	2.7	5.4	1.7	0.006	0.007	0.010	
	CoV	5.7	6.4	15.0	5.1	16.7	2.1	2.1	2.6	
C70B/42-1a core fragment	μ	133.7	131.6	18.2	104.6	9.4	0.292	0.339	0.369	El Chayal
	σ	5.8	4.3	1.5	4.9	1.3	0.009	0.010	0.016	
	CoV	4.3	3.3	8.1	4.6	14.1	3.2	3.1	4.4	
C70B/42-1b core fragment	μ	137.0	133.6	17.8	104.6	10.0	0.292	0.343	0.365	El Chayal
	σ	8.5	10.8	3.8	4.7	0.5	0.012	0.006	0.014	
	CoV	6.2	8.1	21.6	4.5	5.0	4.0	1.9	3.9	
C70B/42-1g core fragment	μ	139.2	126.1	20.1	87.4	9.4	0.321	0.349	0.329	El Chayal
	σ	7.1	8.0	2.3	7.1	1.1	0.009	0.008	0.011	
	CoV	5.1	6.3	11.2	8.1	12.0	2.7	2.2	3.2	
C86C/15-2 core fragment	μ	96.4	139.1	17.6	156.3	9.4	0.189	0.322	0.488	Ixtepeque
	σ	3.3	5.9	1.8	13.0	1.1	0.004	0.008	0.010	
	CoV	3.4	4.2	10.3	8.3	12.0	1.9	2.5	2.1	
C39B/7c blade	μ	103.7	156.2	18.3	171.8	11.3	0.183	0.328	0.489	Ixtepeque
	σ	9.6	8.8	3.4	9.8	1.7	0.011	0.005	0.013	
	CoV	9.3	5.7	18.5	5.7	15.3	5.8	1.4	2.7	

[a]See Hughes 2010 for methodology.[23]

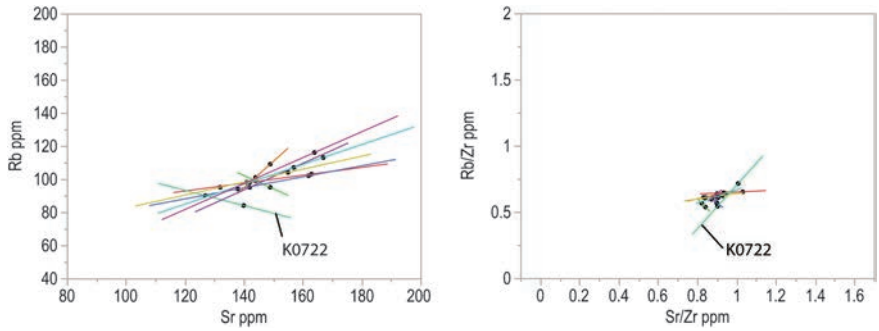

Figure 15.3 Biplot of Ixtepeque sourced obsidian using ppm grouped by pXRF instrument. Left: Rb by Sr. Right: Rb/Zr by Sr/Zr with 95% confidence ellipses. Ellipses display as a line due to small sample size of fitting the Ixtepeque source-group. Note that sample size is in relative proportion to El Chayal in archaeological assemblages from Maya sites in this area of Central America (*e.g.*, about 1 in 10).

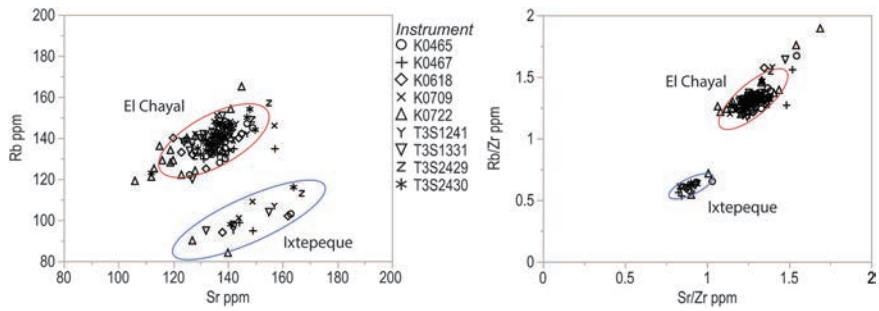

Figure 15.4 Biplot of both El Chayal and Ixtepeque obsidian samples grouped by source assignment with a 95% confidence ellipse included to show the spread of variation.

semi-quantitative analysis using specimen photon counts or peak intensity. This method of elemental photon ratios (Table 15.2) shows even less variation regarding specimen by source even though most artifacts were greater than 4 mm thick. The focus on a photon count relationship between three elements also diminishes the spread in variation seen in the ppm data (see Figure 15.6).

15.4.2 Results of Ethiopian Quarry Sample Obsidian (Baantu Quarry)

The Ethiopian quarry samples show high geochemical variability, ranging from 145–250 ppm Rb, 0–8 ppm Sr, 138–220 ppm Y, 1307–2292 ppm Zr, and 234–393 ppm Nb. Strontium is effectively nonexistent in the sample (*e.g.*, 0–8 ppm). Some of this variation could be due to the presence of sub-sources (*i.e.*, punctuated obsidian formation events in one location), although the

Table 15.3 Summary of CoV statistics expressed as a percent (CoV×100) of variation for Baantu Quarry obsidian in southwestern Ethiopia analyzed on all instruments.

Quarry sample ID		Parts per million (ppm)					Ratio of net photon count[a]		
		Rb	Sr	Y	Zr	Nb	Rb%	Sr%	Zr%
Q109 2	μ	218.7	3.1	181.2	1859.0	342.7	0.073	0.006	0.922
	σ	13.2	1.4	17.7	177.2	29.0	0.002	0.001	0.003
	CoV	6.0	43.9	9.7	9.5	8.5	3.3	17.3	0.3
Q112	μ	219.6	2.9	179.3	1869.6	338.3	0.073	0.006	0.922
	σ	10.6	1.6	12.6	129.1	20.3	0.004	0.001	0.004
	CoV	4.8	55.9	7.0	6.9	6.0	5.0	12.3	0.5
Q113	μ	223.0	3.4	177.2	1886.2	345.4	0.073	0.006	0.921
	σ	13.2	1.6	16.0	156.2	23.6	0.004	0.001	0.004
	CoV	5.9	46.2	9.0	8.3	6.8	5.8	19.4	0.5
Q114	μ	219.4	3.3	180.7	1867.8	343.4	0.073	0.006	0.922
	σ	9.1	1.6	11.7	132.9	17.5	0.003	0.001	0.004
	CoV	4.1	47.4	6.5	7.1	5.1	4.3	17.4	0.4
Q115	μ	160.2	3.1	145.6	1431.9	246.4	0.069	0.006	0.925
	σ	7.3	2.4	11.0	94.2	10.0	0.004	0.001	0.004
	CoV	4.5	77.8	7.5	6.6	4.1	5.2	21.1	0.5
Q2	μ	221.3	4.0	178.8	1858.6	342.3	0.074	0.006	0.920
	σ	12.0	1.9	15.4	154.7	23.7	0.002	0.001	0.003
	CoV	5.4	46.8	8.6	8.3	6.9	2.8	12.3	0.3
Q3 2	μ	217.8	3.0	180.1	1853.0	339.1	0.073	0.005	0.922
	σ	14.8	1.5	15.9	154.2	20.8	0.004	0.001	0.005
	CoV	6.8	50.0	8.8	8.3	6.1	5.8	15.2	0.5
Q3 3	μ	213.9	3.7	174.8	1807.3	335.2	0.073	0.006	0.921
	σ	15.1	1.5	15.2	128.7	17.3	0.003	0.001	0.003
	CoV	7.1	40.9	8.7	7.1	5.1	4.5	11.7	0.4
Q38 1	μ	222.6	2.1	181.9	1855.9	343.3	0.074	0.005	0.921
	σ	14.7	1.1	10.1	126.8	18.4	0.005	0.001	0.005
	CoV	6.6	49.9	5.5	6.8	5.4	6.4	12.0	0.5
Q39	μ	226.6	2.9	186.1	1892.6	350.9	0.074	0.005	0.920
	σ	7.4	1.7	14.1	121.2	13.7	0.003	0.001	0.003
	CoV	3.3	58.6	7.6	6.4	3.9	4.2	21.5	0.4
Q63 4	μ	224.9	2.2	183.7	1881.4	343.8	0.074	0.005	0.921
	σ	10.3	1.1	9.0	128.8	20.7	0.003	0.001	0.003
	CoV	4.6	49.2	4.9	6.8	6.0	4.0	14.2	0.3
Q65 6	μ	226.0	1.9	187.8	1923.3	352.8	0.073	0.005	0.922
	σ	14.1	1.1	13.8	146.7	18.9	0.004	0.001	0.005
	CoV	6.2	55.8	7.3	7.6	5.4	5.0	21.3	0.5
Q95 1	μ	220.6	2.3	180.4	1829.4	339.0	0.074	0.005	0.920
	σ	10.9	2.0	13.1	159.1	15.0	0.003	0.001	0.004
	CoV	4.9	88.1	7.3	8.7	4.4	4.1	16.2	0.4

[a]See Hughes 2010 for methodology.[23]

relationship between Rb, Sr, and Zr photon count might suggest otherwise due to the relationship in photon count percentages (Figure 15.6). The coefficient of variation for strontium is high (*e.g.*, 88.1% variation, Table 15.3) and is due to overall low strontium ppm values. Other elements with higher ppm values have much lower CoV values (<10% variation). Artifact surface

variation is not a consideration in this sample set because all were thick, blocky geological samples with flat surfaces, some of which retained cortical surfaces; no cortical surfaces were scanned.

Again, all instruments performed similarly, with instrument K0722 preforming slightly differently likely due to the lower energy μA current used during analysis (see Table 15.1). Regardless, each instrument plots any two elements (*e.g.*, Rb and Y) similarly (Figure 15.5). Using semi-quantitative photon ratios, the variation becomes even less apparent (Figure 15.6). We cannot say how the variation in the Baantu Quarry source may influence the XRF analysis of archaeological artifacts, but the low strontium ppm values and the ranges in variation reveals a shortcoming in a calibration reference set that does not account for near 0 ppm values for particular elements (see below).

Our analysis indicates that when variation is considered among all specimens and all instruments, variation ranges from 4.9% to 14.9% for any a given element (Table 15.4). Magnitudes range from 3 to 182 ppm in the Caracol data. For many Mesoamerican samples Rb and Sr are used, and our study shows that

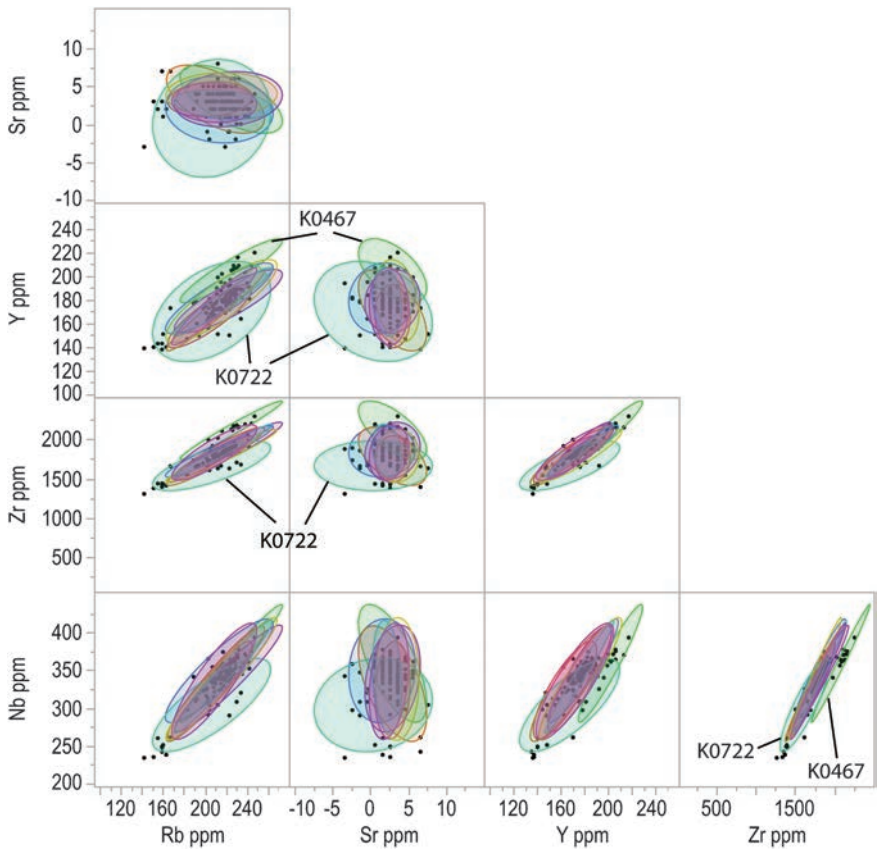

Figure 15.5 Matrix plot of Baantu Quarry obsidian grouped by pXRF instrument and 95% confidence ellipse to show variation by instrument.

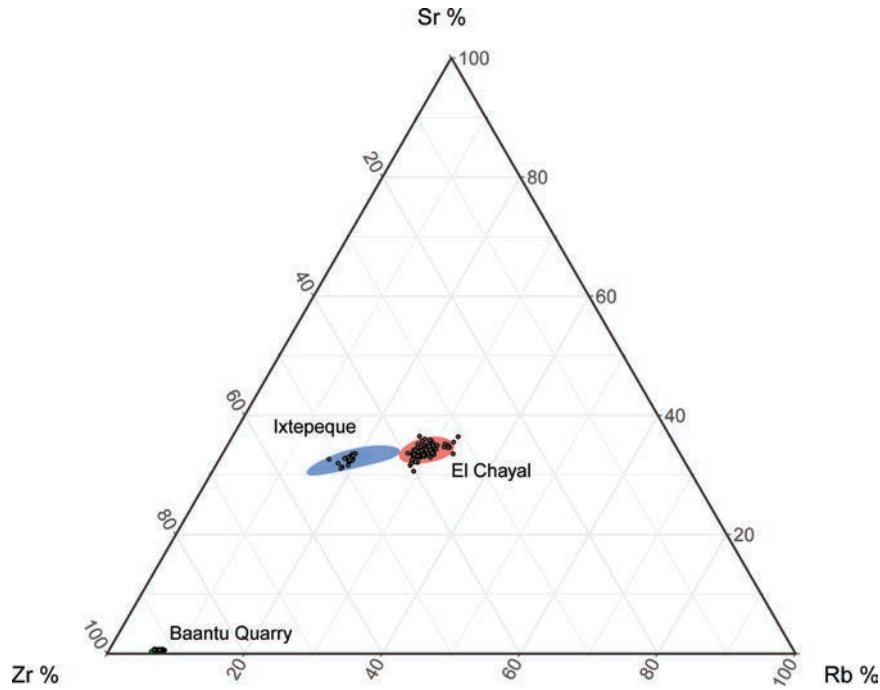

Figure 15.6 Ternary plot of all instruments and all analyzed samples using relative percentages of photon counts for Rb, Sr, and Zr listed in above tables. Note: obsidian sources are grouped by pXRF instrument and 95% confidence regions are calculated and drawn using ggtern in R.[29,30]

Table 15.4 Average CoV statistics expressed as a percent of variation for each source-group on all instruments.

	Parts per million					Photon count ratios		
	Rb	Sr	Y	Zr	Nb	Rb%	Sr%	Zr%
El Chayal	5.1	5.6	13.7	4.9	14.9	2.9	2.4	3.3
Ixtepeque	5.5	4.9	13.4	5.1	12.9	3.1	2.0	2.2
Baantu Quarry	5.4	48.1	8.4	7.2	6.8	4.6	16.3	0.4

variation did not exceed 5.6% variation on average. The variation is greater when including Sr for the Baantu Quarry sample (5.4–48.1% for the range 0–2292 ppm). Lower ppm results correlate with higher variability (*i.e.*, CoV), because smaller differences become more significant in lower abundance. Elemental variation is dependent on the unique characteristics of a given source and is not necessarily dependent on the instrument used in this study. Variation is even less so when inter-elemental semi-quantitative analysis of photon counts are utilized (Figure 15.6), signaling that photon analysis is a useful data output in some cases [though this risks non-reproducible results (Speakman and Shackley 2013)].

15.4.3 Analysis of Variation (ANOVA)

Two-way ANOVA tests with the instrument and artifact as independent variables were used to evaluate the proportion of variance each contributed (*i.e.*, dividing square errors by summed square errors). While our analysis of both instruments and artifacts demonstrated statistically significant levels of variation ($p < 0.05$), the effect was magnitudes of order larger for geochemical sources (>95%) than for instruments (<0.7%). Residual variance was in all cases larger than instrument variance but was notably larger for Rb (Table 15.5).

15.5 Discussion and Significance

Overall, all instruments performed to predictively identify trace element concentrations with only minor variations in ppm outputs. Data from Ethiopia showed low CoV for all elements except Sr despite differences in trace elemental concentrations (Figure 15.5). The low CoVs and large data spread indicate that variation is in the Baantu source and not instrumental measurements. Data from the Belize samples show more variation for Nb due to smaller concentrations (9–11 ppm).

The degree of difference in Sr CoV for the Ethiopian quarry samples is larger (40.9%–88.1%) because Sr is at or below all instrument detection limits. The magnitude of the shift in instruments is quite small, only 1.6 ppm. By contrast, Zr in the same deposit has on average 139 ppm difference, with a CoV of 7%. Higher concentrations of elements (~1800 ppm) tend to increase uncertainty in the upper ranges in this calibration, likely because there are few standards ($n = 1$) that constrain the empirical calibration, though the change in CoV is relatively modest.

Variance of photon data reveals the complexity of both photon-based analysis (*i.e.*, semi-quantitative) and statistically derived quantified data (ppm). For example, photon count ratios in both El Chayal and Ixtepeque source-groups recovered from archaeological contexts from Caracol showed less variation overall when compared to ppm. In the Ethiopian source, photons normalized to the Compton peak and photons per second outperformed the quantitative calibration for strontium – these data were less influenced by the low levels of strontium relative to quantified results.

Table 15.5 Proportion of variance calculated from two-way ANOVA tests. Instrument variation is much lower than either source or residuals.

Element	Instrument (%)	Artifact (%)	Residuals (%)
Rb	0.61	95.89	3.51
Sr	0.23	99.21	0.57
Y	0.26	98.80	0.93
Zr	0.37	98.97	0.66
Nb	0.18	99.42	0.40
Mean	0.33	98.46	1.21

The question of whether instrument quantification uncertainty (e_q) contributes significantly to overall variation is addressed in Table 15.5; for key elements the instrument in question contributes on average 0.33%, while the artifact ($e_q + e_c$) contributes on average 98.6%. Unaccounted for variance (e_i) is 1.2% on average. The simple act of turning spectra into quantitative data is not in itself sufficient for reproducibility; the key step in empirical quantification is having an external control on instrumental variability. While fundamental parameter methods offer the ability to produce quantitative data,[10,13] many algorithms do not correct for instrumental variability. It is incumbent on the researcher to demonstrate precision and accuracy – citing the assurance of an instrument manufacturer is not sufficient.

The results presented here demonstrate the importance of quantitative calibrations in examining precision and replicability. Quantitative values typically outperform photon-based materials analysis due to significant differences in count statistics between instruments. Nonetheless, these results also indicate that photon-based analysis can be replicable between different components, including tubes and detectors, with appropriate normalization. It is important to note that this would only be replicable as far as ED-XRF instrumentation is concerned for the same optical parameters (energy, current, filter, collimation). This is not viable for questions like obsidian sourcing, where relevant literature and other techniques use weight %/ppm as a standard unit. However, in non-destructive applications such as analysis of pigments, residues, and corrosion, this method might prove useful.

As illustrated here, even a well-established obsidian calibration, developed by MURR[15] in conjunction with Bruker Elemental (Nano), need to be taken as a starting point; any calibration's continuing utility should be based on additive processes which address research questions. All pXRF users should bear in mind that instrument manufacturers necessarily make choices about the flexibility of expanding calibrations to satisfy economic concerns; the mere existence of a calibration by a manufacturer does not guarantee instrument suitability to a research question.[7] For example, the occurrence of strontium concentrations below 0.2 ppm at or below the detection limit of XRF suggests a technical limitation. It is not sufficient to only have an internationally accepted reference set of materials, but also to possess the ability to revise and make available updated calibrations as needed to answer new questions or incorporate new regions. This means a large, sharable, and reliable set of global reference standards, that are possibly linked to internal laboratory standards *via* ICPMS or other instrumental data. It is up to the individual researcher to demonstrate XRF results of a reference set is reliable. Any set of reference materials for empirical calibration must be evaluated as appropriate by the researcher and augmented or refined as needed. In practice, a reliance on fundamental parameters and/or a reliance on factory calibrations may be appealing, as they ultimately may reduce the accuracy of analysis, and likely the precision as well. Arguably, they may invalidate determining the geochemical source of artifacts.[6,7]

The broader theoretical issue is how to reconcile validity, in which data is appropriate to a research question in archaeology,[3] and reliability, in which data of measurements are precise and accurate reflecting the true composition of an artifact or geological specimen. Here, we argue that validity and reliability are only accomplished if precision and accuracy is possible. For XRF analysis, validity and reliability is attained *via* the Lucas-Tooth equation in conjunction with an elementary diverse and representative reference set. This approach will most reliably and accurately produce values that correspond to the true chemical composition. The only way to do this is to ensure that results are reliable, or in other words, not dependent upon one machine or instrument but rather dependent upon a set of accepted international reference standards.[3] For non-uniform materials, a normalization procedure for net or gross photons may be appropriate. However, we acknowledge that such results are not as replicable between instruments as quantitative values.

15.6 Conclusion

Handheld and portable ED-XRF systems have the capability to produce precise and accurate results across a range of X-ray tube and detector components, thus being able to distinguish between obsidian sources and answering archaeological inquiries.[3,5,31] As the handheld portable XRF community grows, it becomes increasingly important that further inter-instrument comparisons continue, specifically, those from a single manufacturer should be performed within a regional context with systematically assayed source rock. Concentrating scrutiny on different models of a single XRF instrument type from manufacturers of XRF instruments should ensure the science is reproducible, reliable, and methodologically transparent.

We also argue that this study further reinforces the procedures for disclosing and sharing data between obsidian handheld portable XRF analysts. This is especially true for analysts who do not have access to source rock materials. Sharing data between analysts who use the same analytical techniques should be better than extracting quantitative data from published manuscript tables or appendices. This is particularly true when semi-quantitative analysis is conducted using photon counts rather than ppm. We encourage analysts to publish both ppm and photon count data.

Lastly, XRF technology, like other analytical techniques continue to develop in various aspects. First, portable XRF instruments manufactured after this study show greater resolution (<140 eV @ 250 000 counts per second Mn Ka) and analytical speed (<60 seconds for obsidian with infinite thickness). Perhaps more importantly, various calibration procedures are available beyond manufacturer defaults. At least two developments in this regard are available. The first is CloudCal, which offers an open-source multi-linear regression option based in R.[32] The second is the introduction of the Peabody–Yale Reference Obsidians or PYRO offering another matrix matched reference set to customize an existing calibration or to fine-tune an existing method

(*e.g.*, Fundamental Parameters) *via* post-analysis slope correction.[33,34] Recently, this set along with the calibration set used herein (MURR/Bruker) was evaluated by several instruments to examine accuracy, precision, and validity under nominal operating conditions recommended by manufactures to analyzed rhyolitic glass.[35] The Johnson *et al.* (2021) study offers an effective method for evaluating systematic error in predicting ppm values *via* various calibration procedures prior to the analysis of unknown archaeological materials. The advent of multiple obsidian calibration sets coupled with alternative algorithms should allow greater transparency and further expand open-source data sharing that is not necessarily bound by manufacturer presets. As archaeologists continually search for non-destructive analytical techniques, these developments, while requiring more knowledge by the user, do improve inter-comparability and data transparency.

Acknowledgements

We would like to thank each analyst for taking their time in analyzing each specimen. The Institute of Archaeology in Belize enabled the export of obsidian samples for this study and the subsequent dissertation project that was funded by the National Science Foundation (#1556260) and made possible by the Caracol Archaeological Project directed by Diane Z. Chase and Arlen F. Chase. We would also like to thank Daron Duke of the Far Western Anthropological Group, Inc. for his comments and assistance in refining some of the text before submission.

References

1. J. R. Cann and C. Renfrew, The characterization of obsidian and its application to the Mediterranean Region, *Proc. Prehistoric Soc.*, 1964, **30**, 111–133.
2. *Handheld XRF for Art and Archaeology*, ed. A. N. Shugar and J. L. Mass, Leuven University Press, Belgium, 2012.
3. R. E. Hughes, On reliability, validity, and scale in obsidian sourcing research, in *Unit Issues in Archaeology: Measuring Time, Space, and Material*, ed. A. F. Ramenofsky and A. Steffen, University of Utah Press, 1998, pp. 103–114.
4. N. Craig, R. J. Speakman, R. S. Popelka-Filcoff, M. D. Glascock, J. D. Robertson, M. S. Shackley and M. S. Aldenderfer, Comparison of XRF and PXRF for analysis of archaeological obsidian from southern Peru, *J. Archaeol. Sci.*, 2007, **34**(12), 2012–2024.
5. A. J. Nazaroff, K. M. Prufer and B. J. Drake, Assessing the applicability of portable X-ray fluorescence spectrometry for obsidian provenance research in the Maya lowlands?, *J. Archaeol. Sci.*, 2010, **37**(4), 885–895.
6. E. Frahm, Non-destructive sourcing of Bronze Age Near Eastern obsidian artefacts: Redeveloping and reassessing electron microprobe analysis for obsidian sourcing, *Archaeometry*, 2012, **54**, 623–642.

7. R. J. Speakman and M. S. Shackley, Silo science and portable XRF in archaeology: A response to Frahm, *J. Archaeol. Sci.*, 2013, **40**, 1435–1443.
8. N. Broll and P. de Chateaubourg, Spectral distribution from end window X-ray tubes, *Adv. X-ray Anal.*, 1999, **41**, 393.
9. H. J. Lucas-Tooth and B. J. Price, A mathematical method for the investigation of interelement effects in X-ray fluorescence analysis, *Metallurgia*, 1961, **64**(383), 149–161.
10. G. L. Bosco, Development and application of portable, hand-held X-ray fluorescence spectrometers, *James L. Waters Symposium 2012 Report Trends Anal. Chem.*, 2013, **45**, 121–134.
11. A. F. Ramenofsky and A. Steffan Units as tools of measurement, in *Unit Issues in Archaeology: Measuring Time, Space, and Material*, ed. A. F. Ramenofsky and A. Steffen, University of Utah Press, Salt Lake City, 1998, pp. 3–17.
12. B. L. Drake, Portable XRF: A (very) brief introduction, in *Lights On… Cultural Heritage and Museums*, ed. P. M. Homem, LabCR|FLUP, Porto, 2016, pp. 140–161.
13. R. M. Rousseau, The quest for a fundamental algorithm in X-ray fluorescence analysis and calibration, *Open Spectrosc. J.*, 2009, **3**, 31–42.
14. M. D. Glascock and J. R. Ferguson, *Report on the Analysis of Obsidian Source Samples by Multiple Analytical Methods*, University of Missouri Research Reactor, Columbia, 2012.
15. R. J. Speakman, *Evaluation of Bruker's Tracer Family Factory Obsidian Calibration for Handheld Portable XRF Studies of Obsidian*, Report Prepared for Bruker AXS, Kennewick, WA, 2012.
16. A. F. Chase and D. Z. Chase, *Investigations at the Classic Maya City of Caracol, Belize: 1985–1987. Monograph 3*, PARI, San Francisco, 1987.
17. A. F. Chase, D. Z. Chase, Weishampel and J. F. Lasers, in the jungle: Airborne sensors reveal a vast Maya landscape, *Archaeology*, 2010, **64**(4), 29–31.
18. A. F. Chase, D. Z. Chase, J. F. Weishampel, J. B. Drake, R. L. Shrestha, K. C. Slatton, J. J. Awe and W. E. Carter, Airborne LiDAR, archaeology, and the ancient Maya landscape at Caracol, Belize, *J. Archaeol. Sci.*, 2011, **38**, 387–398.
19. A. F. Chase, D. Z. Chase, J. J. Awe, J. F. Weishampel, G. Iannone, H. Moyes, J. Yaeger, M. K. Brown, R. L. Shrestha, W. E. Carter and J. Fernandez-Diaz, Ancient Maya regional settlement and inter-site analysis: The 2013 west-central Belize LiDAR survey, *Remote Sens.*, 2014, **6**, 8671–8695.
20. D. Z. Chase and A. F. Chase, *Studies in the Archaeology of Caracol, Belize, Monograph 7*, PARI, San Francisco, 1994.
21. D. Z. Chase and A. F. Chase, Caracol, Belize and changing perceptions of ancient Maya society, *J. Archaeol. Res.*, 2017, **25**(3), 15–249.
22. L. R. M. Johnson, *Toward an Itinerary of Stone. Investigating the Movement, Crafting, and Use of Obsidian from Caracol, Belize*, Dissertation, University of Florida, 2016.

23. R. E. Hughes, Determining the geologic provenance of tiny obsidian flakes in archaeology using nondestructive EDXRF, *Am. Lab.*, 2010, **42**(7), 27–31.
24. S. A. Brandt, E. C. Fisher, E. A. Hildebrand, R. Vogelsang, S. H. Ambrose, J. Lesur and H. Wang, Early MIS 3 occupation of Mochena Borago Rockshelter, Southwest Ethiopian Highlands: implications for Late Pleistocene archaeology, paleoenvironments and modern human dispersals, *Q. Int.*, 2012, **274**, 38–54.
25. S. Brandt, E. Hildebrand, R. Vogelsang, J. Wolfhagen and H. Wang, A new MIS 3 radiocarbon chronology for Mochena Borago Rockshelter, SW Ethiopia: Implications for the interpretation of Late Pleistocene chronostratigraphy and human behaviour, *J. Archaeol. Sci.: Rep.*, 2017, **11:**, 352–369.
26. A. M. Warren, *Preliminary Characterization and Provenance of Obsidian Artifacts from Ethiopian Archaeological Sites Using Portable X-Ray Fluorescence*, BA thesis, University of Florida, Gainesville, 2010, unpublished.
27. J. R. Ferguson X-ray, fluorescence of Obsidian: Approaches to calibration and the analysis of small samples, in *Handheld XRF for Art and Archaeology*, ed. A. Shugar and J. L. Mass, Leuven University Press, Belgium, 2012, pp. 401–422.
28. K. M. Davis, T. L. Jackson, M. S. Shackley, T. Teague and J. H. Hampel. Factors affecting the energy-dispersed X-ray fluorescence (EDXRF) analysis of archaeological Obsidian, in *X-Ray Fluorescence Spectrometry (XRF) in Geoarchaeology*, ed. S. Shackley, Springer, New York, 2011, pp. 45–64.
29. N. Hamilton, *ggtern: An Extension to 'ggplot2', for the Creation of Ternary Diagrams. R package version 2.2.2*, 2018, available at https://CRAN.R-project.org/package=ggtern.
30. R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020, Available at http://www.R-project.org/
31. E. Frahm, Evaluation of archaeological sourcing techniques: Reconsidering and re-deriving Hughes' four-field assessment scheme, *Geoarchaeology*, 2012, **27**, 166–174.
32. L. B. Drake, *CloudCal v3.0. GitHub*, 2018, available at: https://github.com/leedrake5/CloudCal
33. E. Frahm, Introducing the Peabody–Yale reference obsidians (PYRO) sets: open-source calibration and evaluation standards for quantitative X-ray fluorescence analysis, *J. Archaeol. Sci. Rep.*, 2019, **27**, 101957.
34. E. Frahm, Corrigendum to ''Introducing the Peabody-Yale Reference Obsidians (PYRO) sets: Open-source calibration and evaluation standards for quantitative X-ray fluorescence analysis'' [*J. Archaeol. Sci.: Rep.*, 2019, **27**, 101957], *J. Archaeol. Sci.: Rep.*, 2022, **41**, 102937.
35. L. R. M. Johnson, J. Ferguson, K. P. Freund, B. Lee Drake and D. Duke, Evaluating obsidian calibration sets with portable X-ray fluorescence (ED-XRF) instruments, *J. Archaeol. Sci. Rep.*, 2021, **39**, 103126.