Commentary

Silo science and portable XRF in archaeology: a response to Frahm

Robert J. Speakmana,*, M. Steven Shackleyb

a Center for Applied Isotope Studies, University of Georgia, Athens, USA
b Geoaarchaeological XRF Laboratory, University of California, Berkeley, USA

ABSTRACT

A recent study by Frahm (in press) on the application of portable XRF (PXRF) for chemical characterization of obsidian ignores fundamental issues of reliability and validity in the measurements, and justifies “internally consistent” measurements as acceptable. We argue this form of science is unacceptable, point out several flaws in Frahm’s paper, and provide some examples of PXRF measurements that are valid and reliable and conform to international standards as published.

Keywords: PXRF Obsidian Reliability Validity Silo science

The recent literature has been spate with experiments and analyses of archaeological materials, often obsidian, with portable X-ray fluorescence spectrometry (PXRF). Many have been focused on the project at hand, and some have been concerned with the quality of the results with regard to the study. A common thread that concerns us is consistently ignoring the issues of reliability and validity in the measurements, and justifying these “internally consistent” results as necessary and sufficient for compositional analysis. While the philosophical underpinnings of this perspective are unfortunate, particularly in Americanist archaeology, the long-term results might be more problematic. Ellery Frahm’s recent study (Frahm, in press) is especially troubling given that it is obvious that he fully understands the importance of reliable and valid studies in archaeology (Frahm, 2012). If, as promulgated by Frahm and others, that it is perfectly fine to provide results that are only internally consistent, and do not conform to established international standards and data, then we are entering a time of silo science” where each researchers’ data is self-contained, independent, and cannot be verified externally. We find this to not only be unacceptable, but another aspect of a social science that is ejecting the “science” aspect of our discipline to “play scientist” with portable XRF technology. Frahm’s “internally consistent” results using the Niton PXRF, while valid as far as it goes, is not necessary or acceptable. There are other PXRF options that conform to international standards as published, and are valid and reliable in toto.

As Grave et al. (2012) have noted (which Frahm acknowledges in his paper), “archaeological applications [of PXRF] continue to be cautiously treated, principally because of a perceived lack of analytic rigor,” Grave et al. are absolutely correct on this point. In publishing his paper, Frahm essentially argues that a lack of analytical rigor is acceptable—trust me my results are internally consistent. Most archaeological PXRF users lack experience—whether it be X-ray physics, basic analytical chemistry, statistics, or the fundamentals of provenance studies. Many users approach PXRF from a “black box” perspective in which the inner workings of the XRF instrument are not understood by the user, nor does the user care to learn how and why the instrument functions—the only importance is that the sample is analysed and numbers are generated (Speakman et al., 2010, 2011). This often results in the user not recognizing the most basic problems that can and do occur with XRF analyses. Because of the manner in which the capabilities of PXRF are [mis]represented by many manufacturers (e.g., promoting ideas of internal consistency, ability to measure low Z elements accurately, etc.), users tend to disregard conventional knowledge concerning XRF fundamentals. Consequently, we believe that many PXRF-based studies of pottery, obsidian, metals, and other materials that are presented at meetings and/or published are founded on poor science. A professional photographer recently told one of us that as a consequence of widespread availability of digital cameras, now everyone is a photographer. It would seem that the same holds true for portable analytical instrumentation—now everyone can be a scientist.
While we agree that archaeological interpretations can be made from Frahm’s data, the ultimate results do not fulfill the rules of validity and reliability, the foundation of good scientific research. If the results of any experiment cannot be compared and evaluated by a subsequent experiment outside the original experiment, then it is unreliable even though it is internally consistent. X-ray fluorescence spectrometry (what Frahm refers to as labXRF) in archaeology has attempted to rely on and refine valid and reliable measurements such that any XRF or NAA laboratory can use those results and analyses of archaeological or source data to arrive at the same result. Most scientists would agree that this is not only a laudable goal, but a necessary one in order to avoid “silo science” in archaeology—i.e., my results are internally consistent and so I don’t care if it’s correct within the discipline or the results can be tested externally. A recent PXRF obsidian analysis exercise at the 2012 Society for American Archaeology meeting (organized by the authors, Michael Glascock, and Arlen Heginbotham) was structured to address this very issue (publication of this study is forthcoming).

One can easily interpret Frahm’s paper as approving this emerging “silo science”, and that reliability and validity no longer is the cornerstone of scientific inquiry—at least in compositional studies in archaeology. Although Frahm’s paper deals specifically with obsidian, while analytically quite straightforward to chemically characterize, there are unintended consequences. Specifically, there are those who would interpret and apply his approach to the study of ceramics, soils, bone chemistry, and other material classes.

In his introduction, Frahm states that to date fewer than two dozen studies based on PXRF of obsidian have been published. This statement is incorrect. We are aware of about 70 publications of PXRF of obsidian (Appendix A); we suspect that many more exist in the Asian, European, and South American literature.

Frahm’s introduction states that to achieve extreme portability, these analysers sacrifice performance. While this is true of certain models of all manufactures, the reality is that many PXRF systems currently on the market have superior detector resolution and electronics than laboratory-based instruments manufactured 5–10 years ago. More specifically, Frahm states that PXRF instruments use less-than-ideal geometric arrangements of miniaturised, low-power components that run on batteries, and the processing of X-rays is done by on-board electronics and software rather than external systems. We are unclear as to what this “less-than-ideal” geometry specifically refers to, but we have serious doubts that any manufacturer would fail to optimize their tube-sample-detector geometry. We know that Bruker PXRF systems are configured with the tube and detector at a 62° angle with sample placement occurring at the convergence of this angle; Thermo laboratory-based systems are about 70° with sample placement occurring at the convergence of this angle. Other XRF systems (portable and fixed laboratory) use angles that range from 45 to 90°, none of which create problems with current XRF engineering. Frahm also cites miniaturization of components as problematic—Why? Miniaturization of electronic components has been ongoing for decades in all areas of technology. For example, radios and televisions no longer use transistors and modern cell phones are only a fraction of the size they were 15 years ago. Did miniaturization of electronic components for radios, television, and phones result in a sacrifice to performance? We think not. Finally, we fail to understand why Frahm would state that the processing of X-rays being done by onboard electronics is a sacrifice—that’s where the X-ray detector is located. There’s no reason that this function needs to occur outside the “box” whether it be a portable or a fixed-laboratory system—we have come a long way from the days when a single instrument required a large space that was dedicated to a mainframe computer, multichannel analysers, amplifiers, and other assorted electronic components.

Although Frahm designates acronyms for laboratory-based energy dispersive XRF (Frahm’s labXRF) and portable-XRF (Frahm’s HHpXRF), it is apparent that Frahm truly does not understand the instrumentation used to generate data for his own publications. Frahm states that his labXRF analyses of obsidian were conducted at the University of Missouri Research Reactor (MURR), and according to Frahm (2010: 403) the instrument used to analyse the artifacts at MURR (and confirmed via personal communication from Michael Glascock to RJS) was an ElvaX EDXRF system. The ElvaX instrument is designed and marketed not as a laboratory XRF system, but rather a portable-XRF instrument. Given that one of us (Speakman) has used this specific instrument to conduct research in Alaska and Panama—outside of the laboratory—and MURR scientists have deployed the ElvaX instrument in Argentina (Barberena et al., 2011; Giesso et al., 2011), the ElvaX system clearly does not fit Frahm’s definition of labXRF. Consequently in Frahm’s manuscript, Figs. 6a–I and 7a–c are not figures that plot traditional laboratory-based EDXRF data (Frahm’s labXRF) versus PXRF data (Frahm’s HHpXRF) or laboratory EDXRF data versus NAA data, but rather these figures are simply comparing (1) data generated from two different portable-XRF instruments and (2) PXRF data to NAA data. To reiterate, these figures do not project any data generated by a laboratory-based EDXRF system as Frahm has indicated.

However, it is of interest here that Frahm’s plots of labXRF (really PXRF) data versus NAA data show higher correlated values than plots comparing the ElvaX PXRF data to Niton PXRF data. The reason for this is quite simple—the MURR ElvaX instrument was calibrated by Mike Glascock using obsidian that had been analysed by INAA and/or other analytical methods at MURR with the express intent of generating data that would be as comparable to MURR’s extant NAA obsidian database. In other words the instrument was calibrated to generate valid and reliable results for obsidian (for an example, see INAA and PXRF data published in Knight et al., 2011). In contrast, the Niton XRF data are poorly correlated with MURR PXRF data because the Niton was not calibrated for obsidian.

In Fig. 1, we provide a comparison of data that compare data from a Thermo Fisher laboratory-based EDXRF system and a Bruker PXRF. The calibration for the Thermo XRF was created by Shackley using international powdered geologic reference materials whereas the PXRF data were generated using a ‘canned’ factory calibration based on the analysis of 40 or so obsidian samples provided to Bruker by MURR (Glascock and Ferguson, 2012). In contrast to Frahm’s XRF comparison figures, we observe highly correlated values for the 12 plotted samples. In some cases there are systematic offsets between the Thermo and Bruker data, but because the two datasets are highly correlated, it is possible to correct the data through the use of a secondary reference material (i.e., RGM-1, RGM-2, SRM-278)—the same method that any two laboratories, XRF or otherwise, would use to intercalibrate data. This underscores the importance of publishing data not only for archaeological samples, but also data generated for international reference standards as has been the practice of Shackley and others for more than 20 years.

Frahm states that “Empirical correction curves using custom calibrations are regarded as the best practice for labXRF but are rarely used in HHpXRF obsidian sourcing”—this statement simply

---

1 In fact, the manufacturer has taken the internal components of this small portable benchtop system and repackaged them into a handheld configuration referred to as the ProSpector. The internal electronics (tube, detector, multichannel analyser) and software for the two systems are virtually identical—the only difference is packaging.
is not accurate. Beginning in 2006, Speakman and Mike Glascock began using empirical calibration schemes based on obsidian reference materials to calibrate the ElvaX PXRF, initially at MURR and later at the Smithsonian. In 2007, Speakman in collaboration with Bruker developed the “green” filter that is commonly used for obsidian analyses involving Bruker instruments and at the same time developed the first empirical obsidian calibration scheme for the Bruker PXRF. Since then, we along with our MURR colleagues have continued to refine calibration standards for empirical analyses. These standards have been used on dozens of instruments since 2008. XRF, portable or lab-based, requires one to create element references, standards libraries, and calibrations that can be constantly modified. XRF instruments rarely have “off the shelf” calibrations, and if they do they are typically instructional and ultimately require the user to make some adjustments. Having said that, some vendors do not allow the user to make such adjustments.

Over the past decade, we have collectively examined almost every major PXRF instrument on the market. The one single issue that we have continually raised with vendors has been centred around this fundamental issue of software and the ability of the user to create and modify empirical calibrations. Ultimately, this is why some manufacturers have been more successful than others in selling instruments for archaeological-based research. We believe that PXRF manufacturers can and should provide better software—there is no compelling reason not to. For example, Thermo-Fisher manufactures several high-quality lab based XRF instruments. The software for these instruments is fully developed and there is absolutely no reason why they could not provide that software for use with spectra generated on their Niton handheld instruments. The same holds true for other manufacturers. Historically there is a reason for this—commercial portable units

---

**Fig. 1. Direct comparison of values (expressed as ppm) obtained for 12 samples using a portable XRF (x-axis) and laboratory-based EDXRF (y-axis).**

---

2 In Appendix A, data for all papers that include Glascock or Speakman as coauthors are based on custom empirical calibrations. Furthermore, all publications in Appendix A that are generated using Bruker instruments also are based on custom empirical calibrations. Likewise, most PXRF measurements reported in Heginbotham et al. (2011)—a round robin of copper-based metals—are based on custom empirical calibrations.
were designed intentionally to be black boxes for use by novices outside of a laboratory environment primarily in the mining and metals recycling industries. They were never really intended, at least initially, to be true research instruments (or really developed with the intention for analysis of silicates), hence the software and user interfaces were “dummied” down to limit analytical choices and facilitate ease of use.

Although Frahm and others (e.g., Dybowski, 2012) have convincingly demonstrated that some PXRF instruments cannot be used to generate valid and reliable obsidian data (using factory calibrations), we believe some viable options do exist. For example, Bruker has recently released an obsidian calibration based on a set of forty slab-cut obsidian source samples commissioned from the MURR Archaeometry Laboratory (Glascok and Ferguson, 2012). Analysed by multiple analytical techniques, the set superseded (in terms of accuracy) earlier calibrations developed by MURR and Speakman. A cursory examination of data generated using this calibration suggests high degrees of accuracy, precision, and reproducibility (see Speakman, 2012 for the full unpublished report).

In an experiment designed to assess reproducibility, a Bruker PXRF was set up to analyse a single obsidian sample continuously for 17 h with spectra being saved every 200 s. All samples were measured at 40 kV, 25 μA, with a 12 mil Al, 1 mil Ti, 6 mil Cu filter placed in the X-ray path for a 200-s live-time count. Peak intensities for the Kα peaks of Mn, Fe, Zn, Ga, Rb, Sr, Y, Zr, Nb, and Lα peak of Th were calculated as ratios to the Compton peak of rhodium, and converted to parts-per-million (ppm) using Bruker’s factory-installed calibration for obsidian. This resulted in data for 307 analyses that are summarized in Table 1. Individual data points for these analyses are plotted in Fig. 2 (Zn, Ga, Th, Rb, Sr, Y, Zr, Nb) and Fig. 3 (Mn and Fe). Data for these 307 analyses exhibit relatively low variation (expressed as %RSD). For Fe, Rb, Y, Zr, and Nb, the %RSD is 2% or lower which is comparable to data generated on most laboratory-based EDXRF instruments. Mn, Zn, and Ga exhibit %RSD values of ca. 3–6% which is typical for these elements. Sr has the highest error, as would be expected given the low concentration in this particular sample (which also approaches the limits of EDXRF detection for Sr in silicate matrices). In other words, the high %RSD for Sr in this sample is entirely due to counting statistics, and in no way a reflection of the instrument itself, beyond inherent instrumental error.

A Horwitz curve (Horwitz et al., 1980) (Fig. 4) for the 40 obsidian calibration standards illustrates our point concerning Sr. Horwitz curves are commonly used to evaluate upper and lower ends of calibration curves. In this plot Sr concentrations (x-axis, based on the average value of 5 measurements) are plotted against %RSD (y-axis). A power trendline is fit against all data points. In this particular plot the x-axis is constrained to samples with less than 100 ppm Sr to facilitate visual evaluation of the low end of the curve. The plot shows (as expected) that as concentration decreases (i.e., lower counting statistics), %RSD increases; thereby illustrating the point that measurement error is correlated with concentration (counting statistics). The closer one gets to instrument detection limits, the greater the analytical error. At concentrations of about 7 ppm Sr we observe less than 5%RSD. At concentrations above 15 ppm Sr the error decreases to about 2–3%RSD. To reiterate, the high %RSD reported in Table 1 is not a reflection of XRF performance; similar results would be obtained with any properly calibrated EDXRF spectrometer regardless of manufacturer.

Of particular import for Figs. 2 and 3 is that the overall trend for each element is “flat”, indicating that instrumental drift (stability) is not an issue. For researchers conducting analyses under field conditions (or in museums), reanalysis of samples is oftentimes not possible. An instrument that drifts throughout the day will impart greater analytical error on the experiment. For obsidian studies this could mean that distinctions among compositional groups could be obscured. Likewise it is important to note that the relative deviation between analyses is minimal and that no outliers are present. In most laboratory systems, the energies can be calibrated at any point to mitigate this problem; many PXRF systems do this as well simply by restarting the instrument.

We also assessed the accuracy of Bruker’s obsidian calibration against independent quality control standards. As stated above and elsewhere (Shackley, 2010), such standards must periodically be analysed and published to establish validity. Tables 2 and 3 present data obtained from replicate analyses of pressed-powdered international obsidian standards (RGM-1 and NIST SRM-278) relative to their recommended values and published literature values. In all cases the Bruker data are acceptable, thereby establishing validity.

Ultimately though, the problem with converting the spectra to ppm and associated problems related to accuracy and reproducibility lie with the user, not the manufacturer. In Frahm’s case an argument could be made that the Niton PXRF will not allow the user to do this. However, the fact of the matter is that Niton spectra

Table 1

<table>
<thead>
<tr>
<th>Element</th>
<th>Mn</th>
<th>Fe</th>
<th>Zn</th>
<th>Ga</th>
<th>Th</th>
<th>Rb</th>
<th>Sr</th>
<th>Y</th>
<th>Zr</th>
<th>Nb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>438.7</td>
<td>7446.2</td>
<td>134.1</td>
<td>270</td>
<td>41.7</td>
<td>364.4</td>
<td>1.1</td>
<td>83.9</td>
<td>163.7</td>
<td>237.5</td>
</tr>
<tr>
<td>St dev</td>
<td>23.7</td>
<td>72.6</td>
<td>4.6</td>
<td>1.7</td>
<td>1.5</td>
<td>3.4</td>
<td>0.5</td>
<td>1.6</td>
<td>1.7</td>
<td>2.4</td>
</tr>
<tr>
<td>%RSD</td>
<td>5.4</td>
<td>1.0</td>
<td>3.4</td>
<td>6.2</td>
<td>3.5</td>
<td>0.9</td>
<td>0.9</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

* High %RSD (% relative standard deviation) a consequence of low Sr concentration in this sample which results in low counting statistics. In no way is this a reflection of the instrument used for the analysis (all EDXRF instruments would exhibit similar errors at this concentration and matrix). For samples containing more than 15 ppm Sr, 2–3%RSD is the norm.

Please cite this article in press as: Speakman, R.J., Steven Shackley, M., Silo science and portable XRF in archaeology: a response to Frahm. Journal of Archaeological Science (2012), http://dx.doi.org/10.1016/j.jas.2012.09.033
Fig. 3. Results in ppm (y-axis) versus time (x-axis) for 307 consecutive 200-s counts of Bruker obsidian standard XR06. Note that the overall trend for each element is “flat” indicating that instrumental drift is not an issue. Likewise note that the relative deviation between analyses is minimal and that no outliers are present. Refer to Table 1 for summary statistics for this experiment.

Fig. 4. Horwitz curve for the 40 obsidian calibration standards used to calibrate Bruker PXRF instrument. Sr concentration (x-axis) is based on the average value of five measurements. Curve is fit to all data points, but the plot is constrained to samples with less than 100 ppm Sr to facilitate the evaluation of the low end of the curve. This plot shows that as concentration decreases (i.e., lower counting statistics), the %RSD increases thereby illustrating the point that measurement error is correlated with concentration. The lower one gets to instrument detection limits, the greater the analytical error. Again, this is not a reflection of instrument performance; similar results would be obtained with any properly calibrated EDSRF spectrometer.
instrumentation. Perhaps, as one of us has argued elsewhere, the acceptance of unreliable and invalid data in archaeology is pervasive and allows for users of PXRF instrumentation to readily accept this “internally consistent” silliness (Shackley, 2010). Why would any archaeologist accept results that could not be verified, unless they simply did not understand the need for verification. If that is indeed the case, 21st century archaeology is moving down a hazardous path. And lastly, we still cannot understand why Frahm would want to analyse materials in this way and publish it in an archaeological science journal.

Conflict of interest statement

Robert J. Speakman maintains a professional relationship with Bruker AXS specifically with respect to instrument and application development in archaeological science. Speakman’s laboratory, The Center for Applied Isotope Studies at the University of Georgia, occasionally conducts work for Bruker AXS on a fee for service basis. Speakman also is a principal in Rusty Trowel, LLC (registered in the State of Georgia), a company that provides consulting services, in the form of XRF instrument training, to Bruker AXS customers. M. Steven Shackley declares no competing interests.

Acknowledgements

We thank Mike Glascock, Eric Dyrdahl, and two anonymous reviewers for commenting on earlier drafts of this paper. Mike Glascock, Mark McCoy, Eric Dyrdahl, and others also are acknowledged for their assistance in compiling the PXRF obsidian bibliography.

appendix A. Bibliography of PXRF Obsidian publications


Please cite this article in press as: Speakman, R.J., Steven Shackley, M., Silo science and portable XRF in archaeology: a response to Frahm, Journal of Archaeological Science (2012), http://dx.doi.org/10.1016/j.jas.2012.09.033

Table 2

<table>
<thead>
<tr>
<th>Mn</th>
<th>Fe</th>
<th>Zn</th>
<th>Th</th>
<th>Rb</th>
<th>Sr</th>
<th>Y</th>
<th>Zr</th>
<th>Nb</th>
</tr>
</thead>
<tbody>
<tr>
<td>布鲁克林（n = 5）</td>
<td>321 ± 8</td>
<td>13,075 ± 69</td>
<td>40 ± 2</td>
<td>16 ± 1</td>
<td>157 ± 3</td>
<td>104 ± 1</td>
<td>26 ± 1</td>
<td>223 ± 3</td>
</tr>
<tr>
<td>USGS推荐</td>
<td>279 ± 50</td>
<td>13,010 ± 210</td>
<td>32</td>
<td>15 ± 1,3</td>
<td>150 ± 8</td>
<td>110 ± 10</td>
<td>25</td>
<td>220 ± 20</td>
</tr>
<tr>
<td>Shackley（2012）</td>
<td>302 ± 14</td>
<td>13,116 ± 308</td>
<td>n.r.</td>
<td>16 ± 3</td>
<td>151 ± 3</td>
<td>106 ± 3</td>
<td>25 ± 2</td>
<td>219 ± 5</td>
</tr>
<tr>
<td>斯金纳和Davis（1996）</td>
<td>291 ± 47</td>
<td>13,480 ± 745</td>
<td>37 ± 7</td>
<td>n.r.</td>
<td>152 ± 3</td>
<td>107 ± 9</td>
<td>24 ± 3</td>
<td>217 ± 8</td>
</tr>
<tr>
<td>休斯（2007）</td>
<td>278 ± 10</td>
<td>13,079 ± 140</td>
<td>n.r.</td>
<td>143 ± 4</td>
<td>105 ± 3</td>
<td>23 ± 3</td>
<td>214 ± 4</td>
<td>8 ± 3</td>
</tr>
</tbody>
</table>

n.r.—不报告。

Table 3

<table>
<thead>
<tr>
<th>Mn</th>
<th>Fe</th>
<th>Zn</th>
<th>Th</th>
<th>Rb</th>
<th>Sr</th>
<th>Y</th>
<th>Zr</th>
<th>Nb</th>
</tr>
</thead>
<tbody>
<tr>
<td>布鲁克林（n = 5）</td>
<td>432 ± 27</td>
<td>14,521 ± 100</td>
<td>55 ± 4</td>
<td>13 ± 2</td>
<td>133 ± 2</td>
<td>62 ± 1</td>
<td>41 ± 1</td>
<td>281 ± 2</td>
</tr>
<tr>
<td>USGS推荐</td>
<td>403 ± 2</td>
<td>14,269 ± 140</td>
<td>n.r.</td>
<td>12.4 ± 0.3</td>
<td>127.5 ± 0.3</td>
<td>83.5 ± 0.1</td>
<td>n.r.</td>
<td>n.r.</td>
</tr>
<tr>
<td>格拉斯科（2006）</td>
<td>397 ± 23</td>
<td>14,500 ± 900</td>
<td>53 ± 5</td>
<td>12.6 ± 0.6</td>
<td>133 ± 6</td>
<td>64 ± 5</td>
<td>39 ± 5</td>
<td>290 ± 30</td>
</tr>
<tr>
<td>Shackley（2012）</td>
<td>383 ± 7</td>
<td>14,329 ± 37</td>
<td>n.r.</td>
<td>15 ± 5</td>
<td>130 ± 2</td>
<td>67 ± 1</td>
<td>40 ± 2</td>
<td>276 ± 2</td>
</tr>
</tbody>
</table>

n.r.—不报告。


Please cite this article in press as: Speakman, R.J., Steven Shackley, M., Silo science and portable XRF in archaeology: a response to Frahm. Journal of Archaeological Science xxxx (2012) 1–9.
References


