RESEARCH PAPER

Cross-validation and evaluation of the performance of methods for the elemental analysis of forensic glass by μ-XRF, ICP-MS, and LA-ICP-MS

Tatiana Trejos · Robert Koons · Stefan Becker · Ted Berman · JoAnn Buscaglia · Marc Duecking · Tiffany Eckert-Lumsdon · Troy Ernst · Christopher Hanlon · Alex Heydon · Kim Mooney · Randall Nelson · Kristine Olsson · Christopher Palenik · Edward Chip Pollock · David Rudell · Scott Ryland · Anamary Tarifa · Melissa Valadez · Peter Weis · Jose Almirall

Abstract Elemental analysis of glass was conducted by 16 forensic science laboratories, providing a direct comparison between three analytical methods [micro-x-ray fluorescence spectroscopy (μ-XRF), solution analysis using inductively coupled plasma mass spectrometry (ICP-MS), and laser ablation inductively coupled plasma mass spectrometry]. Interlaboratory studies using glass standard reference materials and other glass samples were designed to (a) evaluate the analytical performance between different laboratories using the same method, (b) evaluate the

T. Trejos · A. Tarifa · J. Almirall
International Forensic Research Institute,
Florida International University, OE116 11200 SW 8th St,
Miami, FL 33199, USA

R. Koons · J. Buscaglia
Federal Bureau of Investigation, Laboratory Division,
Counterterrorism and Forensic Science Research Unit,
2501 Investigation Parkway,
Quantico, VA 22135, USA

S. Becker · M. Duecking · P. Weis
Bundeskriminalamt (BKA), Forensic Science Institute,
65173 Wiesbaden, Germany

T. Berman · S. Ryland
Florida Department of Law Enforcement,
Orlando Regional Operations Center, 500 W. Robinson St.,
Orlando, FL 32801, USA

T. Eckert-Lumsdon · K. Mooney
US Army Criminal Investigation Laboratory, 4930 N 31st St,
Forest Park, GA 30297-5202, USA

T. Ernst
Michigan State Police-Grand Rapids Laboratory,
720 Fuller Ave NE,
Grand Rapids, MI 49503, USA

C. Hanlon
Miami Dade Police Department,
Forensic Science Services Bureau, 91105 NW 25th St,
Miami, FL 33172, USA

A. Heydon · D. Rudell
Center of Forensic Sciences, 25 Grosvenor St,
Toronto, ON, Canada M7A268

R. Nelson
Tennessee Bureau of Investigation, Forensic Services Division,
901 R.S Gass Blvd,
Nashville, TN 37216, USA

K. Olsson
Johnson County Sheriff's Office Crime Laboratory, 11890 S.,
Sunset Olathe, KS 66061, USA

C. Palenik
Microtrace LLC, 790 Fletcher Drive,
Elgin, IL 60123, USA

E. C. Pollock
Sacramento County District Attorney's Office,
Laboratory of Forensic Services, 4800 Broadway,
Sacramento, CA 95820, USA

M. Valadez
Texas Department of Public Safety Crime Laboratory,
5800 Guadalupe,
Austin, TX 78752, USA

J. Almirall (✉)
Department of Chemistry and Biochemistry,
Florida International University, OE116 11200 SW 8th St,
Miami, FL 33199, USA
e-mail: almirall@fiu.edu

analytical performance of the different methods, (c) evaluate the capabilities of the methods to correctly associate glass that originated from the same source and to correctly discriminate glass samples that do not share the same source, and (d) standardize the methods of analysis and interpretation of results. Reference materials NIST 612, NIST 1831, FGS 1, and FGS 2 were employed to cross-validate these sensitive techniques and to optimize and standardize the analytical protocols. The resulting figures of merit for the ICP-MS methods include repeatability better than 5 % RSD, reproducibility between laboratories better than 10 % RSD, bias better than 10 %, and limits of detection between 0.03 and 9 μg g^{-1} for the majority of the elements monitored. The figures of merit for the μ-XRF methods include repeatability better than 11 % RSD, reproducibility between laboratories after normalization of the data better than 16 % RSD, and limits of detection between 5.8 and 7,400 μg g^{-1}. The results from this study also compare the analytical performance of different forensic science laboratories conducting elemental analysis of glass evidence fragments using the three analytical methods.

Keywords Elemental analysis · Forensic · Glass · X-ray fluorescence · ICP-MS · Laser ablation ICP-MS

Introduction

The comparison of glass fragments recovered from crime scenes to glass sources of known origin has long been recognized as a key examination of physical evidence. The significance of any associations made as a result of these comparisons is improved when more discriminating analytical methods are used [1]. The comparison of elemental composition between glass samples has proven to enhance the value of an association when one is found, and to reduce false associations between different sources that may result when less discriminating methods, such as refractive index [2–19], are used. As the number of forensic science laboratories performing elemental comparisons of glass fragments has increased, the need for consistency among laboratories concerning both analytical methodology and interpretive criteria has been recognized [20]. To address these issues, an Elemental Analysis Working Group (EAWG) consisting of forensic glass examiners and research scientists from North America and Europe was formed under the direction of researchers at Florida International University with funding from the US National Institute of Justice. The goal of the EAWG was to develop analytical protocols and to assess the utility of glass source comparisons by way of several interlaboratory studies. This paper describes the evaluation and validation of the analytical methods for the elemental analysis of glass evidence fragments.

Glass represents a model matrix for trace evidence examiners for several reasons: (a) due to its fragile nature and wide use in society, it is one of the most common types of trace evidence found in case scenarios such as hit-and-run accidents, burglaries, kidnappings, homicides, and shootings; (b) it is easily transferred from the broken source to the scene, victims, and others in the vicinity; (c) it is easily recovered from a scene or object; (d) it can persist after transfer; (e) its chemical composition does not vary over time; (f) the typical recovered fragment size is normally sufficient for analysis by a variety of analytical methods; (g) there are sensitive methods and suitable reference standards routinely used in forensic laboratories to detect chemical and physical properties; (h) the physical properties and elemental composition of glass fragments are relatively homogeneous within a single pane or sheet of glass; (i) despite the standardization of manufacturing processes, detectable variations in the physical/optical properties and chemical composition permit the differentiation of glass samples from different manufacturing sources and from a single source over time; (j) when sensitive methods are used, excellent source discrimination can be achieved on the basis of the optical characteristics and elemental composition; and (k) the framework proposed to construct opinions derived for glass comparisons can also be used by other types of trace evidence [12]. For these reasons, glass was selected as a model material by the EAWG to work towards the standardization of analytical methods and the interpretation of evidence.

A number of analytical methods have been used to measure the elemental composition of glass for forensic purposes. These include multi-elemental determinations either by quantitative or qualitative methods. Currently, the methods most frequently used in forensic science laboratories are scanning electron microscopy-x-ray spectroscopy (SEM-EDX), x-ray fluorescence spectroscopy (XRF) and inductively coupled plasma (ICP)-based methods with either mass spectrometry (MS) or optical emission spectroscopy (OES) as a detection method. Effective sample introduction for ICP-MS and ICP-OES methods has been accomplished using either digestion of glass fragments followed by nebulization of the resulting solution or by laser ablation (LA) of the solid glass material.

SEM-EDX is used both for the classification of the type of glass (soda-lime, borosilicate, alumino-silicate, lead-alkali-silicate, etc.) of recovered fragments and for the comparison of recovered glass fragments with potential sources [7]. This technique is nondestructive and allows the characterization of very small glass fragments such as glass debris on projectiles or pulverized and imbedded in tools and weapons. However, SEM-EDX has limited sensitivity

and therefore can only be used to detect the presence of minor and major elements at concentrations greater than 0.1 % [20–23]. In addition, the precision is generally poorer than other methods such as XRF and ICP-based methods [23]. For these reasons, the interlaboratory exercises reported in this paper do not include SEM-EDX data but instead focused only on the more sensitive and discriminating methods.

In order to accommodate the small size of recovered glass fragments, x-ray fluorescence spectroscopy instruments with either highly collimated or capillary-focused x-ray beams are typically used for analysis. Collectively, these instruments are referred to as micro-XRF instruments (μ-XRF). Emitted x-rays are detected with an energy dispersive detector in μ-XRF instruments. The advantages of μ-XRF are similar to those of SEM-EDX: it is nondestructive, relatively easy to operate, and provides simultaneous multi-elemental information. However, μ-XRF is more sensitive than SEM-EDX especially for elements of energy higher than 3 keV providing better discrimination between glasses of the same type [2–4, 23, 24]. Advantages of μ-XRF over ICP-based methods are that it has a lower instrument cost and easier operation and maintenance; it does not require a pre-determined elemental menu prior to the analysis; it can be used at any point in the analytical scheme due to its totally non-destructive nature; and, although data acquisition is more time-consuming, most instruments can operate unattended [19, 21, 25].

The main drawback to μ-XRF is that the analysis of very small and irregularly shaped samples can produce inaccurate quantitative results and less precise replicate measurements than ICP-methods, both within a given fragment and between fragments from the same source [12]. Also, μ-XRF is not sensitive enough to measure several trace elements that have been shown to have good source discrimination capability [12]. Accurate quantitation typically requires matrix-matched standards and use of a method such as embedding and polishing of the sample in order to present a flat surface to the x-ray beam [24]. As a result, most forensic laboratories compare x-ray data taken from glass fragments by spectral overlay and/or semi-quantitative comparison of the ratios of the intensities of the x-ray emission peaks. However, the best comparisons can only be made between samples having relatively flat surfaces and similar shape morphologies [4–6, 19, 24].

Several methods based upon inductively coupled argon plasmas (ICP) are gaining in popularity for the analysis of glass samples in forensic science laboratories. ICPs are well-controlled, high discharges that are used to excite and ionize elements that make up samples introduced into the plasma. Detection is made either by optical emission in ICP-OES instruments or mass spectrometry in ICP-MS instruments [12]. ICP methods benefit from features such as nearly simultaneous multi-elemental capability, reduced matrix interference effects, wide linear dynamic ranges, and excellent precision and sensitivity. These attributes result in superior discrimination power compared to other methods of glass analysis [5, 9, 26–32].

Initially, protocols using ICP-OES or ICP-MS for glass fragment analysis required dissolving the glass in a hydrofluoric acid-based mixture followed by evaporation to dryness to remove excess HF, and then reconstitution of the dissolved material in an acid matrix [30]. The resulting digest is aspirated into the plasma for analysis [5, 10, 17, 30]. The major drawbacks to these protocols are that they are rather time-consuming, require the use of hazardous reagents, and allow for potential introduction of contaminants into the solution. ICP-MS instruments are normally one to two orders of magnitude more sensitive than ICP-OES, therefore allowing for the use of smaller glass fragments. A typical digestion of glass for ICP-OES analysis consumes 5 to 8 mg per replicate, whereas ICP-MS requires only about 1 to 2 mg per replicate measurement [12].

To avoid the problems associated with dissolution, direct analysis of a solid glass sample can be accomplished by LA with introduction of the resulting aerosol directly into the ICP torch. Laser ablation can be coupled to either ICP-OES or ICP-MS instruments to simplify the analysis, significantly reducing not only the time and complexity of sample preparation but also the amount of sample consumption (<0.3 to 2 μg per replicate) [28, 31–40]. The main drawbacks to any ICP-based techniques are more expensive instrumentation, more challenging to operate, and currently available in only a few forensic science laboratories.

Although the aforementioned techniques are routinely used in forensic science laboratories worldwide, there is still a need for improved standardization of the methods within the forensic community. A preliminary effort towards this goal was reported by Becker et al. [8], where the discrimination potential of different techniques such as SEM-EDX, μ-XRF, and ICP-MS was described. However, the work did not include comparisons to laser-based methods. The European Working Group (NITECRIME), using LA-ICP-MS only, conducted an analogous study on glass standards in the period 2001–2005 [36]. To the best of our knowledge, this is the first time that all three of these sensitive methods are directly compared to each other, not only based on their analytical performance but also based on their discrimination potential for glass evidence.

In this paper, important considerations in analytical method validation for μ-XRF and ICP-based methods will be discussed that may be used as guidance by scientists for the standardization of methods of analysis and for providing a

better understanding of the capabilities of these techniques, including reporting figures of merit, match criteria, and their informing power. This will be especially useful in the context of quality management, accreditation, and interpretation of the significance of evidence, which have become matters of increasing relevance in trace evidence examination in recent years [20].

Experimental

Instrumentation and measurement parameters

Several different instruments were used within the interlaboratory studies. The ICP and XRF instruments and analytical parameters used in this study are summarized in Tables 1 and 2, respectively.

The following element list was used by the LA-ICP-MS participants: ^{7}Li, ^{25}Mg, ^{27}Al, ^{29}Si (as internal standard), ^{39}K, ^{42}Ca, ^{49}Ti, ^{55}Mn, ^{57}Fe, ^{85}Rb, ^{88}Sr, ^{90}Zr, ^{118}Sn, ^{137}Ba, ^{139}La, ^{140}Ce, ^{146}Nd, ^{178}Hf, and ^{206,207,208}Pb (reported as total Pb). The participant that conducted acid digestion followed by ICP-MS used the same menu with the exception of ^{29}Si, ^{7}Li, ^{139}La, ^{140}Ce, ^{146}Nd, ^{178}Hf, and ^{206,207,208}Pb. The digestion and ICP-MS method followed the ASTM method E2330 [41].

Due to the nature of the technique, the XRF participants did not have a pre-determined element list but were asked to report data for any detected elements with atomic number greater than ten, including at least Na, Mg, Al, Si, K, Ca, Ti, Fe, Sr, and Zr. Participants were asked to report peak area intensity data for the following ratios: Ca/Mg, Ca/Ti, Ca/Fe, Sr/Zr, Fe/Zr, and Ca/K.

Reagents, standards, and samples

The standard reference materials NIST SRM 612, NIST SRM 1831 (National Institute of Standards and Technology, Gaithersburg, MD), and the matrix-matched float glass standard (FGS) glasses FGS 1 and FGS 2 (Bundeskriminalamt, Wiesbaden, Germany) were provided to each participant for the interlaboratory studies. The glass DGG 1 (Deutsche Glastechnische Gesellschaft, Offenbach, Germany) was also used as a control check in an extended study. In addition, glass samples were submitted as mock casework comparisons. Those samples were selected from a set of different sources collected and analyzed at Florida International University between 1998 and 2010.

Analytical protocols and descriptions of interlaboratory tests

Thirteen participants reported the first interlaboratory test results. One participant performed acid digestion followed

Table 1 Instrumental parameters used for the elemental analysis of glass fragments by either LA-ICP-MS or ICP-MS

Lab ID	B-ICP	C-ICP	D-ICP	F-ICP	G-ICP	H-ICP	I-ICP[a]
Method	LA-ICP-MS	LA-ICP-MS	LA-ICP-MS	Digestion-ICP-MS	LA-ICP-MS	LA-ICP-MS	LA-ICP-MS
ICPMS instrument	Perkin Elmer Elan DRC II	Thermo X series II	Perkin Elmer Elan DRC II	Agilent Technologies 7500	Agilent Technologies 3500	Agilent Technologies 7500	Perkin Elmer-Elan DRC II
RF power (W)	1,500	1,450	1,500	1,600	1,350	1,390	1,300
Ar gas flow (Lmin^{-1})	1.0	1.0	0.90	0.50	0.9	0.9	0.95
Laser ablation instrument	New Wave UP213	New Wave UP213	New Wave UP213	n.a.	New Wave UP213	New Wave UP213	CETAC LSX500
Carrier gas (Lmin^{-1})	He, 1.0	He, 1.0	He, 0.90	n.a.	He, 0.70	He, 0.93	He, 0.95
Wavelength (nm)	213	213	213	n.a.	213	213	266
Spot size (μm)	60	55	55	n.a.	60	55	50
Fluence (J cm^{-2})	17.5	19	30	n.a.	22	12	nr
Tube length (m)	0.91	1.8	1.2	n.a.	1.2	1.2	0.75
Ablation cell volume (cm^3)	80	25	25	n.a.	25	25	50

[a] Data reported only for interlaboratory test # 2

Table 2 Instrumental parameters used for the elemental analysis of glass fragments by μ-XRF

Lab ID	A-XRF	B-XRF	C-XRF	D-XRF	E-XRF	F-XRF	G-XRF	H-XRF[a]	I-XRF[a]
Instrument model	EDAX Eagle III	EDAX Eagle II	EDAX Eagle III	IXRF coupled to Phillips XL30	EDAX Eagle III	EDAX Eagle III	EDAX Eagle II	EDAX Orbis	EDAX Eagle III
Dead time (%)	35	40	35	35–40	35–40	18–22	Avg.	40	25
Beam size (μm)	114	300	100	100	210	100	50	30	100
Beam energy (kV)	40	40	50	45	45	50	38	50	50
Time constant (μs)	17	17	35	2	17	35	17	12.8	35
Acquisition time (live seconds)	1,200–1,500	1,200	1,200	1,200	1,200	1,200–1,800	1,200	1,200	1,200
Resolution (eV)	138.4	161.9	145.5	146.0	147.2	148.3	164.3	177.5	148.0
Collimation of capillary beam	Poly-capillary	Mono-capillary	Mono-capillary	Poly-capillary	Poly-capillary	Mono-capillary	Poly-capillary	Poly-capillary	Mono-capillary
Tube material	Rh	Rh	Rh	Mo	Rh	Rh	Rh	Rh	Rh

[a] Data reported only for interlaboratory test # 2

by ICP-MS, five participants conducted the analysis using LA-ICP-MS, and seven participants used μ-XRF. Sixteen participants reported the second test results. One participant performed acid digestion followed by ICP-MS, six participants conducted the analysis using LA-ICP-MS, and nine used μ-XRF.

Each interlaboratory test contained the instructions for analysis and reporting according to the analytical method. The protocol of analysis was standardized for each analytical method as much as possible to facilitate interlaboratory comparisons. However, each laboratory was allowed some latitude in setting instrumental parameters according to their own optimized method.

First interlaboratory test

The first glass interlaboratory test was designed to conduct analyses on glass standard materials NIST 612 and NIST 1831 and also to conduct analyses on glass fragments that simulate glass transfer evidence in order to answer the question "Does the glass from the known sample (K1) and the questioned sample (Q1) share the same elemental composition?"

Items were packaged individually in weighing paper and placed in pillboxes properly identified with labels. Glass samples that were packaged and labeled as item 1 (K1) and item 2 (Q1) originated from the same source. The fragments were obtained from a windshield glass from the FIU glass collection. The windshield was manufactured by PPG industries, Pittsburgh USA in August 2002 and displays the logo: TOYOTA. Participants in the study were not informed as to the source of the samples or that they originated from the same source in this blind study.

Pieces of ~2–3 cm^2 were collected from an area of about 30 cm^2 of the inside panel of the windshield. The glass samples were then washed with methanol, nitric acid (0.8 M), and DI water. Once the samples were dry, they were broken into small fragments. Sample size was selected to be representative of typical fragments received in casework. About ~3–5 fragments of 3 to 7 mm length were placed in pillboxes and labeled as K1. About seven to ten small fragments of 1 to 5 mm in length were placed in pillboxes and labeled as Q1. One pair of pillboxes along with the test instructions was provided to each participant, for each analytical method used.

Second interlaboratory test

The second glass interlaboratory test was designed to conduct elemental analyses on glass standard materials NIST 1831, FGS 1, and FGS 2 to study both the intralaboratory and interlaboratory variation in the measurements. Glass fragments of NIST 1831 were submitted as full thickness

fragments (ranging from 5 to 12 mm in length) and small fragments (ranging from 1 to 3 mm in length) to evaluate the effects of fragment size and shape.

An expanded study was conducted to evaluate the homogeneity of the elemental composition of glass standard SRM 1831 at bulk and surface fragments by LA-ICP-MS. A sample fragment taken from SRM NIST 1831 was broken into four full-thickness fragments that were then used for the full thickness measurements (surface and bulk). The full-thickness fragments were analyzed in different orientations (surface 1 up focused to the laser beam, surface 2 up focused to the laser beam and bulk material tilted (cross section) focused to the laser beam). Four small fragments were also sampled from the bulk area. All fragments were analyzed in six replicates. Reference standard materials SRM NIST 612 and/or FGS 2 were used as calibrators. The glass DGG 1 was used for quality control verification.

In addition, a set of glass fragments was submitted for comparison in order to permit further evaluation of different match criteria and to address the interpretation. Items were packaged individually in weighing paper and then in envelopes properly identified with labels. Glass samples that were packaged and labeled as item 1 (K1), item 2 (Q1), and item 3 (Q2) were architectural float glass manufactured at the same manufacturing plant (Cardinal Glass Industries, Portage, WI, USA). Glass samples labeled K1 and Q1 shared a common origin. They were sampled from a 4×4-cm glass fragment collected from a glass pane sampled at the Cardinal manufacturing plant on April 1, 2001. Glass samples labeled Q2 originated from a different glass sheet of glass from those labeled sample K1; however, they were compositionally similar. Although they were manufactured at the same manufacturing plant, the glass Q2 was manufactured 2 years and 8 months before glasses K1 an Q1 (August 12, 1998).

A total of three fragments, all of them full thickness ranging from 2 to 7 mm across, were submitted as known samples (K1). Three fragments were submitted for each of the questioned samples; at least two of them were full thickness fragments ranging from 1 to 4 mm. The glass samples were washed with methanol, nitric acid (0.8 M), and deionized water and examined microscopically to assure full thickness and/or original surfaces were present when required. Once the samples were dry, they were carefully broken and measured with a caliper to group them by size and make sure all participants had series of fragments of similar size and shape. Each sample was prepared in a separate clean area to avoid cross contamination.

The participants were not told of the sources of the samples for this blind test. The only information provided to them was that the results of preliminary tests (color, microscopic examination, and refractive index) showed no significant differences between K1 and items Q1 and Q2.

Data analysis

Five ICP-participant laboratories processed their TRA signal from laser ablation with GLITTER™ software (GEMOC, Macquarie University, Sydney, Australia), which allows reduction of transient signal to quantitative data. One of the participants used Plasmalab (Thermo Fisher XSeriesII, Bremen, Germany) and Microsoft Excel (Microsoft Corp, WA, USA), and one used in-house software for the data reduction. The XRF data were processed using manufacturer's software (EDAX, NJ, USA) for spectral overlay and Microsoft Excel (Microsoft Corp, WA, USA).

Statistical analyses were performed by either the use of SYSTAT for windows (v.8.0, SPSS Science, IL, USA), JMP (v.5.0.1 SAS, NC, USA), Excel 2003 (v9.0.2719, Microsoft Corp., WA, USA), Plot for mac OSX (v.0.997, Berlin, Germany), or Mathematica (v. 5.2.0.0, IL, USA).

Results and discussion

The interlaboratory tests were intended to assist participating forensic laboratories in improving elemental analysis of glass comparisons by cross-validating their methods and evaluating their analytical protocols. The main objective of these studies was to conduct elemental analysis of glass with different analytical techniques in order to provide standardized methods and a basis for discussion of the utility of elemental analysis comparison methods, the effectiveness of different methods of statistical analysis, and the interpretation of results.

Both studies consisted of two main tasks: (a) analysis of reference standard materials to evaluate the analytical performance within and among methods and (b) analysis of glass fragments submitted as "blind" tests to evaluate the capabilities of the techniques to correctly associate glass that originated from the same source and/or discriminate glasses that originated from different sources.

The glass standard reference materials NIST 612, NIST 1831 and the glass standards FGS 1 and FGS 2 were used to evaluate the accuracy and precision of individual laboratory measurements. Glass fragments were submitted with a simulated casework scenario to assist the selection of match criteria and the reporting of comparison results between questioned and known fragments.

Evaluation of the analytical performance

The results for the elemental analysis of glass standards were separated into two sub-groups based on the techniques used by the participants: (1) the "ICP Group" consisted of six to seven laboratories that performed elemental analysis by ICP-MS or LA-ICP-MS and (2) the "XRF Group"

consisted of seven to nine laboratories that conducted elemental analysis by μ-XRF.

Due to the nature of the techniques used for the analysis of the standards and samples, the ICP Group reported quantitative data, whereas the XRF Group reported semi-quantitative data; therefore, different statistical methods were used to evaluate the results for each group.

Analytical performance of ICP-MS methods

The bias and precision obtained by each laboratory were compared to the interlaboratory results as well as to the certified or reference values for the glass standards. All LA-ICP-MS laboratories were asked to use the standard SRM NIST 612 as a single calibrator for the analysis of verification control standards and samples. Concentration values for SRM NIST 612 were used as reported by Pearce et al. [42]. The participant that conducted acid digestion followed the dissolution and calibration methods described in ASTM E2330 [41].

The glass reference materials NIST 1831, FGS 1, and FGS 2 were used to monitor the analytical performance of the methods and the assessment of the fitness for purpose of the method. These reference materials were selected due to the similarity of their compositions to the typical soda-lime glass found in forensic casework. The interlaboratory test results for precision and bias obtained for the three reference

standard materials are shown in Tables 3, 4, and 5. Each of the ICP laboratories made seven replicate sample measurements each on SRM NIST 1831, FGS 1, and FGS 2. The precision measures included repeatability standard deviation and reproducibility standard deviation. Both repeatability and reproducibility were calculated as specified in ASTM Practice E 177 [43] and international standards [44, 45].

The majority of the 18 isotopes monitored showed study bias and interlaboratory reproducibility better than 10 %, demonstrating that ICP-MS methods (solution and laser ablation-based) can provide accurate and precise quantitative information that can be used for forensic comparison of glass samples. Moreover, the results showed absence of systematic errors when ICP-based methods were employed for the measurement of bias for the range of relevant reference glass standards used in this study. The determination of bias was also important to demonstrate traceability of the measurements.

Although accuracy is important in the decision to include data in glass databases or data collections, for purposes of typical forensic comparisons between known and questioned fragments, precision is more critical. As shown in Tables 3, 4, and 5, repeatability within replicates measured by a single laboratory is typically better than 5 %. Reproducibility better than 10 % was achieved between participants in different laboratories that used different instruments, operating parameters, and operators, demonstrating consistency of results.

Table 3 Bias and precision obtained by ICP-methods for SRM NIST 1831 from the second interlaboratory study

Element	Reported value, μg g^{-1}	Study average, μg g^{-1a}	Study bias %	Repeatability-within s$_r$ (%)	Reproducibility-between s$_R$ (%)
Li	5.00 [b]	5.3	6	5.1	5.6
Mg	21,200 [c]	23,200	9.4	1.1	10.8
Al	6,380 [c]	6,400	0.3	1.1	8.7
K	2,740 [c]	2,680	−2.2	2.3	6.7
Ca	58,600 [c]	58,000	−1.0	2.6	3.6
Fe	608 [c]	540	−11	2.7	24.9
Ti	114 [c]	130	14	2.6	6.5
Mn	15.0 [d]	13.3	−11	1.8	2.5
Rb	6.11 [d]	6.0	−1.8	2.4	3.5
Sr	89.1 [d]	86	−3.5	2.0	5.6
Zr	43.4 [d]	37	−15	2.2	6.8
Ba	31.5 [d]	30	−4.8	2.6	7.9
La	2.12 [b]	2.2	3.8	2.6	6.7
Ce	4.54 [d]	4.4	−3.1	2.6	3.8
Nd	1.69 [b]	1.8	6.5	2.3	7.1
Hf	1.10 [d]	1.0	−9.1	3.7	8.5
Pb	1.99 [d]	1.8	−9.5	5.0	4

[a] Average value obtained from seven participant laboratories using different manufacturer LA and ICP-MS instruments

[b] Historical data from a single lab over a 1-year period (n=42 days)

[c] Certified by NIST

[d] Reported in reference [41], values obtained by acid digestion ICP-MS interlaboratory test

Table 4 Bias and precision obtained by ICP methods in FGS 1 from the second interlaboratory study

Element	Reported value [a], µg g^{-1}	Study average, µg g^{-1} [b]	Study bias %	Repeatability-within s_r (%)	Reproducibility-between s_R (%)
Li	6.0	5.9	−1.8	4.5	8.2
Mg	23,900	26,100	9.2	1.6	9.0
Al	1,500	1,560	4.0	2.2	3.5
K	920	1,000	8.7	4.1	4.7
Ca	60,600	59,200	−2.3	1.0	5.5
Fe	580	530	−8.6	1.6	26
Ti	69	80	16	4.5	8.8
Mn	43	45	3.5	0.8	2.8
Rb	8.6	7.8	−8.8	2.8	3.1
Sr	57	56	−1.6	2.3	6.7
Zr	49	46	−5.9	2.4	8.9
Sn	19	20	4.7	1.9	1.9
Ba	40	41	2.3	2.8	8.0
La	4.3	4.2	−2.1	4.2	6.7
Ce	5.2	5.0	−4.6	0.9	7.8
Nd	5.1	5.0	−2.4	3.8	7.7
Hf	3.20	3.0	−6.6	2.4	10
Pb	5.8	5.2	−10	1.6	2.9

[a] Consensus values reported in reference [36] (n=5–10)

[b] Average value obtained from seven participant laboratories using different manufacturer LA and ICP-MS instruments

The ICP-MS and LA-ICP-MS methods have been previously developed and validated prior to these interlaboratory studies by various members of this working group, and the possible sources of uncertainty have been previously identified and reported elsewhere [8, 9, 13, 18, 30–37, 40]. In addition, Table 3, 4, and 5 show that the bias observed in

Table 5 Bias and precision obtained by ICP methods in FGS 2 from the second interlaboratory study

Element	Reported value [a], µg g^{-1}	Study average, µg g^{-1} [b]	Study bias %	Repeatability-within s_r (%)	Reproducibility-between s_R (%)
Li	29	26	−10.3	1.7	4
Mg	23,400	25,600	9.4	1.1	10
Al	7,400	7,600	2.7	1.1	7.2
K	4,600	4,900	6.5	0.8	6.5
Ca	59,300	59,000	−0.5	1.3	6.5
Fe	2,600	2,600	0.8	1.5	15.3
Ti	326	370	13.5	1.3	10.6
Mn	221	222	0.5	2.3	2.1
Rb	35	38	7.7	1.5	3.8
Sr	253	256	1.2	0.7	6.1
Zr	223	221	−0.9	1.8	9.7
Sn	94	97	3.4	1.4	2.3
Ba	199	198	−0.5	1.3	8.3
La	18	19	5.0	1.5	8.2
Ce	23	24	3.5	3.5	6.4
Nd	25	25	1.6	4.1	8.6
Hf	15	14	−5.3	2.3	7
Pb	24	24	1.7	1.8	4.4

[a] Consensus values reported in reference [36]

[b] Average value obtained from seven participant laboratories using different manufacturer LA and ICPMS instruments

this study was not significantly different than the reproducibility of the measurements between the participant laboratories. As a result, the combined standard uncertainty can be reported as the reproducibility standard deviation as recommended by EURACHEM/CITAC GUIDE CG4 [46].

An exception for reproducible results was observed for iron. Even though good repeatability was achieved by individual laboratories for replicate measurements, poor interlaboratory reproducibility was observed between participants. The inferior performance for iron, in terms of bias and reproducibility, was not surprising because standard quadrupole ICP-MS instruments suffer from polyatomic interferences including oxides and hydroxides such as $^{40}Ar^{16}O^1H^+$, $^{40}Ca^{16}O^1H^+$, $^{41}K^{16}O$, $^{40}Ar^{16}O^+$, $^{40}Ca^{16}O^{1+}$ that compromise the analytical determination of $^{56}Fe^+$ and $^{57}Fe^+$. Due to the nature and abundance of these interferences, standard unit resolution ICP-MS instruments cannot measure the most abundant iron isotope $^{56}Fe^+$ (91.72 % abundant); therefore, limits of detection for the lower abundant isotope $^{57}Fe^+$ (2.2 % abundant) are typically high (>10 $\mu g\ g^{-1}$) [47]. Moreover, the concentration of iron in the standard SRM NIST 612 used as calibrator for LA-ICP-MS is close to the limit of quantitation for some of the instrument configurations, introducing a source of error and inconsistency.

In addition to the interlaboratory measures of precision and bias reported, each laboratory was later provided with detailed information of (a) the individual mean values and standard deviations reported by each laboratory for each element, (b) certified values, (c) acceptance study range, (d) interlaboratory variation of the measurements, and (e) z scores. This information allowed an effective way for each participant to evaluate their own protocol, cross-validate their methods, and detect outliers or systematic bias, if any.

The z score corresponds to how far the reported value from each laboratory was from the study mean, divided by the standard deviation of the study [48]. The acceptance range for the purposes of this interlaboratory study was defined as the study mean±three times the study standard deviation [48].

Strontium results for FGS 1 are shown in Fig. 1 as an example of the interlaboratory statistics. In general, all laboratories had excellent accuracy and precision for most elements. All laboratories were within the control criteria for the interlaboratory comparison (reported as z score), with few exceptions for few elements. One participant laboratory presented a systematic bias for Zr (for the three reference standard materials), which led to improvement of their method of analysis.

One of the participants experienced inconsistencies of the results of the concentrations of Ce and La for the glass reference FGS 1, which led to an interesting finding for the forensic laser ablation community. It was made clear by the participant that these values derived from measurements that

Laboratory ID	Mean value	Standard deviation	z-scores	Inter-Lab statistics	
B-ICP	57.8	1.4	0.46	Study Mean	56.1
C-ICP	53.7	0.5	-0.63	Study Standard Deviation	3.8
D-ICP	54.3	1.2	-0.47	high limit (mean + 3sd)	67.5
F-ICP	62.1	0.04	1.59	low limit (mean - 3sd)	44.6
G-ICP	57.3	1.6	0.32		
H-ICP	57.2	0.8	0.30		
I-ICP	50.0	5.0	-1.61		
Reference value [a]	57.0	4			
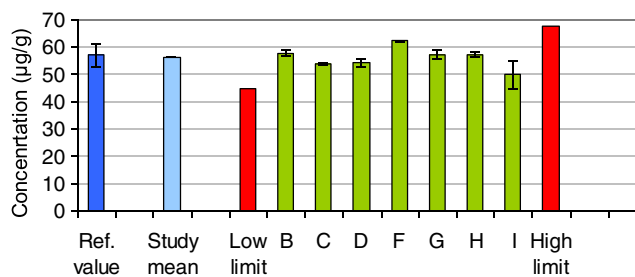
[a] Consensus value reported in reference [36]

Fig. 1 Example of interlaboratory statistics of ICP-MS participants for strontium in FGS 1

were taken from a fragment that had originated from the frosted rim of the FGS 1 glass disk. The TRA signal of these ablations exhibits a large peak in the beginning, followed by tailing, suggesting surface contamination.

Triggered by these observations, several experiments were carried out by the issuer of the FGS glasses (BKA/Germany). All eight FGS 1 and FGS 2 glasses that were examined exhibited a pre-peak-like signal for Ce and to a smaller extent also for La, combined with spiking of the TRA signal. Based on communication with SCHOTT AG/Germany, the producer of the glass, this is most certainly caused by a partial removal of cerium oxide that was used during the polishing stages of the FGS 1 and FGS 2 disks.

Moreover, several sets of analyses have been carried out by BKA, ablating on the polished surface very close to the rim of FGS 1 and FGS 2. When ablating on the rim or very close to the rim (up to 250 μm) in several cases spikes can be detected for Ce and La, inspecting the TRA signal. These spikes led to incorrect high concentrations for cerium and lanthanum. After removal of these peaks using the time-resolved analysis software GLITTER™, the concentrations for Ce and La were correct.

It can be concluded that measurements/ablations on the rim and very close to the rim of the FGS standards (FGS 1 and FGS 2) should be avoided. The interlaboratory exercises showed that the analytical methods used by ICP participants are fairly standardized and provide consistent results between laboratories regardless of the instrument configuration. The analytical performance of the method proved to be fit for purpose.

Analytical performance of μ-XRF methods

The μ-XRF group reported results based on semi-quantitative analysis (i.e., intensities or ratios of intensities for the analytes). Although some calibration strategies can be used to conduct quantitative analysis of glass by μ-XRF, this is not typically performed in forensic laboratories as part of their glass examinations. Quantitative accuracy and precision are dependent on algorithm ZAF corrections that can vary significantly for uneven surfaces and varying sample thicknesses. Instead, comparisons of spectra and/or of ratios of intensities, the latter intended to mitigate the effects of varying take-off angles, are common practice among forensic examiners.

All the individual laboratories were asked to report intensities for a pre-determined list of elements. A large variation in the analytical signal was observed amongst participant laboratories due to differences of instrument configurations and acquisition parameters, making the evaluation of the interlaboratory performance particularly challenging.

Although these interlaboratory differences do not affect the interpretation of the individual comparison results, a direct comparison between labs was unattainable at this stage. For this reason, a normalization of the data was conducted versus the standard reference material 1831 measured by each participant as a way to attempt to standardize the responses from different laboratories. In order to conduct the normalization for each laboratory, measurements of the glass samples and the SRM 1831 were conducted on the same day. The mean intensity of an element measured on the glass standard was divided by the mean intensity of the same element measured on the SRM 1831:

$$E\text{normalized} = \frac{\left[\frac{1}{n}\sum_{i=1}^{n} Ei\right]\text{sample}}{\left[\frac{1}{n}\sum_{i=1}^{n} Ei\right]\text{SRM1831}} \quad (1)$$

where E is the peak area intensity of the analyte of interest and n is the number or replicate measurements.

This approach relies on the premise that if a certain instrument configuration produces a lower intensity for a specific element, the response will be lower for both the sample and the 1831 reference standard SRM, and vice versa. Therefore, by using the ratios, these relative interlaboratory differences can be minimized.

Figure 2 illustrates this effect, where significant differences between laboratories were observed, before normalization, in the response of calcium and magnesium on FGS 1. After normalization with SRM NIST 1831, the responses between participants were comparable. Standard deviations of the ratios were estimated as a propagation of random errors for multiplicative expressions as reported elsewhere [46, 48].

This approach allowed a comparison of the response between laboratories for the following ratios on standards FGS 1
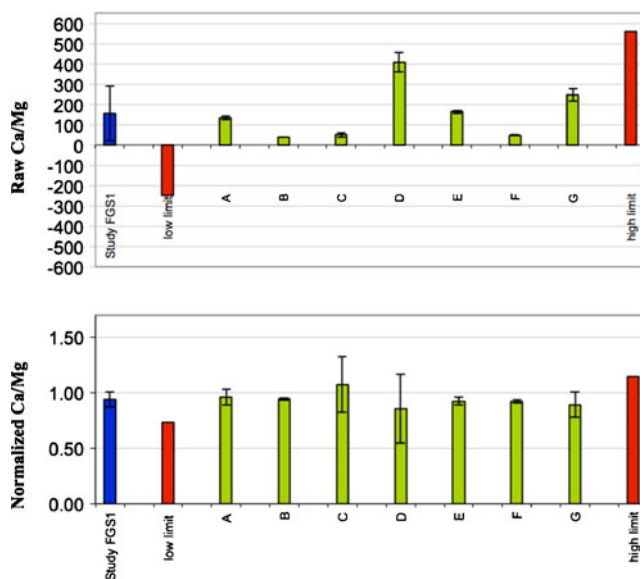

Fig. 2 Interlaboratory comparison of the Ca/Mg ratio measured by μ-XRF, for FGS 1 without normalization (*top*) and after normalization (*bottom*) with SRM NIST 1831

and FGS 2: Ca/Mg, Ca/Ti, Ca/Fe, Sr/Zr, Fe/Zr, and Ca/K. The semi-quantitative normalized data expressed as ratio of the peak area intensities were used to estimate z score values and to detect systematic errors within laboratories. Table 6 illustrates that data obtained by different participants were very consistent after normalization, with variation between laboratories within the acceptance criteria (absolute z score value equal to or less than 3). The normalization not only facilitated interlaboratory comparisons but also opened an opportunity to share XRF databases in the future.

Table 6 Values of z score obtained from the interlaboratory comparison of elemental ratios by μ-XRF for FGS 1 and FGS 2

Lab ID	Ca/Mg	Ca/K	Ca/Fe	Sr/Zr	Fe/Zr	Ca/K
z scores for FGS 1						
A-XRF	0.36	0.04	0.98	1.20	0.94	0.09
B-XRF	0.39	0.26	0.16	0.72	1.23	1.61
C-XRF	1.41	1.89	0.81	1.05	0.19	0.42
D-XRF	1.64	0.91	2.09	na	na	1.51
E-XRF	0.24	0.32	0.05	1.18	0.49	0.83
F-XRF	0.16	0.16	0.25	0.31	0.15	0.08
G-XRF	0.95	0.18	0.18	0.63	0.86	0.51
z scores for FGS 2						
A-XRF	0.33	0.88	1.17	0.17	0.04	0.46
B-XRF	0.05	0.66	0.38	0.77	1.34	1.09
C-XRF	1.97	1.67	0.98	1.84	1.51	1.10
D-XRF	1.21	0.36	1.33	na	na	1.75
E-XRF	0.23	0.66	0.45	0.45	0.06	0.62
F-XRF	0.23	1.00	0.21	0.07	0.43	0.37
G-XRF	0.68	0.64	1.42	0.87	0.26	0.02

The efficiency of the normalization approach is also reflected in Table 7 where the reproducibility is presented for the FGS standards. With the exception of Fe/Zr, reproducibility among laboratories was better than 12 %. The poorer precision of Fe/Zr could be a result of the x-ray energies for Fe and Zr that are widely divergent and much more prone to take-off angle variations.

Comparison of figures of merit of μ-XRF and ICP-based methods

Figures of merit such as repeatability, reproducibility, bias, and limits of detection were evaluated in these interlaboratory tests. Precision and bias figures obtained by ICP and μ-XRF methods were suitable for purposes of glass comparisons in the forensic context.

The precision in terms of repeatability and reproducibility is reported for ICP-MS (Tables 3, 4, and 5) and μ-XRF methods (Table 7). Although good precision is observed by all the studied methods, better repeatability between replicate measurements is attainable by the ICP-based methods.

Reproducibility and repeatability in the measurements by μ-XRF methods are more affected than ICP-MS measurements by changes in the instrument configurations, acquisition parameters, limits of detection, and sample fragment size and orientation. The concentrations of some elements in the standards analyzed in this study were close to the limits of detection

Table 7 Precision data obtained by μ-XRF methods for FGS 1 and FGS 2

Element ratio	Average[a]	Repeatability-within s_r (%)[b]	Reproducibility-between s_R (%)[c]
FGS 1 precision			
Ca/Mg	0.89	5	11
Ca/Ti	1.44	9	8
Ca/Fe	1.07	1	3
Sr/Zr	0.60	9	12
Fe/Zr	0.81	11	16
Ca/K	2.61	3	7
FGS 2 precision			
Ca/Mg	0.93	4	9
Ca/Ti	0.36	3	6
Ca/Fe	0.23	1	3
Sr/Zr	0.55	2	8
Fe/Zr	0.86	5	15
Ca/K	0.55	2	4

[a] Average value obtained from 9 different μ-XRF instrument configurations, 7 replicates per configuration for a total of 63 replicates

[b] Average %RSD value obtained from 9 different μ-XRF instrument configurations, 7 replicates per configuration for a total of 63 replicates

[c] Variation estimated as %RSD from mean values for nine different μ-XRF instrument configurations. Mean values were estimated from ratios normalized to SRM NIST 1831

(LOD) and/or quantitation limits for some XRF systems, which affected the overall precision. However, most monitored elements in μ-XRF are typically observed at higher concentrations than present in the standard reference materials and, therefore, better precision (<10 %) was observed on the K/Q comparisons.

The LOD has been used consistently in the area of analytical chemistry as an objective way of evaluating and reporting the performance of the methods. For this reason, the LODs were reported for ICP and μ-XRF data as a means to monitor and compare the methods and techniques used in these interlaboratory tests. The evaluation of the LODs played an important role in the optimization and standardization of the methods, helping participants to (1) evaluate the performance of their instrumentation and optimize their parameters to achieve expected threshold values, (2) make informed decisions about the selection of elements for the comparison of glass samples, and (3) validate the methodology through interlaboratory comparison of the sensitivity for a suite of relevant elements.

Table 8 shows the expected LODs of the different methods. The LOD reported here is the concentration at which the analyte signal is three times the system noise. The LODs were determined for several elements in NIST SRM 1831, FGS 1, and FGS 2 [49, 50].

The background count level in μ-XRF is affected by the sample and uses counting statistics; therefore, to estimate the signal-to-noise ratio, the noise in a μ-XRF spectrum is calculated as the square root of the background counts under the peak of interest. Limits of detection were estimated as the concentration of each analyte corresponding to three times the noise. More detail in data treatment was recently reported by Ernst et al. [51].

The limits of detection of the method for LA-ICP-MS data were determined for each element by measuring procedure blanks. Blanks corresponded to the background signal prior to the laser interaction with the glass. The LODs were calculated by three times the standard deviation of 21 instrumental replicates from the standards NIST 1831, FGS 1, and FGS 2.

ICP-based methods showed superior limits of detection than μ-XRF (one to three orders of magnitude) allowing the analysis of greater number of trace elements. As expected, the LODs for μ-XRF improved with increasing atomic number as a consequence of the increase in critical escape depth and excitation efficiency of the generated x-rays from these elements in thicker samples [24].

Regardless of the differences in sensitivity, most elements monitored by each method are above the typical concentration range observed in soda-lime glass (Table 8). Therefore, it is anticipated that all methods will provide information about the elemental composition that is sensitive to variations in the composition of glass manufactured in different plants or at the same plant at different time intervals. In order to evaluate whether or not the differences in figures of merit among techniques affect the

Table 8 Expected limits of detection (LOD) for glass analysis by ICP-MS, LA-ICP-MS, and μ-XRF methods, respectively

Method/element	LA-ICP-MS [a]	Digestion-ICP-MS[b]	μ-XRF [c]	Range of sample concentrations (μg g^{-1})
Li	0.75	n.a	n.a	0.8–7 [d]
Na	n.a	n.a	7,400	n.a
Mg	0.52	7.5	1,300	6,273–51,076 [e]
Al	1.85	5.0	890	298–11,940 [e]
K	2.22	n.a	100	45–6,328 [e]
Ca	145	n.a	49	46,086–69,767 [e]
Ti	3.15	0.18	20	39–3,226 [e]
Fe	9.21	n.a	11	461–6,063 [d]
Mn	0.77	0.17	14	9–468 [e]
Rb	0.19	0.04	6.4	0.3–33 [e]
Sr	0.07	0.06	7.8	19–576 [e]
Zr	0.13	0.91	5.8	19–269 [e]
Sn	0.52	n.a.	n.a	11–2,180 [d]
Ba	0.30	0.04	n.a	3–384 [e]
La	0.05	0.02	n.a	1–19 [e]
Ce	0.03	0.05	n.a	2–1,896 [e]
Nd	0.17	n.a	n.a	0.8–8 [d]
Hf	0.09	0.15	n.a	0.5–7 [e]
Pb	0.16	0.05	n.a	0.3–251 [e]

[a] Average limits of detection for measurements of glass standards FGS 1, FGS 2, and SRM NIST 1831, values expressed as micrograms per gram of the elemental concentration in the solid glass

[b] Values reported for a set of 50 soda-lime glass samples all expressed as nanograms per gram of the elemental concentration in the final solution [41]

[c] Average limits of detection for instrument configurations A-I, for data collected for glass standards FGS 1, FGS 2, and SRM NIST 1831 expressed as micrograms per gram of the elemental concentration in the solid glass

[d] From actual measurement of a set of 127 soda-lime glass samples from vehicle and architectural windows by LA-ICP-MS

[e] From actual measurement of a set of 286 soda-lime glass samples from vehicle and architectural windows by digestion-ICP-MS

discrimination capabilities, a set of glass samples were analyzed in both interlaboratory studies as described below.

Evaluation of association and/or discrimination capabilities of the methods

Another aim of these studies was to evaluate and compare the discrimination capabilities of the different techniques and methods in traditional glass samples. Blind test samples were submitted to each participant along with a simulated casework scenario and preliminary analysis results (color, microscopic examination, and refractive index) to assist their selection of match criteria and reporting.

Results from the first interlaboratory test

As detailed in the experimental section, samples submitted as known and questioned items (K1 and Q1) originated from the same source, so it was expected that respondents associate those fragments based on their elemental composition and their selected match criteria. The glass from K1 and Q1 was analyzed prior its distribution and found to be indistinguishable by refractive index and elemental analysis. Pre-distribution elemental analysis conducted by LA-ICP-MS revealed no significant differences, using the *t* test at 95 % confidence, in the content of the following elements: Al, K, Ti, Mn, Fe, Rb, Sr, Zr, Ba, La, Ce, Nd, Hf, and Pb.

All 13 respondents correctly reported that item 1 (K1) was found to be indistinguishable from item 2 (Q1) based on LA-ICP-MS or μ-XRF. Each participant was asked to use the match criteria commonly used in their casework. Although there was agreement in the reporting of results, a lack of standardization in the match criteria was observed for this first interlaboratory test. The participants reported a variety of match criteria, including *t* test, ±2 s, ±3 s, ±4 s, modified ±4 s, range overlap, and spectral overlay.

Results from the second interlaboratory test

Glass samples that were submitted as item 1 (K1), item 2 (Q1), and item 3 (Q2) were architectural float glass manufactured at the same manufacturing plant. Glass samples sent as K1 and Q1 shared a common origin; they were sampled from a glass pane manufactured in 2001. Glass samples sent as Q2

originated from a different source than sample K1. Although they were manufactured at the same manufacturing plant, the glass Q2 was manufactured 2 years and 8 months before.

The glass samples were analyzed prior to their distribution and found to be indistinguishable by RI. These particular glass sources were selected specifically because they had similar refractive indices but different elemental composition of some of their trace elements. Concentration of the trace discriminating elements in these glass sources ranged from 0.5 to 125 μg g^{-1}, with exception of iron that was present at ~600 μg g^{-1}. Major elements such as Al, K, Mg, and Ca were present at concentrations above 1 %.

All the participating laboratories correctly reported that item 1 (K1) was indistinguishable from item 2 (Q1), and all the labs correctly reported that item 1 (K1) was distinguishable from item 3 (Q2). For this second trial, there was a consensus amongst the μ-XRF participants towards using spectral overlay and ± 3 s as match criteria. The ICP participants still reported a large variety of match criteria for this test.

In this test, the basis for discrimination (differences) between the elemental compositions of glasses manufactured at different times depends on the LODs of the methods. Significant differences were found by ICP-MS on a large number of elements (7 to 15 out of the 16 to 18 elements analyzed were found to be distinguishable based on their selected match criteria). The XRF participants detected differences primarily on major elements (K, Ca) and trace elements that were present in these samples above 70 μg g^{-1} (Ti, Mn, and Fe).

The results of these two studies demonstrate that each of the evaluated methods (ICP-MS, LA-ICP-MS, and μ-XRF) can be successfully applied to determine the elemental composition of glass fragments as a tool to improve discrimination capabilities of preliminary screening tests, such as RI. Despite the use of a variety of analytical methods and match criteria, all laboratories were able to correctly associate samples that originated from a single source and discriminate between glasses manufactured in the same plant at different periods of time.

The lack of standardization of the match criteria used by the participants motivated the design of additional interlaboratory exercises that would permit a thorough evaluation of the effect of match criteria on the incidence of type I and type II errors. Those results will be presented in a separate publication.

Comparison of composition data from SRM 1831 full thickness versus small fragments

The effects of size of glass fragments on the analytical measurements by LA-ICP-MS and its performance in forensic comparisons were also studied. Data reported in the literature have shown that fragment size and shape do not affect the performance of the quantitative data on glass fragments by LA-ICP-MS. These studies have been reported on standard reference materials NIST 612, NIST 610, and several flat glass samples but, to the best of our knowledge, have not been reported on SRM NIST 1831 [33, 34] .

In this interlaboratory exercise, quantitative data obtained from fragments of SRM NIST 1831 having different thicknesses and sizes showed good precision and accuracy (repeatability <1–5 %, bias <10 %). Nevertheless, significant differences were detected between full thickness and small fragments using the most common match criteria reported by the participants (ANOVA ($p = 0.05$), t test comparison ($p = 0.05$), and ± 3SD; see Table 9).

Significant differences were also found between small and full thickness data collected by μ-XRF. These differences were expected due to the well-known effects of the take-off angle and critical depth on XRF measurements [4]. For this reason, the study was then focused on ICP-MS data only.

For the purposes of forensic glass comparisons, if the two fragments being compared are significantly different by at least one element (or ratio), these can be excluded as having come from the same source. In this exercise, full thickness fragments were used for the known source, and the small fragments were used for questioned samples. The results presented here indicate that the application of multiple t tests for multivariate datasets obtained by LA-ICP-MS measurements might be problematic (Table 9). The possible reasons for these type I errors (false exclusions) might be day-to-day variations of measurement conditions, sample orientation or position in the ablation cell, sample heterogeneity, and small variations between replicate measurements.

In an effort to identify the sources of type I errors in this set, an additional experiment was conducted to evaluate whether the differences in elemental composition were due to: (a) fragment size, (b) surface versus bulk heterogeneity, and/or (c) match criteria used for comparisons.

Analyses were conducted on full thickness fragments at both original surfaces (S1, S2), at the bulk area of full thickness fragments (B1 and B2), and on four small fragments taken from the bulk of a SRM NIST 1831 fragment. Six replicate measurements were acquired from each fragment. Pairwise comparisons by ANOVA ($p = 0.05$) show significant differences between the small fragments, bulk areas, and surface areas.

A recent study published by the Bundeskriminalamt/Federal Criminal Police Office, Forensic Science Institute [52] reported that wider match criteria are recommended for LA-ICP-MS measurements of glass due to the excellent precision between replicates. The authors conducted an extensive study on the elemental variability of 34 glass fragments that originated from the same glass sheets and found that tight match criteria, such as the t test, produced high rates of false exclusions. The best results for glass casework were achieved using a broader match

Table 9 Detail of elements with differences in elemental composition for full thickness vs small fragment of SRM NIST 1831 measured by LA-ICP-MS

Lab ID	Elements distinguished by t test ($p=0.05$)	Elements distinguished by t test with Bonferroni correction ($p=0.05$)	Elements distinguished by ±3 s	Elements distinguished by ±4 s	Elements distinguished by ±4 s with 3 % min RSD	Elements distinguished by ±4 s with 4 % min RSD	Elements distinguished by ±4 s with 5 % min RSD
B	Mg, Al, Sr, Zr, Sn, Nd, Hf, K, Ca, Ti, Mn, La, Pb	Mg, Al, Sr, Zr, Hf, Ca, Ti, Pb	Mg, Al, Sr, Zr, Sn, Hf	Hf, Zr, Sn	Hf, Zr, Sn	Sn	Sn
C	Mg, K, Fe, Li, Al, Ti, Mn, Rb, Sr, Zr, Sn, La, Nd, Hf, Pb	Mg, K, Fe, Mn,	Mg, K, Fe, Zr, Sn, Nd	K, Fe, Zr, Sn, Nd	Fe, Zr, Sn	Fe, Sn	none
D	Mg. Sr, Zr, Al, K, Ca, Ti, Fe, Rb, Ba, La	Mg, Sr, Zr, Ti	Al, Mg, Sr, Zr	Al, Mg, Sr, Zr	Sr, Zr	Sr, Zr	none
G	Sr, Zr, Sn, Nd, Hf, Li, Mg, Al, K, Ca, Ti, Mn, Fe, Ba, La, Ce, Pb	Sr, Zr, Sn, Nd, Hf, Al, Ti, Mn, Ba, La, Ce, Pb	Ti, Sr, Zr, Sn, Nd, Hf	Ti, Sr, Zr, Sn, Nd, Hf	Sr, Zr, Sn, Pb, Hf	Sr, Zr, Sn, Hf	Sr, Zr, Sn,
H	Mg, Al, Zr, K, Ca, Mn, Fe, Sr, Nd, Hf	Mg, Al, Zr, Mn, Fe, Hf	Mg, Al, Zr, Hf	Mg, Al, Zr, Hf	Zr, Hf	none	none
I	Mg. Sr, Zr, Al, Ca, Ti, Fe, Rb, Ba, La, Sn	none	Mg, Al, Sr, Zr, Pb	Mg, Al, Sr, Zr,	Sr, Zr, Sn	Sr, Zr, Sn	Sn

criterion, such as a modified ±4 s approach, based on fixed relative standard deviations.

Due to the close precision obtained and reported by most of the ICP-based participants (≤1–5 % RSD), it was observed that some match criteria, such as the t test, may be too sensitive to false exclusions, depending on the data set under evaluation. For this reason, a modified ±4 s criterion was applied to these two sets of samples. Table 9 shows that, for most participants, the number of elements distinguished is reduced by using a 4 s criterion with a minimum of 3–5 % RSD. Further discussion of this recommendation will be included in a separate publication.

Some ICP laboratories still detected differences on the tin content, even after applying wider match criteria. Although SRM NIST 1831 was not produced by the float glass process, ICP methods detected a slightly different composition on the original surfaces versus the cross section of the glass. Original surfaces were only present on the full thickness fragments. Nevertheless, in casework, tin is typically monitored to detect the float versus the non-float side of a glass and is not typically included as part of the elements used for comparison between samples.

The results in Table 10 demonstrate that the differences detected between the SRM NIST 1831 fragments submitted for the interlaboratory tests were due to a combination of the heterogeneity between surface and bulk composition on SRM NIST 1831 and the selection of match criteria used for comparisons.

First, the use of wider match criteria, such as ±4 s with minimum 3 % RSD, reduced the number of false exclusions. Using ANOVA, 18 out of 28 possible comparison pairs were excluded (64 %); using ±4 s criterion, the number of exclusions was reduced to 13 out of 28 possible comparison pairs (46 %), whereas using the wider match criteria, the number of

exclusions was limited to 7 out of 28 possible pairs (25 %). Second, when using wider criteria (i.e., ±4 s criteria with a minimum of 3 % RSD) significant differences are still detected between one of the original surfaces (S2) and the rest of the fragments, while no significant differences are detected between the rest of the fragments regardless of their size.

The results revealed that one of the original surfaces of the SRM NIST 1831 is depleted in Sr, Zr, Hf, and Pb which causes a significant heterogeneity for microsampling techniques like LA-ICP-MS. Although this study implies that fragment size does not affect comparison of the elemental composition of glass by LA-ICP-MS, caution should be taken when using full thickness fragments to avoid possible differences in the composition of original flat surfaces. The effects of expanding the match criteria on type I and type II errors were further studied by the working group with the aim to provide standardized recommendations and will be reported in a separate publication.

Conclusions

This study allowed for a direct comparison between three of the most sensitive methods currently available for the forensic elemental analysis of glass samples (LA-ICP-MS, solution ICP-MS, and μ-XRF). The methods were compared in terms of analytical performance and discrimination capability.

ICP-based methods (ICP-MS and LA-ICP-MS) are the most sensitive methods, with limits of detection on the order of sub-parts per million in the solid material. Advantages of these methods are that they are fairly standardized among participant laboratories, they are currently used in forensic

Table 10 Pairwise comparison of SRM NIST 1831 glass fragments using ANOVA ($p=0.05$), 4 s interval, and 4 s with minimum 3 % RSD, respectively

	B1	B2	F5	F6	F7	F8	S1	S2
ANOVA $p=0.05$								
B1								
B2	Fe							
F5		Fe						
F6	Li	Li, Fe						
F7	Li, Fe		Fe					
F8			Fe					
S1		Na, Fe		Li, Na	Na, Fe	Ti, Fe		
S2	Fe, Sr, Pb, Zr	Sr, Pb, Zr	Fe, Sr, Hf, Pb, Zr	Fe, Sr, La, Hf, Pb, Zr	Sr, Hf, Pb, Zr	Sr, Hf, Pb, Zr	Na, Fe, Sr, Pb, Zr	
±4 s								
B1								
B2								
F5		Fe						
F6		Li						
F7								
F8								
S1		Fe		Li, Na	Li, Fe, Zr	Li, Na		
S2	Sr, Pb, Zr, Fe	Sr, Pb, Zr	Fe, Sr, Pb, Zr	Fe, Sr, Pb, Zr	Sr, Pb, Zr	Sr, Zr	Fe, Sr, Pb, Zr	
±4 s (min 3 % RSD)								
B1								
B2								
F5								
F6								
F7								
F8								
S1								
S2	Sr, Pb, Zr, Hf	Sr, Pb, Zr, Hf	Sr, Pb, Zr, Hf	Sr, Pb, Zr. Hf	Sr, Pb, Zr, Hf	Sr, Pb, Zr, Hf	Sr, Pb, Zr	

Elements listed were significantly different using the specified match criteria

laboratories, and they have been accepted in court. A standardized ASTM method already exists for the digestion and analysis by ICP-MS (ASTM E2330) [41], and the EAWG developed a standardized method for LA-ICP-MS which was submitted to ASTM [53]. Both methods are fairly mature with several publications previously reporting the evaluation of their capabilities and limitations. In addition, laser ablation sampling has unique advantages over digestion-based methods, such as reducing the sample consumption from milligrams to just few a hundred nanograms, reducing the time for analysis, and eliminating the use of hazardous digestion reagents. Interlaboratory comparisons of glass reference standard materials demonstrated that ICP methods provide accurate and precise quantitative data with deviations lower than 10 % for nearly all elements measured in the studies.

Important findings from LA-ICP-MS methods include: (a) the detection and report of heterogeneity of Ce and La close to the rim on FGS standards (<250 μm) and (b) the awareness

that possible differences between surface and bulk composition in compared glasses may lead to false exclusions if sampling and data interpretation are not carefully evaluated.

XRF methods provided consistent data among participants after normalization with a reference standard material such as SRM NIST 1831. The EAWG also used the experience gained from these interlaboratory tests to work towards the standardization of a μ-XRF method for the elemental analysis of glass, which was submitted to ASTM [54]. Limits of detection are two to three orders of magnitude higher than ICP-based methods; therefore, the number of trace elements typically detected in glass samples is more limited. Nevertheless, good performance was also observed among XRF laboratories. The measurement of LODs provided a better understanding of the capabilities of the technique and permitted a means of quantitatively comparing the performance of different instrument configurations. Relevant observations derived from the studies include: (a) the use of normalized data to a glass standard such as SRM NIST 1831

provides a means to account for differences among instrumental configurations and to conduct interlaboratory comparisons, (b) the use of a glass standard as a "control" glass is recommended to check method performance prior to analysis, and (c) the use of K and Q fragments with similar size and shape is necessary to improve precision and thus increase discrimination power.

Mock case samples allowed an inter-method comparison of the capabilities to associate samples that originated from the same source and to discriminate among samples that were manufactured in the same plant line at different time periods. Excellent agreement between laboratories was achieved in both blind tests with 100 % correct conclusions. The interlaboratory tests also provided an excellent opportunity for participants to fine-tune their methods and protocols and cross-validate their methodology. The study revealed that a wide variety of match criteria are currently employed by forensic laboratories to conduct statistical comparisons of elemental composition data. Extensive discussions between the group members led to the design of additional interlaboratory tests to address the interpretation of evidence and the systematic selection of match criteria for elemental comparisons of glasses, based on simultaneously minimizing the frequency of both false exclusions and false inclusions. Results of these studies will be presented in a separate publication.

References

1. Almirall JR, Buckleton J, Curran JM, Hicks TN (2000) In: Curran JM, Champod TNH, Buckleton JS (eds) Forensic interpretation of glass evidence examination of glass. CRC Press, Boca Raton, FL
2. Reeve V, Mathiesen J, Fong W (1976) Elemental analysis by energy dispersive X-ray: a significant factor in the forensic analysis of glass. J Forensic Sci 21:291–306
3. Dudley RJ, Howden CR, Taylor TJ, Smalldon KW (1980) The discrimination and classification of small fragments of window and non-window glasses using energy-dispersive X-ray fluorescence spectrometry. X-Ray Spectrom 9:119–122
4. Howden CR, Dudley RJ, Smalldon KW (1978) The analysis of small glass fragments using energy dispersive x-ray fluorescence spectrometry. J Forensic Sci Soc 18:99–112
5. Koons RD, Peters CA, Rebbert PS (1991) Comparison of refractive index, energy dispersive X-ray fluorescence and inductively coupled plasma atomic emission spectrometry for forensic characterization of sheet glass fragments. J Anal At Spectrom 6:451–456
6. Hicks T, Monard Sermier F, Goldmann T, Brunelle A, Champod C, Margot P (2003) The classification and discrimination of glass fragments using nondestructive energy dispersive X-ray μ fluorescence. Forensic Sci Int 137:107–118
7. Andrasko J, Machly A (1978) The discrimination between samples of window glass by combining physical and chemical techniques. J Forensic Sci 22:250–262
8. Becker S, Gunaratman L, Hicks T, Stoecklein W, Warman G (2001) The differentiation of float glass using refractive index and elemental analysis: comparisons of techniques. Problems of Foren Sci 47:80–92
9. Duckworth DC, Morton SJ, Bayne K, Montero S, Koons RD, Almirall JR (2002) Forensic glass analysis by ICP-MS: a multi-element assessment of discriminating power via analysis of variance and pairwise comparisons. J Anal At Spectrom 17:662–668
10. Koons RD, Fiedler C, Rawalt C (1988) Classification and discrimination of sheet and container glasses by inductively coupled plasma atomic emission spectrometry and pattern recognition. J Forensic Sci 33:49–67
11. Montero S, Hobbs A, French T, Almirall JR (2003) Elemental analysis of glass fragments by ICP-MS as evidence of association: analysis of a case. J Forensic Sci 48:1101–1107
12. Almirall JR, Trejos T (2006) Advances in forensic analysis of glass fragments with a focus on refractive index and elemental analysis. Forensic Sci Rev 18(2):74–96
13. Suzuki Y, Sugita R, Suzuki S, Marumo Y (2000) Forensic discrimination of bottle glass by refractive index measurement and analysis of trace elements with ICP-MS. Anal Sci 16:1195–1198
14. Hughes JC, Catterick T, Southear G (1976) The quantitative analysis of glass by atomic absorption spectroscopy. J Forensic Sci 8:217–227
15. Coleman RF, Goode GC (1973) Comparison of glass fragments by neutron activation analysis. J Radiol Chem 15:367–372
16. Koons RD, Buscaglia J (2001) Distribution of refractive index values in sheet glasses, Forensic Science Commun 3,1 http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/jan2001/koons.htm. Accessed 3 April 2013
17. Catterick T, Hickman DA (1981) The quantitative analysis of glass by inductively coupled plasma-atomic-emission spectrometry: a five-element survey. Forensic Sci Int 17:253–263
18. Buscaglia J (1994) Elemental analysis of small glass fragments in forensic science. Anal Chim Acta 288:17–24
19. Ryland S (2011) Discrimination of flat (sheet) glass specimens having similar refractive indices using micro X-ray fluorescence spectrometry. J Am Soc Trace Evid Examiners 2(1):2–12
20. Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council (2009) Strengthening forensic science in the United States: a path forward. The National Academies Press, Washington DC
21. Brozel-Mucha Z, Zadora G (1998) Differentiating between various types of glass using SEM-EDX elemental analysis: a preliminary study. Problems of Foren Sci 37:68–89
22. Kuisma-Kursula P (2000) Accuracy, precision and detection limits of SEM-WDS, SEM-EDS and PIXE in the multi-elemental analysis of medieval glass. X-Ray Spectrom 29:111–118
23. Krüsemann H (2001) SEMs and forensic science. Probl Forensic Sci XLVII:110–121
24. Roedel TC, Bronk H, Haschke M (2003) Investigation of the influence of particle size on the quantitative analysis of glasses by energy-dispersive micro x-ray fluorescence spectrometry. X-Ray Spectrom 31:16–26
25. Ryland S (1986) Sheet or container? Forensic glass comparisons with an emphasis on source classification. J Forensic Sci 31:1314–1329

26. Naes B, Umpierrez S, Ryland SG, Barnett C, Almirall JR (2008) A comparison of laser ablation inductively coupled plasma mass spectrometry, micro X-ray fluorescence spectroscopy, and laser induced breakdown spectroscopy for the discrimination of automotive glass. Spectrochim Acta Part B 63(10):1145–1150

27. Wolnik KL, Gaston CM, Fricke FL (1989) Analysis of glass in product tampering investigations by ICP-AES with a hydrofluoric acid resistant torch. J Anal At Spectrom 4:27–31

28. Moenke-Blankenburg L, Schumann T, Günther D, Kuss HM, Paul M (1992) Quantitative analysis of glass using ICP-AES and ICP-MS, laser micro-analysis ICP-AES and laser ablation ICP-MS. J Anal At Spectrom 7:251–254

29. Zurhaar A, Mullings L (1990) Characterization of forensic glass samples using inductively coupled mass spectrometry. J Anal At Spectrom 5:611–617

30. Parouchais T, Warner IM, Palmer LT, Kobus H (1996) The analysis of small glass fragments using inductively coupled plasma mass spectrometry. J Forensic Sci 41:351–360

31. Duckworth DC, Bayne CK, Morton SJ, Almirall JR (2000) Analysis of variance of forensic glass analysis by ICP-MS: variance within the method. J Anal At Spectrom 15:821–828

32. Imbert JL, Telouck P (1993) Application of laser ablation ICP-MS to elemental analysis of glasses. Microchim Acta 110:151–160

33. Trejos T, Almirall JR (2005) Sampling strategies for the analysis of glass fragments by LA-ICPMS. Part I. Microhomogeneity study of glass and its application to the interpretation of forensic evidence. Talanta 67:388–395

34. Trejos T, Almirall JR (2005) Sampling strategies for the analysis of glass fragments by LA-ICPMS. Part II. Sample size and shape considerations. Talanta 67:396–401

35. Trejos T, Montero S, Almirall JR (2003) Analysis and comparison of glass fragments by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) and ICP-MS. Anal Bioanal Chem 376(8):1255–1264

36. Latkoczy C, Dücking M, Becker S et al (2005) Evaluation of a standard method for the quantitative elemental analysis of float glass samples by LA-ICP-MS. J Forensic Sci 50:1327–1341

37. Becker S, Watzke P, Dücking M, Stoecklein W (2003) Laser ablation in forensic glass analysis: the use of matrix matched standards for quantitative float glass analysis. Forensic Sci Int 136(1):361

38. Almirall JR, Trejos T, Hobbs A, Furton K (2003) In: Carapezza EM (ed), Elemental analysis of glass and paint samples of forensic interest by ICP-MS using laser ablation solid sample introduction, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Defense and Law Enforcement II. Proc Int Soc Opt Eng (SPIE) (SPIE) 5071:193–196

39. Bajic S, Aeschliman D, Staetveit NJ, Baldwin DP, Houk RS (2005) Analysis of glass fragments by laser ablation-inductively coupled plasma-mass spectrometry and principal component analysis. J Forensic Sci 50:1123–1127

40. Berends-Montero S, Wiarda W, de Joode P, van der Peijl G (2006) Forensic analysis of float glass using laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS): validation of a method. J Anal At Spectrom 21:1185–1193

41. American Society for Testing and Materials Method E2330-04 (2004) Standard test method for determination of trace elements in glass using inductively couple plasma mass spectrometry (ICP-MS). In: ASTM Annual Book of ASTM Standards, 14.02. American Society for Testing and Materials, West Conshohocken

42. Pearce NJ, Perkins WT, Westgate JA, Gorton MP, Jackson SE, Neal CR (1996) A compilation of new and published major and trace element data for NIST SRM 610 and NIST SRM 612 glass reference materials. Geostandard Newsletter 1:115–144

43. American Society for Testing and Materials ASTM Practice E177 (2004) Standard practice for the use of the terms precision and bias in ASTM methods. American Society for Testing and Materials, West Conshohocken

44. International Organization for Standardization ISO 3534–1 (2006) Statistics—vocabulary and symbols—part 1: general statistical terms and terms used in probability. ISO, Geneva, Switzerland

45. International Organization for Standardization ISO 5725–2 (1994) Accuracy (trueness and precision) of measurement methods and results, part 2: basic methods for the determination of repeatability and reproducibility of a standard measurement method. ISO, Geneva, Switzerland

46. EURACHEM/CITAC Guide CG4 (2000) Quantifying uncertainty in analytical measurements, Ellison SLR, Rosslein M, Williams A (eds), EURACHEM/CITAC, UK

47. Castro W, Trejos T, Naes B, Almirall JR (2008) Comparison of high resolution and dynamic reaction cell ICP-MS capabilities for the forensic analysis of iron in glass. Anal Bioanal Chem 392:663–672

48. Miller JN, Miller JC (2000) Statistics and chemometrics for analytical chemistry, 4th edn. Pearson Education Limited, Harlow, England

49. ACS Committee on Environmental Improvement (1980) Guidelines for data acquisition and data quality evaluation in environmental chemistry. Analytical Chemistry 52:2242–2249

50. International Union of Pure and Applied Chemistry (IUPAC) (1976) Commission on Spectrochemical and Other Optical Procedures for Analysis. Pure Appl Chem 45:99–103

51. Ernst T, Berman T, Buscaglia J et al. (2012) Signal to noise ratios in forensic glass analysis by micro x-ray fluorescence spectrometry. Journal of XRay Spectrom. doi: 10.1002/xrs.2437

52. Weis P, Dücking M, Watzke P, Menges S, Becker S (2011) Establishing a match criterion in forensic comparison analysis of float glass using laser ablation inductively coupled plasma mass spectrometry. J Anal At Spectrom 26:1273–1284

53. American Society for Testing and Materials ASTM (2012, WK39818) Standard test method for the determination of trace elements in soda-lime glass fragments using laser ablation inductively coupled plasma mass spectrometry for forensic comparisons

54. American Society for Testing and Materials ASTM (2012, WK39817) Standard test method for forensic comparison of glass using micro x-ray fluorescence (μ-XRF) spectrometry