JAAS

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

Cite this: *J. Anal. At. Spectrom.*, 2013, **28**, 1270

Forensic analysis of glass by μ-XRF, SN-ICP-MS, LA-ICP-MS and LA-ICP-OES: evaluation of the performance of different criteria for comparing elemental composition

Tatiana Trejos,[a] Robert Koons,[b] Peter Weis,[c] Stefan Becker,[c] Ted Berman,[d] Claude Dalpe,[e] Marc Duecking,[c] JoAnn Buscaglia,[f] Tiffany Eckert-Lumsdon,[g] Troy Ernst,[h] Christopher Hanlon,[i] Alex Heydon,[j] Kim Mooney,[g] Randall Nelson,[k] Kristine Olsson,[l] Emily Schenk,[a] Christopher Palenik,[m] Edward Chip Pollock,[n] David Rudell,[j] Scott Ryland,[d] Anamary Tarifa,[a] Melissa Valadez,[o] Andrew van Es,[p] Vincent Zdanowicz[q] and Jose Almirall*[a]

Four interlaboratory tests were designed to evaluate the performance of match criteria for forensic comparisons of elemental composition of glass by μ-XRF, solution nebulization SN-ICP-MS, LA-ICP-OES and LA-ICP-MS. A total of 24 analysts in 18 laboratories participated in the tests. Glass specimens were selected to study the capabilities of the techniques to discriminate glass produced in the same manufacturing plant at different time intervals and to associate samples that originated from a single source. The assessment of the effectiveness of several match criteria included: confidence interval (±6*s*, ±5*s*, ±4*s*, ±3*s*, ±2*s*), modified confidence interval, *t*-test, range overlap, and Hotelling's T^2. Error rates are reported for each of these criteria. Recommended match criteria were those found to produce the lowest combinations of type 1 and type 2 error rates. Performance of the studied match criteria was dependent on the homogeneity of the glass sources, the repeatability between analytical measurements, and the number of elements that were measured. The best results for μ-XRF data were obtained using spectral overlay followed by a ±3*s* confidence interval or range overlap. For ICP-based measurements, a wider match criterion, such as a modified confidence interval based on a fixed minimum relative standard deviation (±4*s*, >3–5% RSD), is recommended due to the inherent precision of those methods (typically <1–5% RSD) and the greater number of elements measured. Glass samples that were manufactured in different plants, or at the same plant weeks or months apart, were readily differentiated by elemental composition when analyzed by these sensitive methods.

Received 6th April 2013
Accepted 10th June 2013

DOI: 10.1039/c3ja50128k

www.rsc.org/jaas

Introduction

Elemental composition when combined with optical and physical properties has been shown to provide excellent discrimination between glass samples originating from different manufacturing sources.[1–10]

A number of analytical methods are used for the elemental composition of glass for forensic purposes, including scanning

[a]*Florida International University, International Forensic Research Institute, 11200 SW 8th St, Miami, FL 33199, USA. E-mail: almirall@fiu.edu; Fax: +1 305 3484485; Tel: +1 305 3483917*

[b]*Federal Bureau of Investigation, Laboratory Division, Counterterrorism and Forensic Science Research Unit, 2501 Investigation Parkway, Quantico, VA 22135, USA*

[c]*Bundeskriminalamt (BKA), Forensic Science Institute, 65173 Wiesbaden, Germany*

[d]*Florida Department of Law Enforcement, Orlando Regional Operations Center, 500 West Robinson Street, Orlando, Florida 32801, USA*

[e]*Royal Canadian Mounted Police Department, Forensic Science and Identification Services, 1200 Vanier Parkway, Ottawa, ON, Canada*

[f]*Federal Bureau of Investigation, Laboratory Division, Counterterrorism and Forensic Science Research Unit, 2501 Investigation Parkway, Quantico, VA 22135, USA*

[g]*US Army Criminal Investigation Laboratory, 4930 N 31st St, Forest Park, GA 30297-5202, USA*

[h]*Michigan State Police-Grand Rapids Laboratory, 720 Fuller Ave NE, Grand Rapids, MI 49503, USA*

[i]*Miami Dade Police Department, Forensic Science Services Bureau, 91105 Nw 25th St, Miami, FL, 33172, USA*

[j]*Center of Forensic Sciences, 25 Grosvenor St, Toronto, ON M7A268, Canada*

[k]*Tennessee Bureau of Investigation, Forensic Services Division, 901 R.S Gass blvd, Nashville, TN, 37216, USA*

[l]*Johnson County Crime Laboratory, 6000 Lamar Ave, Mission, KS, 66202, USA*

[m]*Microtrace LLC, 790 Fletcher Drive, Elgin, IL, 60123, USA*

[n]*Laboratory of Forensic Science, Sacramento, California, USA*

[o]*Texas Department of Public Safety Crime Laboratory, 5800 Guadalupe, Austin, TX, 78752, USA*

[p]*Netherlands Forensic Institute, Laan van Ypenburg, 2497 6B The Hague, The Netherlands*

[q]*Department of Homeland Security, CBP Research Laboratory, 7501 Boston Blvd, Springfield, VA 22153, USA*

This journal is © The Royal Society of Chemistry 2013

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

electron microscopy-X-ray spectroscopy (SEM-EDX),[7,13,14] X-ray fluorescence spectroscopy (XRF),[2,15] inductively coupled plasma optical emission spectroscopy (ICP-OES),[5,10,13–16] inductively coupled plasma mass spectrometry (ICP-MS)[18–21] and laser ablation methods coupled to ICP-MS or to optical emission spectrometers (LA-ICP-MS[17,18,22,23] and LIBS[23,24]).

With the increasing use of methods of elemental comparisons of glass fragments by forensic laboratories, the need for consistent analytical protocols and interpretive criteria has been recognized.[25] The Elemental Analysis Working Group (EAWG), made up of forensic glass examiners and research scientists, was formed to develop robust analytical protocols and to assess the accuracy of various criteria used for source comparison. The tests concerning development of analytical protocols have been addressed in a separate paper.[15] Analytical protocols using μ-XRF and ICP-based methods that provide quantitative data for 6 to 18 elements or intensity ratios were assessed. In order to make a decision as to whether the elemental compositions of two fragments are analytically indistinguishable, a match criterion must first be selected. In this study, if the known and the questioned glasses are found to be significantly different by the selected criterion in at least one of the monitored element concentrations or intensity ratios, then it can be concluded that the two samples did not originate from the same broken source. Conversely, a lack of significant differences in the elemental compositions suggests that the samples could have originated from the same broken source. This paper describes the results of interlaboratory tests that measured the error rates obtained when using a variety of selected match criteria on glasses having similar elemental compositions.

Several match criteria have been used to compare the elemental compositions of two or more glass fragments to determine if they could have originated from the same broken glass source.[12,27–34] Match criteria evaluated in this study are briefly described below.

Spectral overlay is typically one of the first steps used in comparisons of μ-XRF spectra. Reproducible differences in the elemental peaks present or marked differences in their relative peak heights indicate that the samples have different sources. The benefits of this technique are its simplicity and the ability to recognize the presence of unusual elements, such as Mo, which may be present at measurable concentrations. It is, however, somewhat subjective in that the ability to differentiate among visually similar spectra depends upon the similarity of background shapes, the signal to noise levels, and the experience of the person making the comparison. Moreover the glass specimens used for μ-XRF comparisons should be of similar size, shape, and thickness to each other to avoid false exclusions derived from take-off-angle and critical depth effects.[40] One disadvantage to spectral overlap is that its use does not provide a measure of the statistical significance when the result is that two samples are indistinguishable.

Range overlap is a relatively simple assessment that consists of determining whether or not the overall range of values of a measured parameter from one sample overlaps the range of values of that parameter in another sample. The range is calculated as the range between the highest and the lowest value in a data set. If the two measured ranges for all of the measured variables overlap, it may be concluded that the two samples could have originated from the same source. Advantages of range overlap are that there is no need for the assumption that the data are normally distributed and it is easy to explain to a jury. On the other hand, the statistical significance of any observed overlap may be difficult to determine since the distributions of the observed measurements may not be known, particularly when the number of measurements is small.[35]

The other criteria for comparison of elemental compositions that were evaluated in this study can conveniently be grouped into two general approaches. The first consists of tests of the hypothesis of equality of means of two sets of measurements. Examples of this approach are Student's *t*-test and its multivariate version, Hotelling's T^2 test. One advantage of the *t*-test comparison is that the statement regarding a match can be supported with a significance or probability value. Forensic practitioners have utilized various significance levels for the *t*-test. When using a *t*-test successively for each of a number of measured elements, a probability adjustment, such as the Bonferroni correction, can be applied to limit a cumulative increase in the family-wise error rate for multivariate comparisons.

Hotelling's T^2-test is a multivariate equivalent of Student's *t*-test.[35–37] It has the limitation that the number of measurements must be at least two greater than the number of variables. For example, if the concentrations of 18 elements are determined during the analysis, then Hotelling's T^2-test requires a total of at least 20 replicate measurements on the two samples. This is not always practical, depending on the analytical technique used and/or the sample size.

For comparisons of more than two samples or groups, Analysis of Variance (ANOVA) provides a statistical test of whether or not the means of several samples are all significantly different. ANOVA estimates differences between and within the samples, which is referred to as the variance.[35]

The second approach is designed to more closely reflect the common forensic situation where many measurements can be made on one sample (the known sample, K, typically a large fragment of a broken glass object from a known source) but only a few measurements can be made on the other sample (the questioned sample, Q, usually a small fragment recovered from a crime scene or suspect). The measurements from the broken object are used to characterize an acceptance interval for each element or intensity ratio using the mean and some measure of dispersion of measurements about that mean obtained from multiple measurements. The smaller number of measurements made from each recovered glass fragment are used to find mean values for that fragment which can then be compared to the acceptance interval. The glass samples are considered indistinguishable only when mean concentrations of each of the elements or element intensity ratios of the questioned sample fall within the matching known sample acceptance intervals. Conversely, if one or more of the measured values for the Q sample fall outside of their acceptance intervals, then K can be excluded as the source of that Q sample. When the data are normally distributed about the mean, a confidence interval and

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

a corresponding significance level can be assigned. Intervals of ±2, 3, 4, 5 and 6 standard deviations about the mean were examined in this study. Of course this approach does not apply to all situations. If the Q sample consists of a sufficient number and size of fragments that can reasonably be assumed to be from one source, it too can be characterized like the broken object and their acceptance intervals compared to see if they overlap.

For LA-ICP-MS measurements, the standard deviations of concentration values determined for a given piece of glass are sometimes so small that an unacceptably high rate of false exclusions can occur. To correct for this effect, it has been suggested that the acceptance interval should be set to the appropriate multiple times the greater of either the actual standard deviation or a value equal to 3% to 5% of the mean concentration for each element in the known sample.[33]

Regardless of which match criterion is used, there are two types of errors that can occur, either incorrect exclusion of samples from the same source (type 1) or incorrect association of samples from different sources (type 2). Two comprehensive studies have recently been reported for the evaluation of error rates for several match criteria for univariate refractive index data[38] and for multi-elemental concentration data.[33]

The aim of the research presented here is to determine directly the error rates associated with various match criteria using data obtained by several of the elemental techniques currently being used for the forensic analysis of glass. This study involved elemental analysis of glass fragments *via* four interlaboratory tests with each part of the study designed based on the previous interlaboratory test. The data for these studies was obtained by the participants using analytical instrumentation in current use in their laboratories. The development of standardized analytical protocols was described in a previous paper.[26] The blind testing of glass fragments under several case scenarios was conducted in the last three collaborative tests and these results are described in this paper. Topics discussed include blind test results for individual laboratories using their testing and interpretation criteria, evaluation of several match criteria by determination of the number of type 1 and type 2 errors that result when each criterion is applied to the data provided by the participants, and a homogeneity study of one of the glass sheets used in these studies. These interlaboratory tests were designed to determine which match criteria are appropriate for the interpretation of the data generated from the elemental analysis of glass and whether results obtained using these criteria are dependent on the technique used to acquire the data. This study represents an effort towards the standardization of the match criteria and sampling strategies used by glass examiners, and represents another step towards assessing the significance of glass associations.

Experimental

Instrumentation and measurement parameters

Several different instruments were used for the interlaboratory tests. The μ-XRF, SN-ICP-MS, LA-ICP-MS and LA-ICP-OES instruments and analytical parameters used in this study are reported in Tables 1 and 2. Typical intensity ratios reported by μ-XRF users included Ca/Mg, Ti/Ca, Ca/Fe, Sr/Zr, Ca/K, and several other minor elements, when detectable. Participants using ICP-based methods reported between 10 and 18 element concentrations from the following list: Li, Mg, Al, K, Ca, Fe, Ti, Mn, Rb, Sr, Zr, Sn, Ba, La, Ce, Nd, Hf, and Pb.

Standards and glass samples

The standard reference materials SRM NIST 612 and SRM NIST 1831 (National Institute of Standards and Technology, Gaithersburg, MD, USA) were provided to each participant for the interlaboratory tests. The SRM NIST 1831 standard was provided as full thickness fragments of 5 to 12 mm across to be used for evaluation of the analytical performance of each laboratory and also to normalize the μ-XRF data across participants. The matrix-matched float glass standards FGS 1 and FGS 2 (Bundeskriminalamt, Wiesbaden, Germany) were included in interlaboratory test 2.

Glass test samples were submitted as mock casework items. Those samples were selected from a collection of glass samples from different sources collected and analyzed by Florida International University researchers between 1998 and 2010. Samples with similar refractive indices but distinctive elemental compositions were selected for these exercises. The samples were carefully broken and grouped by size to insure that all participants had a series of fragments of similar size and shape. Each sample was prepared in a separate clean area to avoid potential cross contamination. Typically, one of the original flat surfaces of glass manufactured by the float process is enriched in tin as a result of the direct contact with tin baths.[1] This can also produce depletion of other elements near the float surface. For this reason, the participants were instructed to orient the glass fragments avoiding a float original glass surface to prevent false exclusions derived from sampling strategies rather than significant differences in the bulk composition of the glass.[1]

The project consisted of four interlaboratory tests conducted by the EAWG members. The first and second tests were designed to develop and characterize the μ-XRF and LA-ICP-MS analytical protocols. The results of these tests are detailed in a separate paper.[26] The second, third, and fourth tests contained sample fragments that each participant analyzed and compared to determine which ones could be distinguished as having come from different sources. In every case, participants made three measurements on each of three fragments provided for each sample. The participants were not told of the sources of the samples for these blind tests. Each participant was instructed to conduct elemental analysis in order to compare each of the questioned items with the know item to determine if any of the questioned items could have originated from the same broken source as the known standard. Previous analytical results obtained at FIU for the samples included in these tests are shown in Table 3. Due to space limitations, the data in Table 3 are shown without analytical uncertainties. The information is given to show the general differences between the samples and show which elements are most likely to be distinguishable

 This journal is © The Royal Society of Chemistry 2013

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

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

Lab ID	A-ICP[a]	B-ICP	C-ICP	D-ICP	E-ICP[a]	F-ICP	G-ICP	H-ICP	I-ICP[b]	J-ICP[c]	K-ICP[c]	L-ICP[c]
Method	LA-ICP-MS	LA-ICP-MS	LA-ICP-MS	LA-ICP-MS	LA-ICP-MS	Digestion-SN-ICP-MS	LA-ICP-MS	LA-ICP-MS	LA-ICP-MS	LA-ICP-OES	LA-ICP-MS	LA-ICP-MS
ICPMS instrument	Perkin Elmer Elan DRC II	Perkin Elmer Elan DRC II	Thermo X series II	Perkin Elmer Elan DRC II	Thermo Element 2	Agilent Technologies 7500	Agilent Technologies 3500	Agilent Technologies 7500	Perkin Elmer-Elan DRC II	Perkin Elmer Optima 7300DV	Perkin Elmer Elan DRC II	Thermo Element 2 XR
RF power (W)	1325	1500	1450	1500	1250	1600	1350	1390	1300	1500	1400	1260
Ar gas flow (L min^{-1})	1.0	1.0	1.0	0.90	1.0	0.50	0.9	0.9	0.95	0.7	0.9	0.8
Laser ablation instrument	New Wave UP213	New Wave UP213	New Wave UP213	New Wave UP213	New Wave UP213SS	n.a.	New Wave UP213	New Wave UP213	CETAC LSX500	New Wave UP213	CETAC LSX500	New Wave UP-193fx
Carrier gas (L min^{-1})	He, 0.75	He, 1.0	He, 1.0	He, 0.90	He, 0.7	n.a.	He, 0.70	He, 0.93	He, 0.95	He, 0.6	He, 0.92	He, 0.415
Wavelength (nm)	213	213	213	213	213	n.a.	213	213	266	213	266	193
Spot size (μm)	80	60	55	55	50	n.a.	60	55	50	100	50	50
Fluence (J cm^{-2})	8	17.5	19	30	9.5	n.a.	22	12	nr	25	3.5	6.1
Carrier gas tube length (m)	1.0	0.91	1.8	1.2	1.0	n.a.	1.2	1.2	0.75	1.2	1.0	1.1
Ablation cell volume (cm^3)	33	80	25	25	40	n.a.	25	25	50	25	33	50

[a] Data reported only for interlaboratory test # 3 and 4. [b] Data reported only for interlaboratory test # 2. [c] Data reported only for interlaboratory test # 4.

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

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)	1200–1500	1200	1200	1200	1200	1200–1800	1200	1200	1200
Resolution[b] (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	Rh	Rh	Rh	Rh	Rh	Rh

[a] Data reported only for interlaboratory test # 2. [b] Resolution measured as the full width at half the maximum height of the Mn Kα peak.

among samples by the various techniques. Specific information concerning the samples for each test is as follows.

Second interlaboratory test

The purpose of this test was to serve as a proficiency test during the development of the analytical protocols. The samples for this test consisted of fragments of architectural float glass manufactured by Cardinal Glass Industries (WI, USA) that were packaged and labeled as known (K1) and questioned samples (Q1 and Q2). Glass samples labeled K1 and Q1 were sampled from a 4 × 4 cm glass fragment collected from a glass pane sampled at the plant on April 1, 2001. Glass samples labeled Q2 originated from the same float line as those labeled sample K1, but they were manufactured 2 years and 8 months before glasses K1 and Q1. Three full-thickness fragments, from 2 to 7 mm across, were provided as sample K1. The Q1 and Q2 samples each consisted of three fragments, at least two of which were full thickness fragments from approximately 1 to 4 mm across.

For this test, participants were told to group the three fragments together for each of the samples. All comparisons were to be made using ten measurements on the K sample and nine measurements on each Q sample (three replicates on each of three fragments). Fifteen participants reported analytical results for this test. One laboratory performed acid digestion followed by SN-ICP-MS, seven participants conducted the analysis using LA-ICP-MS, and seven laboratories used μ-XRF.

Third interlaboratory test

The glass samples for this test were selected to study the capabilities of the techniques to discriminate glass produced in the same manufacturing plant at different time intervals (*i.e.* manufactured years apart, months apart and weeks apart). The dates of manufacture and the elemental profile of each of the samples as recorded in the FIU glass database are shown in Table 3. They were sampled from a 2 × 2.5 cm glass fragment originally sampled from a glass ribbon at the manufacturing plant. Samples labeled as K1 and K2 consisted of fragments that were 2 to 7 mm in size and those labeled as Q1, Q2, and Q3 consisted of fragments that were approximately 1 to 4 mm in size. Each sample contained three fragments. As in test 2, participants were again told to group the data from the three fragments together for each of the samples when making interpretations. Fourteen participants reported analytical results for this test. One participant performed acid digestion followed by SN-ICP-MS, six conducted the analysis using LA-ICP-MS, and seven used μ-XRF.

Fourth interlaboratory test

The set of glass samples for this test was selected primarily to study the capabilities of the techniques to associate glass that originated from the same source and also to discriminate glass produced in the same manufacturing plant at different time intervals. This set of glass fragments, consisted of 2 known samples and 3 questioned samples. Samples K1, K2, Q2 and Q3 originated from the same source. The glass fragments originated from two pieces, one 4.0 × 4.3 cm and the other 5.0 × 4.5 cm,

 This journal is © The Royal Society of Chemistry 2013

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

Table 3 Description of manufacturing dates of float glasses and the composition of the inter-laboratory test samples measured by acid digestion SN-ICP-MS[a] (tests 2 and 3) and by LA-ICP-MS (test 4) prior to the distribution of the tests. All values reported as mean concentration in μg g^{-1} ($n = 5$)

Test	Test 2	Test 2	Test 3	Test 3	Test 3	Test 3	Test 3	Test 4	Test 4
Sample	Q2	K1, Q1	K1	K2	Q1	Q2	Q3	Q1	K1, K2, Q2, Q3
Plant	Cardinal	Cardinal	Cardinal	Cardinal	Cardinal	Cardinal	Cardinal	Pilkington	Pilkington
Manufacture date	Aug/12/98	April/01/01	Aug/17/01	April/15/98	Aug/8/31/01	May/17/98	July/17/98	February/18/10	March/03/10
Li	—	—	—	—	—	—	—	6.79	6.14
Mg	22 100	22 000	11 500	21 700	13 300	19 900	21 000	29 300	30 500
Al	—	—	—	—	—	—	—	847	906
K	—	—	—	—	—	—	—	146	191
Ca	—	—	—	—	—	—	—	61 200	62 300
Ti	123	62.9	58.9	118	57.51	155	125	504	315
Mn	101	20.0	17.3	92.3	17.1	158	100	18.75	12.08
Fe	150	530	391	147	420	137	156	4300	3100
Rb	0.73	1.66	1.64	0.63	1.62	0.92	0.72	0.68	0.76
Sr	38.4	31.6	31.7	37.3	29.3	41.9	37.6	47.8	47.7
Zr	34.0	34.3	28.9	32.7	31.2	36.4	34.2	24.9	21.3
Sn	—	—	—	—	—	—	—	21.3	12.8
Ba	19.7	10.2	9.34	15.9	11.9	23.0	17.6	8.31	6.90
La	3.49	2.89	2.77	3.48	2.70	4.04	3.51	1.47	1.48
Ce	6.28	3.79	4.54	6.08	4.49	7.42	6.27	2.30	2.17
Sm	0.51	0.43	0.35	0.46	0.34	0.62	0.51	—	—
Nd	—	—	—	—	—	—	—	1.25	1.12
Hf	0.82	1.06	0.70	0.78	0.77	0.87	0.74	0.67	0.60
Pb	0.75	1.36	1.66	0.93	1.02	0.86	0.71	0.67	0.65

[a] ASTM method for acid digestion and SN-ICP-MS analysis (E330-04).[39]

which were once part of a single piece of glass. The glass was manufactured at a Pilkington plant (CA, USA) on March/03/10. The glass items labeled as Q1 originated from glass manufactured in the same plant approximately two weeks before the other samples (February/18/10). Known samples, K1 and K2 consisted of three full thickness fragments. Questioned samples Q1, Q2, and Q3 were each three small irregular fragments of approximately 0.5 to 1 mm in size, smaller than the fragments used in the prior tests.

For this test, each Q fragment was to be considered separately, rather than grouping the three fragments as in the previous tests. Seventeen participants submitted results for this test. One laboratory performed acid digestion followed by SN-ICP-MS, eight conducted the analysis using LA-ICP-MS, one used LA-ICP-OES, and seven used μ-XRF.

Data analysis

Statistical analyses were performed using 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., Redmond, WA, USA), Plot for mac OSX (v.0.997, Berlin, Germany), Mathematica (v. 5.2.0.0, IL, USA) and R (v2.13.2).

Results and discussion

It has been well established that major, minor, and trace element profiles can provide excellent discrimination among glass sources.[1–13] In order to assess the extent to which this discrimination can be made, participants in three interlaboratory tests provided elemental data measured in their laboratories using several analytical instrumental methods. The μ-XRF users provided fluorescence peak intensity ratios, typically reporting between six and eight ratios for each sample. The ICP-MS and ICP-OES users reported the measured concentrations of up to 18 elements. Analytical data were received from 24 participants in 18 laboratories. The pool of participants used a suite of different instruments, brands, configurations and analytical parameters that represent instrumental techniques currently used by the forensic community. The data were utilized to assess the ability of the participants to correctly associate glass fragments from the same source and to distinguish between fragments from different sources. Throughout this study, a K and Q pair of samples was considered to be indistinguishable when every measured parameter, either element concentration or intensity ratio, for the two samples could not be distinguished using the pertinent match criterion. For the purposes of error rate analysis, the "correct" result was that two samples were considered indistinguishable only when they came from the same small panel of glass in the FIU collection. Two samples produced on the same float line at different times were considered as different sources in assessing the accuracy of conclusions. This approach was taken because in most cases, the question of forensic interest is whether or not two fragments can be associated with the same window, rather than made in the same manufacturing plant.

Results as reported by each participant laboratory using their selected match criteria

Second interlaboratory test. This test was organized like a traditional proficiency test with one K sample and two Q

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

samples. One deficiency detected early in these studies was that the participants did not have a standardized match criterion to reach their conclusions. However, it is noteworthy that all methods gave correct results in this rather simple test. This result was anticipated, since, as shown in Table 3, the concentrations of Ti, Mn, Fe, and Rb are quite different between K1 and Q2. With these results, the protocols for both μ-XRF and LA-ICP-MS were considered robust and further, more difficult tests were designed to evaluate the performance of different match criteria and to work towards standardization of statistical data treatment.

Third interlaboratory test. The purpose of this study was to evaluate the capabilities of each method to discriminate samples manufactured at the same plant at different time intervals (weeks, months and years apart from each other).

There were three pairs of samples that were produced over three years apart; K1/Q2, K1/Q3, and K2/Q1. Based on the results of pre-distribution analysis shown in Table 3, these sample pairs have differences in elemental compositions that were expected to be recognized using sensitive analytical methods.

All fourteen respondents correctly reported that items K2 and Q1, and items K1 and Q2, manufactured more than 3 years apart were distinguishable. Thirteen of the fourteen respondents correctly reported that items K1 and Q3, samples that were manufactured 3 years and 1 month apart were distinguishable. The participant that used SN-ICP-MS reported an inconclusive result due to some uncontrolled problems during the digestion of sample Q3. Thus, each of the participants that completed the analysis was able to correctly discriminate between samples that were manufactured approximately 3 years apart in the same manufacturing plant, despite their indistinguishable refractive indices and physical properties.

The match criteria that were used by the participating forensic glass examiners included spectral overlay, range overlap, several different forms of confidence intervals, the *t*-test, and ANOVA. These criteria were used in various forms either individually or in combinations.

Test 3 also contained three pairs of samples that were produced several weeks to months apart; K1/Q1, K2/Q2, and K2/Q3. The results of pre-distribution analysis shown in Table 3 indicate that these sample pairs have very similar elemental compositions with relatively small differences in the concentrations of some elements. It was expected that these differences could only be detected by those techniques that have good precision of the measurements and low limits of detection.

Only four of the fourteen respondents reported significant differences between items manufactured in the same plant 2 weeks apart (K1 *vs.* Q1). These four respondents used LA-ICP-MS methods to arrive at that conclusion. The discriminating elements reported by the few laboratories that found significant differences between K1 and Q1 were Ba (by three of the four laboratories using LA-ICP-MS) and Mn, K, Zr, Fe, Sr, Sn or Rb. Of the latter, the only other element that was common to two of the laboratories was Zr. Table 3 shows that these elements were present in those samples at concentrations ranging from <2 to 30 μg g^{-1} and therefore only sensitive methods with excellent precision between measurements would be able to detect those differences in concentration. For example, the reported Ba concentration difference between the two samples is approximately 2.5 μg g^{-1}. Iron, the trace element having the largest difference in concentration, was reported as significantly different in the two samples by only one of the LA-ICP-MS laboratories. This result is consistent with the fact that not only the differences in mean element concentrations, but also the variations of the measured data within a sample, are important in defining the ability of a method to distinguish two different sources of glass.

Thirteen of the fourteen respondents reported that item K2 was distinguishable from item Q2. The only respondent that could not distinguish between item K2 and item Q2 used μ-XRF. However, after discussion of the results, this respondent re-examined their data and found significant differences in the Fe/Mn peak intensity ratios that were missed during the test. These samples were manufactured at the same plant 1 month apart and their elemental compositions are similar, but significant differences were detected for some elements, in particular Mn and Ti for μ-XRF measurements and between 6 and 12 elements for the ICP measurements.

Five of the fourteen respondents reported that item K2 was distinguishable from item Q3. None of the seven μ-XRF users were able to differentiate these two samples. The participants that were able to detect differences between these samples used LA-ICP-MS techniques. Two of the ICP-MS users did not differentiate this pair. Lab ICP-A was unable to differentiate the samples probably as a consequence of their smaller number of elements measured, and their match criteria that allows one element to differ and still call the results indistinguishable. Lab ICP-F, who used SN-ICP-MS, reported an inconclusive result due to problems with the digestion of sample Q3. Samples K2 and Q3 were manufactured at the same plant 3 months apart. However, the elemental compositions of K2 and Q3 are more similar than those of K2 and Q2, which were manufactured only one month apart. The concentrations of discriminating elements were present at levels below 30 μg g^{-1} in the glass.

This third interlaboratory test allowed the study of the ability of the different analytical methods to discriminate among samples that shared very similar composition. All techniques were able to differentiate samples manufactured three years apart in the same plant, regardless of the match criteria employed by each respondent. Samples manufactured weeks to months apart could only be differentiated in some instances by the more sensitive analytical techniques.

Fourth interlaboratory test. The EAWG members felt that the results of the third interlaboratory test were encouraging, particularly in the excellent ability of the ICP-MS methods to discriminate glass sources produced over fairly short time periods. However, the high degree of source discrimination could lead to the incorrect source exclusion of glass fragments that came from the same source. To address this, a fourth interlaboratory test was designed and carried out. The set of samples for this test was selected primarily with the aim of studying type 1 errors, although one sample was also included to evaluate type 2 errors on samples produced in the same

 This journal is © The Royal Society of Chemistry 2013

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

manufacturing plant at different times. There were seventeen participants in this test, including one who used an additional technique, LA-ICP-OES.

Samples K1, K2, Q2 and Q3 all originated from glass manufactured at the Pilkington plant on March/03/2010. Sample Q1 was manufactured at the same plant two weeks later (February/18/2010). As shown in Table 3, the pre-distribution analysis indicated that the composition of Q1 is significantly different from that of the other samples. To simulate typical casework, known samples K1 and K2 consisted of three small full thickness fragments, while questioned samples were each three small irregular fragments of approximately 0.5-1 mm in size. The participants were instructed to make a comparison and reach an opinion concerning possible source for each Q fragment separately, rather than grouping them as was done in the prior tests. The participant (ICP-F) who used SN-ICP-MS was unable to analyze the small Q fragments individually due to the sample size requirements in the ASTM method, and thus combined the three fragments for digestion and analysis.

Sixteen of seventeen respondents correctly reported that all of the Q1 fragments were distinguishable from items K1 and K2. Although these samples were manufactured only 2 weeks apart on the same float line, significant differences exist in composition for Fe, Al and Ti and several trace elements that were readily detected by all participant methods. One of the ICP participants, designated as ICP-L, does not conduct glass comparisons on a routine basis at their laboratory and therefore only reported their measured concentration data and did not make a decision of association or exclusion. Their data were only utilized for comparison of match criteria in the next part of this study.

All of the K1, K2, Q2, and Q3 fragments came from the same pieces of a single glass sheet, so they should have been associated by the participants. Using their selected match criteria, all seven respondents that used μ-XRF correctly reported that each fragment labeled as item Q2 or Q3 were indistinguishable from both K1 and K2. Therefore, all participants who used μ-XRF were correct in both their distinguishable and indistinguishable conclusions in this fourth interlaboratory test. It is noteworthy that by the completion of this fourth interlaboratory test most μ-XRF participants agreed on the selection of match criteria for their comparisons, based on previous results and discussion from the interlaboratory tests. All participants used spectral overlay as a preliminary assessment of similarity followed by a $\pm 3s$ criterion for comparison of intensity ratios with the exception of one laboratory that used range overlap.

On the other hand, of the 88 reported comparisons for these four samples made by the ten participants using ICP-based methods, there were 16 incorrect discriminations of fragment pairs. Labs A, E, H, and K correctly found each of the Q2 and Q3 fragments to be indistinguishable from both K1 and K2. Lab C had only one incorrect result for a K1/Q3 comparison. Lab F, the one that used SN-ICP-MS had one incorrect result, but it was out of only four comparisons because the limited fragment size forced grouping of the fragments for each sample for digestion. The majority of incorrect exclusions were made by Lab D with six and Lab J with eight. These two participants used the *t*-test with Bonferroni correction for their match criterion.

These false exclusion results raised a flag for further discussion by the EAWG members concerning the appropriate match criteria for ICP-based methods. Past experience of ICP-MS and ICP-OES users was that false exclusions rarely occur when an appropriate number of elements are used with reasonable match criteria. The observed rate of false exclusions, particularly for Labs D and J were unexpectedly high. It was felt that this high false exclusion rate was a result of participants using match criteria that were too narrow when considering the relatively large number of elements measured. Therefore, further data analysis was conducted to assess the error rates for a number of match criteria with the aim of finding an optimum match criterion that would simultaneously minimize both type 1 and type 2 errors. The results of these studies are discussed in the following section.

Evaluation of performance of different match criteria

In order to evaluate how the choice of match criterion affects error rates, the data provided by each participant were used to assess the error rates for the following criteria for the μ-XRF methods: range overlap, *t*-tests ($p = 0.05$, 0.01 and Bonferroni correction to 0.05), confidence intervals ($\pm 2s$, $3s$, $4s$), Hotelling T^2, and for the ICP-based methods: these plus modified confidence intervals ($\pm 2s$, $3s$, $4s$, $5s$, $6s$ with minimum 3% RSD). The calculations of error rates were performed for data collected for the second, third and fourth interlaboratory tests. The data from each of the individual Q fragments were used when making the comparison to the known sample for the purpose of this error rate analysis. Therefore, each K/Q comparison was made between nine or more measurements from the K sample and three measurements from the Q sample. The result of a comparison was declared as indistinguishable when the values for all measured variables met the match criterion, otherwise the samples were deemed to be distinguishable.

The second test had one K/Q pair that originated from the same source (K1 *vs.* Q1), which allowed the evaluation of false exclusions, or type 1 errors and one pair of samples that originated from different sources (K1 *vs.* Q2), which allowed the evaluation of false inclusions or type 2 errors. The third test had five glass items, 2 known samples and 3 questioned samples, all of them originating from the same plant manufactured on different dates. Because this test did not have pairs of samples that originated from the same source, it did not have the possibility for type 1 errors. There were six sample pair comparisons that could result in false associations, or type 2 errors; (K1/Q1, K1/Q2, K1/Q3, K2/Q1, K2/Q2, and K2/Q3). The fourth test had five glass items, two known and 3 questioned samples. Two K/Q comparison pairs allowed the evaluation of type 2 errors (K1/Q1 and K2/Q1) and 4 K/Q sample pairs (K1/Q2, K1/Q3, K2/Q2, and K2/Q3) were used to evaluate the rate of type 1 errors. For each sample pair, the number of errors was determined three times for the individual fragments of each Q sample and summed across all participants reporting results for that sample pair. All reported sample pairs were used to calculate the number of incorrect associations using each of the tested match criteria.

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

Error rates for μ-XRF data

The summary results of error rate analyses obtained using μ-XRF data for different match criteria expressed as the percentages of incorrect associations or exclusions are shown in Table 4. The number of comparisons used to calculate each percentage is given in the footnote to the table. False inclusions, or type 2 errors, were determined for the data from all three interlaboratory tests. The rate of false inclusions on this test was very low regardless of the match criteria employed for μ-XRF data. For the 68 sample pair comparisons made for the second and fourth interlaboratory tests, only one pair resulted in a type 2 error. This error only occurred for the *t*-test at $p = 0.01$, the *t*-test with Bonferroni correction, and the 4*s* test for the second test. The sample pairs used for the evaluation of type 2 error rates on the second and fourth interlaboratory tests were manufactured in the same plant more than 2 years apart and 2 weeks apart, respectively. Their elemental composition was fairly distinctive and significant differences were detectable using μ-XRF methods.

As expected, the type 2 error rates on the third interlaboratory test are larger than for the other tests because the samples for this test were manufactured on the same float line and, in some cases, at relatively short date intervals. As a result, these samples have only minor differences in elemental composition. The *t*-test at the 0.05 and 0.01 levels resulted in the lowest numbers of type 2 errors in this set, 6% and 15%, respectively.

Type 1 error rates (false exclusions) were determined for the second and fourth tests. At least one false exclusion was observed for all of the match criteria except for 4*s* in the fourth test. The number of type 1 errors when using the narrower match criteria of the *t*-tests and the 2*s* test are generally quite high. The range, 3*s*, 4*s*, and Hotelling's T^2 tests, with their wider match criteria, result in more acceptable type 1 error rates. The high number of type 1 errors is somewhat surprising considering that μ-XRF measurements have repeatability values of approximately 10% for elements present at lower concentrations, such as Ti, Sr, and Zr. A likely reason as to why all pairs of samples from the same source are not correctly associated is that the irregular shapes and small sizes of the Q fragments result in poorer precision or biases in measured intensities when compared to the data from the larger, multiple K fragments. As can be seen in Table 4, the tests estimating population distributions (2*s*, 3*s*, and 4*s*) tend to be more forgiving of these impairments than tests that rely more on a representative sampling of both items (range overlap and *T*-tests). It is also significant that the conclusions reported by individual participants in the interlaboratory tests were all correct. There are several possible reasons for the better performance by the participants than that indicated by the various match criteria. First, the participants used spectral overlay as a pretest prior to comparison of analytical data. As a result, they may have removed some elements that were present at close to limits of quantitation from further quantitative comparison. Some of the false exclusion errors seen here result from comparisons of ratios involving these barely detectable elements which are more prone to sample size and orientation errors than elements present at higher concentrations.[40] The second reason for the lower number of type 1 error rates for individual participant is that they grouped the data for Q fragments in Test 2, which improved the error rates compared with treating fragments individually.

For a compromise between type 1 and type 2 error rates, the optimum match criteria were 3*s*, range overlap, and Hotellings T^2. As shown in Table 4, 3*s* and Hotellings T^2 criteria had higher rates of false inclusions than range overlap for the data from the third test, which employed samples having very similar elemental compositions. However, the range overlap criterion had a significantly greater false exclusion rate than the 3*s* or Hotellings T^2 criteria for the small irregular shaped fragments encountered in the fourth test. One advantage of μ-XRF data is that the typical number of variables (6–8 ratios) allows the fulfillment of the requirement of Hotellings T^2 to have more replicate measurements than variables (*i.e.*, at least 5 to 7 replicate measurements for the known sample and at least 3 for each questioned sample). However, in instances with small Q fragments such as debris cases, it may not be practical, or even possible, to collect the required number of replicate measurements on each fragment unless the position of the X-ray beam remains stationary between measurements. Spectral overlay was not included in the tested match criteria because it is a qualitative comparison. However, based on the experience of EAWG members and the results reported by the participants in the interlaboratory tests, spectral overlay is one of the best match criteria. A protocol for μ-XRF that was recently approved by ASTM as a standard test method recommends the use of spectral overlay followed by either a 3*s* or range overlap match criterion using element intensity ratios. Although the results of this study cannot be applied directly to other manufacturers or even other dates for the float lines studied, they should be generally applicable. That is, the μ-XRF methods are capable of detecting differences in composition of flat glass from the same line within a float glass plant when they are produced over time periods of weeks to months apart.

Table 4 Error rate results for the application of different match criteria on data acquired by XRF methods

	Type 1 error rate (%)			Type 2 error rate (%)		
Match criteria	Test 2[a]	Test 3[b]	Test 4[c]	Test 2[d]	Test 3[e]	Test 4[f]
Range	11	—	19	0	21	0
t-Test .05	52	—	60	0	6	0
t-Test .01	22	—	30	4	15	0
t-Test Bonf.	15	—	26	4	21	0
±2*s*	41	—	24	0	18	0
±3*s*	11	—	6	0	27	0
±4*s*	7	—	0	4	36	0
Hotellings T^2	15	—	9.5	0	26	0

[a] Percent rate calculated out of 27 comparisons from 9 laboratories. [b] Design of the round robin 3 did not account for estimation of type 1 errors. [c] Percent rate calculated out of 84 comparisons from 7 laboratories. [d] Percent rate calculated out of 26 comparisons from 9 laboratories. [e] Percent rate calculated out of 124 comparisons from 7 laboratories. [f] Percent rate calculated out of 42 comparisons from 7 laboratories.

 This journal is © The Royal Society of Chemistry 2013

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

Error rates for ICP data

The summary of error rates obtained for ICP data for different match criteria are shown in Table 5. Because of the good precision of most ICP data, additional broader match criteria were included in this study (5*s*, 6*s* and modified confidence intervals with minimum of 3% RSD).[26] False inclusions or type 2 errors were estimated for the three tests. The only type 2 error that was made for the samples of the second and fourth interlaboratory tests was from the *t*-test with Bonferroni correction for one fragment from one participant in the second test. The samples used for the evaluation of type 2 error rates on these sets were manufactured in the same plant more than 2 years apart or 2 weeks apart, respectively. However, differences in their elemental composition as measured by ICP methods are detectable by any of the match criteria.

The type 2 error rate in the third interlaboratory test was expected to be larger than the other tests because the samples were purposely selected to be closer in manufacture date and also very similar in elemental compositions. In spite of this, the false inclusion rate was very low. All the K/Q pairs that showed false inclusions came from the pair of samples manufactured only 2 weeks apart, demonstrating that the sensitivity and precision of ICP data allows for the discrimination of samples manufactured at the same plant during short time intervals. Confidence intervals greater than 5*s* provided the largest number of type 2 errors in this set.

Type 1 error rates, or false exclusions, were determined for the second and fourth tests. Some false exclusions were observed for the majority of the match criteria, with lower rates provided by broader match criteria (>4*s*). Failure to associate samples with the same origin was observed in the second test only for 2 out of 7 participant laboratories and in all cases the differences were found only for one out of the 16–18 elements monitored. Repeatability between measurements in the discriminating element was lower than 2% RSD. The larger number of type 1 errors on the fourth test is attributed mainly to the atypical heterogeneity discovered in the samples, which is discussed in more detail in the following section.

The best performance for a compromise between type 1 and type 2 error rates is found for 4*s* and modified 4*s* interval. Using these broader match criteria reduced significantly the Type 1 errors without sacrificing the capability to discriminate samples (type 2 errors). Hotellings T^2 is not as practical for ICP data as for the μ-XRF data due to the larger number of variables measured (16–18 elements). However, Hotellings T^2 could be applied in cases where the questioned sample is large enough to allow the requisite number of replicate measurements. Since that was not the case in these studies, no statement can be made as to the error rates that might result when using Hotelling's T^2 with ICP data.

The need to widen the match criteria for ICP measurements is a consequence of the high precision of the measurements (typically less than 2% RSD). Using the broader criteria, the ICP methods were still able to correctly discriminate between samples with similar elemental profiles that originated from the same plant and were manufactured more than 2 weeks apart. It should be noted here that the significant factor affecting changes in composition of float glass is not time, *per se*, but rather changes in the compositions of raw materials and internal processes within the manufacturing plant that occur over time. Again, the results of this study cannot be applied directly to other manufacturers or even other dates for the float lines studied. However, they should be generally applicable in that the ICP-based methods, when applied to many major, minor, and trace elements, are capable of detecting differences in composition of flat glass originating from one plant over time periods of weeks to months.

Table 5 Error rate results for the application of different match criteria on data acquired by ICP methods

	Type 1 error rate (%)			Type 2 error rate (%)		
Match criteria	Test 2[a]	Test 3[b]	Test 4[c]	Test 2[d]	Test 3[e]	Test 4[f]
Range	42	—	81	0	0	0
t-Test .05	74	—	93	0	1	0
t-Test .01	53	—	84	0	1	0
t-Test Bonf.	53	—	69	0	2	0
±2*s*	53	—	85	0	0	0
±2*s* (*s* > 3%)	26	—	75	0	0	0
±3*s*	42	—	66	0	2	0
±3*s* (*s* > 3%)	0	—	47	0	2	0
±4*s*	26	—	42	0	5	0
±4*s* (*s* > 3%)	0	—	28	0	5	0
±5*s*	11	—	30	0	9	0
±5*s* (*s* > 3%)	0	—	18	0	11	0
±6*s*	11	—	27	0	12	0
±6*s* (*s* > 3%)	0	—	13	0	15	0

[a] Percent rate calculated out of 19 comparisons from 7 laboratories. [b] Design of the round robin 3 did not account for estimation of type I errors. [c] Percent rate calculated out of 120 comparisons from 10 laboratories. [d] Percent rate calculated out of 19 comparisons from 7 laboratories. [e] Percent rate calculated out of 126 comparisons from 7 laboratories. [f] Percent rate calculated out of 60 comparisons from 10 laboratories.

Homogeneity study: Pilkington and Cardinal plants

The samples selected for the fourth interlaboratory test originated from a Pilkington glass manufacturing plant that

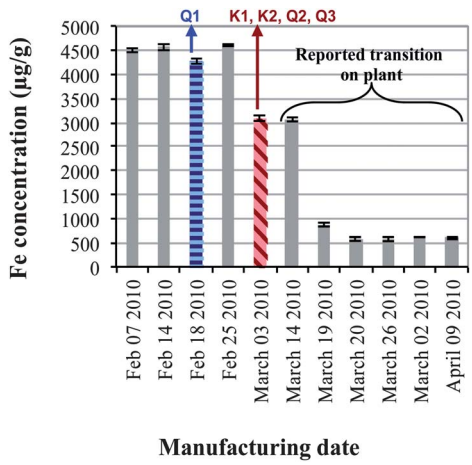

Fig. 1 Elemental variation of iron composition between samples manufactured at the Pilkington plant from February/07/2010 to April/09/10.

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

experienced changes in the formulation of the glass due to market requirements. Fig. 1 shows the variation of concentration of iron in glass samples collected over a 2-month period. Error bars represent the variation (as standard deviation) obtained from 5 replicates of a single sample measured by LA-ICP-MS. Drastic concentration changes in iron content were observed in glass manufactured between February 25, 2010 and March 19, 2010. Nevertheless, the plant reported that their "transition period", where the glass was not released to the market, was between March 14, 2010 and April 09, 2010.

Samples selected for the interlaboratory test were manufactured approximately two weeks and one month before the transition period, respectively. Due to the unexpectedly high rates of false exclusions found in the fourth interlaboratory test by ICP-based methods, it was suspected that one cause for the occurrence of this type of error could be an atypical heterogeneity at the micro-scale of the samples submitted for analysis.

To test this hypothesis, homogeneity studies were conducted on the original source samples from the Pilkington plant. In addition, a set of glass samples from another plant was included for comparison purposes. The Cardinal sample manufactured on August 17, 2001, which was used as K1 for the third interlaboratory test, was selected for this study because in that interlaboratory test none of the participants reported false exclusion errors in their findings.

The homogeneity study was designed to compare the variation: (a) between fragments from the same source (*i.e.* 6 fragments from the same source, 3 replicates each) and (b) within fragments (spatial variation, *i.e.* float side. non-float side, and different areas through a cross-section of the glass).

Variation between fragments from the same source (non-float sides only)

Six fragments were randomly selected from each sample, 3 replicates were conducted on each of the non-float original surfaces of the fragments for a total of 18 measurements per sample.

Comparisons between the six fragments *versus* each other were conducted using ANOVA with Tukeys post-hoc test. The results show more heterogeneity in samples from the Pilkington plant than in the samples from the Cardinal plant. Significant differences for 7 out of 18 elements monitored (Mn, Al, K, Ca, Ti, Fe and Sr) were observed between fragments from the Pilkington glass manufactured on February 18, 2010 and for 3 out of the 18 elements monitored (Al, Ca and Hf) for the Pilkington sample manufactured on March 3, 2010. In contrast, no significant differences were observed between any of the fragments sampled from the Cardinal glass using the same criterion.

In order to simulate the statistical treatment given to the data during the fourth interlaboratory test, the comparison between fragments was also conducted by randomly choosing 3 of the fragments to act as the Known sample (K) and the remaining fragments as independent Questioned samples (Q), with 3 measurement replicates each. Only non-float surfaces were analyzed during this experiment.

In general, the Cardinal glass sample showed evidence of uniform distribution of elemental composition among non-float surfaces. No significant differences were detected between fragments using different match criteria (except *t*-test $p = 0.05$). On the other hand, the Pilkington samples showed more heterogeneity, as evidenced by more false exclusions than the Cardinal sample. The false exclusions were reduced to zero for the three Pilkington subsets when applying broader match criteria, such as $\pm 4s$ and modified $\pm 4s$ (minimum 3–5% RSD).

Variation between original surfaces and across the thickness of the fragment

The sampling scheme used to study spatial variation within a single fragment include five replicate analyses that were conducted on each of the sampling sites (original surfaces such as float *versus* non-float side, and fracture surfaces as different areas across the thickness of the fragment).

Significant differences in composition were found between float and non-float surfaces on all tested samples (Cardinal and Pilkington) using all match criteria. Differences between the surfaces were detected not only for the content of Sn, but also for other elements, such as Al, Ca, Sr, Zr, Ti and Fe.

These observations highlight the relevance of sampling. Whenever possible, sampling from fracture surfaces is preferred over original surfaces. Otherwise, if analyzing original surfaces, either all non-float surfaces or all float surfaces should be used for the comparisons of known and questioned sources. An easy way to detect if the analysis is being done on the float side is to monitor the content of Sn, which will typically be 1–2 orders of magnitude larger on the float side of the glass. This observation can be done *in situ* during the analysis and the sample can be easily turned to the non-float side if needed.

The study of elemental variability across the thickness of the interior portion of glass fragments also revealed more heterogeneity in the Pilkington samples. Significant differences were detected, depending on the match criteria, particularly close to the non-float surface (<200 μm). This variability was detected regardless of the match criteria applied to sample PK030310, which was manufactured close to the time of the reported transition in Fe formulation. This sample was the one selected for the fourth interlaboratory test to examine false exclusion errors.

Significant differences across the thickness of the sample were also detected for the Cardinal sample, depending on the match criteria. Nevertheless, no significant differences were detected using $\pm 4s$ or modified $\pm 4s$ match criteria.

The heterogeneity between fragments was more pronounced for the Pilkington samples than for the Cardinal samples. Due to the shape and small fragment size chosen for the Q samples on the fourth interlaboratory test, it is likely that participants received, Q samples from different areas across the thickness of the fragment and encountered heterogeneous compositions. This could have contributed to the elevated number of false exclusions obtained with sensitive ICP-based methods.

Consequently, based on post distribution experiments and various statistical evaluations of the data, the rate of false

 This journal is © The Royal Society of Chemistry 2013

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

exclusions found in the fourth test for ICP-based methods is attributable to several factors: (a) the limited number of replicates for questioned samples, a common casework concern, (b) unusual heterogeneity of the samples of Pilkington glass, and (c) match criteria too sensitive for methods achieving very high precision between replicates.

In the fourth interlaboratory test, participants were asked to compare the known fragments to each of the individual questioned fragments (instead of grouping all questioned fragments). This approach was selected to be a more realistic simulation of a case where small, irregular fragments recovered from surface debris must be treated as individual fragments. Due to the small size of the fragments submitted for analysis, only 3 replicates were requested per questioned fragment. As a result, for the sensitive ICP-based methods that have high precision, only some of the fragments were correctly identified as indistinguishable from the known sources. The precision and sensitivity of μ-XRF techniques, in combination with the selected match criteria, were shown to be appropriate for these types of samples.

The heterogeneity observed on the Pilkington samples, both within a fragment and between fragments originating from a single source is atypical of what has been observed in the float glass encountered in several manufacturing plant studies analyzed by SN-ICP-MS and LA-ICPMS at FIU over the last decade. Heterogeneity of these samples is also inconsistent with previous within-sheet homogeneity studies conducted at FIU, the BKA, and the FBI. Nevertheless, as with any commercial product, the variability of its elemental composition is dependent on market requirements and the manufacturing history of the specific plant. For these reasons, samples such as those encountered in this fourth interlaboratory test may be present in a real case and should be taken into account during the selection of match criteria and interpretation of the data.

Conclusions

The interlaboratory test results can be used as a guide for forensic scientists in the selection of the match criteria for comparison of elemental composition of glass fragments. The analytical technique used to analyze both known and questioned samples as well as the number of replicates used to characterize the variability of the known samples should be considered when selecting the match criteria.

For μ-XRF analysis, the following observations are derived from the studies. Spectral overlay, $\pm 3s$, range overlap, and Hotellings T^2 performed well in terms of both false exclusions and inclusions. Excellent consistency of reported comparison results among participants was achieved for all the interlaboratory tests, not only for comparison conclusions but also for the elements reported to be responsible for discrimination. Participants who used μ-XRF methods were able to detect significant differences between fragments of glass that were manufactured in the same plant within short periods of time. That period of time is dependent on the variability of the formulation of the glass within a plant. For instance, participants who used μ-XRF were able to detect differences in samples manufactured a month apart at the Cardinal plant in this study. Differences were not detected between samples manufactured at this plant 2 weeks apart and 3 months apart when their elemental compositions were extremely similar. However, all participants were able to detect significant differences between samples manufactured 2 weeks apart at the Pilkington plant in this study. Users of μ-XRF must take into consideration that small, irregularly shaped fragments may result in false exclusions when they are compared to larger fragments, particularly when they are thin enough that high energy X-rays penetrate completely through them.

For ICP-based methods (digestion SN-ICP-MS, LA-ICP-MS and LA-ICP-OES), the following conclusions are derived from the studies. Most participants reported precisions between replicates of 2% RSD or less. This good analytical precision may be one of the factors that contribute to higher false exclusion rates when sensitive match criteria such as the *t*-test or a low multiple of standard deviations are used. Due to the sensitivity of the method, the capability of multi-elemental analysis of trace elements, the typical high precision, and the concerns for heterogeneity, the use of broader match criteria such as $\pm 4s$ is recommended, either with or without minimum precision values depending on the reproducibility within replicates. These match criteria still allow detection of significant differences between samples manufactured in the same plant over short time intervals, even for samples with quite similar elemental profiles. The performance of these match criteria is in agreement with recent published data.[33] As with μ-XRF methods, the time interval over which samples cannot be distinguished depends upon the variability of the formulation of the glass within a plant. For instance, ICP participants were able to detect differences in samples from the same float line at the Cardinal plant manufactured a month apart and some participants detected differences of samples manufactured 2 weeks apart.

In terms of interpretation of elemental comparisons of glass, it can be concluded from the study that glass samples that are manufactured in different plants, or even at the same plant years apart, are clearly differentiated by elemental composition when μ-XRF or ICP-based methods are used for analysis. Samples produced in the same plant over time intervals of weeks to months may also be differentiated. This level of differentiation can be used to add significance to an association, when one is found, and to assist in assigning recovered fragments to a source when selecting among several potential sources.

Acknowledgements

This project was supported by Award no. 2010-DN-BX-K179 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this presentation are those of the authors and do not necessarily reflect those of the Department of Justice, the Department of the Army or the Department of Defense. This is publication number 12-15 of the FBI Laboratory Division. Names of commercial manufacturers are provided for identification purposes only,

Published on 11 June 2013. Downloaded by University of North Texas on 12/18/2020 10:30:14 PM.

and inclusion does not imply endorsement of the manufacturer, or its products or services by the FBI. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the FBI or the U.S. Government. Thanks to all of the Elemental Analysis Working Group members and their agencies for devoting the time necessary to complete this project.

References

1 J. R. Almirall, J. Buckleton, J. M. Curran and T. N. Hicks, in *Forensic Interpretation of Glass Evidence*, ed. J. M. Curran, T. N. Hicks and J. S. Buckleton, CRC Press, Florida, 2000, pp. 1–24.
2 R. D. Koons, C. A. Peters and P. S. Rebbert, *J. Anal. At. Spectrom.*, 1991, **6**, 451.
3 T. Hicks, S. F. Monard, T. Goldmann, A. Brunelle, C. Champod and P. Margot, *Forensic Sci. Int.*, 2003, **137**, 107.
4 J. Andrasko and A. Machly, *J. Forensic Sci.*, 1978, **22**, 250.
5 S. Becker, L. Gunaratman, T. Hicks, W. Stoecklein and G. Warman, *Probl. Forensic Sci.*, 2001, **47**, 80.
6 D. C. Duckworth, S. J. Morton, K. Bayne, S. Montero, R. D. Koons and J. R. Almirall, *J. Anal. At. Spectrom.*, 2002, **17**, 662.
7 R. D. Koons, C. Fiedler and C. Rawalt, *J. Forensic Sci.*, 1988, **33**, 49.
8 S. Montero, A. Hobbs, T. French and J. R. Almirall, *J. Forensic Sci.*, 2003, **8**, 1101.
9 J. R. Almirall and T. Trejos, *Forensic Sci. Review.*, 2006, **18**(2), 74.
10 S. Ryland, *J. Am. Soc. Trace Ev. Exam.*, 2011, **2**, 2.
11 Z. Brozel-Mucha and G. Zadora, *Probl. Forensic Sci.*, 1998, **37**, 68.
12 P. Kuisma-Kursula, *X-Ray Spectrom.*, 2000, **29**, 111; V. Reeve, J. Mathiesen and W. Fong, *J. Forensic Sci.*, 1976, **21**, 291.
13 R. J. Dudley, C. R. Howden, T. J. Taylor and K. W. Smalldon, *X-Ray Spectrom.*, 1980, **9**, 119.
14 C. R. Howden, R. J. Dudley and K. W. Smalldon, *J. Forensic Sci. Soc.*, 1978, **18**, 99.
15 T. Catterick and D. A. Hickman, *Forensic Sci. Int.*, 1981, **17**, 253.
16 K. L. Wolnik, C. M. Gaston and F. L. Fricke, *J. Anal. At. Spectrom.*, 1989, **4**, 27.
17 L. Moenke-Blankenburg, T. Schumann, D. Günther, H. M. Kuss and M. Paul, *J. Anal. At. Spectrom.*, 1992, **7**, 251.
18 A. Zurhaar and L. Mullings, *J. Anal. At. Spectrom.*, 1990, **5**, 611.
19 T. Parouchais, I. M. Warner, L. T. Palmer and H. Kobus, *J. Forensic Sci.*, 1996, **41**, 351.
20 D. C. Duckworth, C. K. Bayne, S. J. Morton and J. R. Almirall, *J. Anal. At. Spectrom.*, 2000, **15**, 821.
21 C. Latkoczy, *et al.*, *J. Forensic Sci.*, 2005, **50**, 1327.
22 S. Becker, P. Watzke, M. Dücking and W. Stoecklein, *Forensic Sci. Int.*, 2003, **136**(1), 361.
23 B. Naes, *et al.*, *Spectrochim. Acta, Part B*, 2008, **63**(10), 1145.
24 E. Cahoon and J. R. Almirall, *Applied Optics*, 2010, **49**(13), C49.
25 National Research Council. *Strengthening Forensic Science in the United States: A Path Forward*, The National Academies Press, Washington DC, 2009.
26 T. Trejos, *et al.*, *Anal. Bioanal. Chem.*, 2013, **405**(16), 5393.
27 T. Trejos and J. R. Almirall, *Talanta*, 2005, **67**, 388.
28 T. Trejos and J. R. Almirall, *Talanta*, 2005, **67**, 396.
29 J. M. Curran, T. Hicks and J. S. Buckleton, *Forensic Interpretation of Glass Evidence*, CRC Press, Florida, 2000, pp. 27–73.
30 J. M. Curran, C. M. Triggs, J. R. Almirall, J. S. Buckleton and K. A. J. Walsh, *Sci. Justice*, 1997, **37**, 241.
31 R. D. Koons and J. Buscaglia, *J. Forensic Sci. Soc.*, 2002, **47**, 505.
32 S. Bajic, D. Aeschliman, N. J. Staetveit, D. P. Baldwin and R. S. Houk, *J. Forensic Sci.*, 2005, **50**, 1123.
33 P. Weis, M. Dücking, P. Watzke, S. Menges and S. Becker, *J. Anal. At. Spectrom.*, 2011, **26**, 1273.
34 S. Berends-Montero, W. Wiarda, P. de Joode and G. van der Peijl, *J. Anal. At. Spectrom.*, 2006, **11**, 1185.
35 J. N. Miller and J. C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, Pearson Education Limited, UK, 4th edn, 2000, pp. 42, 57–67 and 74–80.
36 R. A. Johnson and D. W. Wichern Applied *Multivariate Statistical Analysis*, Prentice Hall, New Jersey, 3rd edn, 1982.
37 C. G. G. Aitken and F. Taroni, *Statistics and the Evaluation of Evidence for Forensic Scientists*, John Wiley & Sons Ltd., England, 2nd edn, 2004.
38 E. J. Garvin and R. D. Koons, *J. Forensic Sci.*, 2011, **56**(2), 491.
39 American Society for Testing and Materials Method E2330-04, 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, PA, 2004.
40 T. Ernst, T. Berman, J. Buscaglia, *et al.*, *X-Ray Spectrom.*, 2012, 1–9.

 This journal is © The Royal Society of Chemistry 2013