

**The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:**

**Document Title:           Significance of Elemental Analysis from Trace Evidence**

**Author(s):                 Jose Almirall**

**Document No.:           242325**

**Date Received:           June 2013**

**Award Number:          2009-DN-BX-K252**

**This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant report available electronically.**

**Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.**

**Final Technical Report for grant number 2009-DN-BX-K252 from the  
Department of Justice, Office of Justice Programs, National Institute  
of Justice**

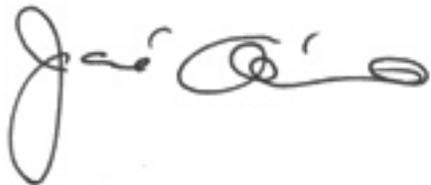
**Title: Significance of Elemental Analysis from Trace Evidence**

**Award Period:** December 2009 – December 30, 2012

**Principal Investigator:** Jose Almirall, Professor, *Department of Chemistry and Biochemistry* and Director, International Forensic Research Institute  
Florida International University

[almirall@fiu.edu](mailto:almirall@fiu.edu) - 305.348.3917 (Phone) - 305.348.4483 (Fax)

Submitted on October 19, 2012



**Executive Summary:**

This research project aims to improve the **scientific validity** of the **probative information** resulting from applying elemental analysis methods in the examination of trace evidence materials. The first part of the project coordinated a series of five (5) inter-laboratory exercises (also called round-robin exercises in this report) involving a large group of experienced examiners from several established operational forensic laboratories using mature forensic techniques and certified reference materials to demonstrate that the quality of the chemical information derived from a variety of analytical methods is extremely good, regardless of the method used. The focus of the inter-laboratory exercises was on the characterization of glass evidence (RR1-RR4) but the soil matrix was also analyzed during one of the inter-laboratory trials (RR5). Many studies have been conducted over the last three decades illustrating the **discrimination power** of elemental analysis for glass (and other types of trace evidence), particularly when LA-ICP-MS was used. The technique of  $\mu$ XRF is preferred in the US forensic laboratory community with approximately 40 forensic laboratories

employing the technique currently but, until this project, there was no standardization of the  $\mu$ XRF technique nor was there a systematic evaluation of the performance of  $\mu$ XRF for forensic glass examination by a large number of users. More than 30 experienced forensic examiners, academics and other interested scientists representing 23 organizations (Elemental Analysis Working Group (EAWG)) worked together to enhance the understanding of the selection of a ***match criteria*** (determination of match/no match based on objective and statistically valid reasoning) for univariate and multivariate comparisons. Two of the round robin exercises were specifically designed to evaluate the performance of the different match criteria currently in use by forensic scientists in an attempt to elucidate the best-performing match criteria for both  $\mu$ XRF data comparisons and LA-ICP-MS data comparisons. The results of these exercises are reported in chapters 2 and 3 of this report. There is also growing evidence that the emerging analytical technique of Laser Induced Breakdown Spectroscopy (LIBS) can produce similar information to both  $\mu$ XRF and LA-ICP-MS for the analysis of materials while also provide significant advantages in the form of ease of use, cost and speed of analysis and therefore also evaluated during this phase of the project.

The forensic utility of LA-ICP-OES is also expected to provide similar advantages (in comparison to LA-ICP-MS) and was also be evaluated during the second part of the project which also resulted in a publication submitted to a scientific journal (Appendix B of this report). Finally, a number of interested EAWG members also completed an interlaboratory exercise (RR5) focused on the bulk analysis of soils for forensic application. This exercise resulted in the submission of a publication to a scientific journal and is referenced in the report.

The major accomplishments of this project were: two (2) ASTM methods of analysis (submitted as appendix C and appendix D in this report) and five (5) publications related to the project research goals (see complete list of publication and presentation references as appendix E of this report).

### **EAWG Working Group Participants:**

Jose Almirall<sup>1</sup>, Tatiana Trejos<sup>1</sup>, Erica Cahoon<sup>1</sup>, Emily Shenck<sup>1</sup>, Sarah Jantzi<sup>1</sup>, Stefan Becker<sup>2</sup>, Peter Weis<sup>2</sup>, Marc Dücking<sup>2</sup>, JoAnn Buscaglia<sup>3</sup>, Robert Koons<sup>4</sup>, Scott Ryland<sup>5</sup>, Ted Berman<sup>5</sup>, Kristine Olsson<sup>6</sup>, Tiffany Eckert-Lumsdon<sup>7</sup>, Melissa Valadez<sup>8</sup>, Randall Nelson<sup>9</sup>, Edward Chip Pollock<sup>10</sup>, Christopher Hanlon<sup>11</sup>, David Rudell<sup>12</sup>, Alex Heydon<sup>12</sup>, Adolfo Caballero<sup>13</sup>, Vincent Zdanowicz<sup>14</sup>, Shirly Montero<sup>15</sup>, Andrew van Es<sup>15</sup>, Jhanis Rodriguez<sup>16</sup>, Troy Earnst<sup>17</sup>, Theresa Hosick<sup>18</sup>, Chris Palenik<sup>19</sup>, Claude Dalpe<sup>20</sup>, Igor Gornushkin<sup>21</sup>, Steve Buckley<sup>22</sup>, Ela Bakowska<sup>23</sup>

1. Department of Chemistry and Biochemistry and International Forensic Research Institute, Florida International University, Miami, FL, USA
2. Forensic Science Institute, Bundeskriminalamt (BKA), Wiesbaden, Germany
3. Federal Bureau of Investigation (FBI laboratory CFRSU), Quantico, Va
4. Consultant (Retired from FBI Laboratory CFRSU), Virginia
5. Florida Department of Law Enforcement, Orlando, FL
6. Johnson County Crime Lab, KS
7. US Army Criminal investigation Lab, Atlanta, GA
8. Texas Department of Public Safety, TX,
9. Tennessee Bureau of Investigation, TN,
10. Laboratory of Forensic Science, Sacramento, CA,
11. Miami Dade Police Department, Miami, FL,
12. Center of Forensic Sciences, Canada
13. Procuraduría General de Justicia, Nuevo León, México,
14. Department of Homeland Security, CBP Research Laboratory, VA,
15. Netherlands Forensic Institute, The Hague, Netherlands,
16. Applied Spectra, CA
17. Michigan State Police - Grand Rapids Forensic Laboratory, MI
18. US EPA – NEIC Laboratory, CO
19. Microtrace LLC, Illinois
20. RCMP-GRC, Ottawa, Canada
21. BAM, Berlin, Germany
22. Photon Machines, Seattle, WA
23. Corning Incorporated, Corning, NY

*This project was supported by Award No. 2009-DN-BX-K252 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. Names of commercial manufacturers are provided for identification purposes only, and inclusion does not imply endorsement of the manufacturer, or its products or services by any of the above organizations. The views expressed are those of the authors and do not necessarily reflect the official policy or position of any of the above organizations 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.*

## Chapter 1. Introduction and Review of the Current Literature

Many research groups have worked to improve the value of trace evidence analyses including the application of mature analytical techniques such as solution inductively coupled plasma optical emission spectroscopy (ICP-OES) [1,3-6,8-9], micro X-ray fluorescence ( $\mu$ XRF) [2], laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) [7,10-17] and, more recently, laser induced breakdown spectroscopy (LIBS) [18-20] for the elemental analysis of the materials glass and paint. Efforts in our research group have resulted in the optimization of both solution ICP-MS analysis of glass [21] and LA-ICP-MS analysis of glass [22,23], paint [16] and soils [24], including a recent review of the literature summarizing the utility of elemental analysis methods for glass analysis [25] and the approval of a “standard method” for the analysis of glass by solution digestion ICP-MS by the ASTM in 2004 [26]. While there have been several reports in the literature successfully using ICP-AES to discriminate glass samples (this is the current method used by the FBI laboratory for casework analysis) and there are many successful reports of coupling laser ablation to ICP-MS instruments, there are, as yet, no publications to our knowledge describing the utility of coupling laser ablation to an ICP-AES instrument to determine the utility of LA-ICP-AES for forensic glass analysis. One of the aims of this research was to evaluate the utility of coupling LA-ICP-AES for forensic glass characterization and comparison. The results of this research are described in the next chapter (chapter 2) of this report.

Several authors and researchers have proposed criteria to use to determine whether or not two samples of glass can be found to be indistinguishable from each other including for the comparison of refractive index values and chemical composition measurements. Experienced glass examiners have recently compiled a list of the possible criteria examiners can choose as follows:

1. **Fixed Difference** – Fixed differences in refractive index or density beyond which a conclusion of two distinct sources is made have been used by many glass examiners for at least 25 years, based on the

suggestion of Miller [27]. Miller suggests a “positive opinion of nonidentity” if flat glass specimens fail to match within the limits of 0.0002 for  $n_D$ , 0.0004 for  $n_C$  or  $n_F$ , or 0.00010 g/cm<sup>3</sup> for density. It is not stated to be a standard or rule but rather “judgement” is to be used in each case. These numbers are not derived from the variability of the broken glass object in question. Instead, they are suggested limits based on an assumption of no measurable variation in modern flat glass and assumed operator and equipment errors. Miller suggests his numbers are only good for modern flat glass. Wider variations, particularly in refractive index, have been measured in headlamp lenses, containers, and in float and tempered flat glasses. A limitation of the fixed difference approach, is that for other parameters, such as element concentrations, fixed difference match criteria have not been established nor recommended. In fact, fixed differences of element concentrations are not appropriate as match criteria, because the measurement precision and source variability both vary as a function of concentration [28]. The fixed difference approach is included here for discussion primarily because of its historical significance. Its use in evaluating glass evidence is *not recommended*.

2. **Fixed sigma ( $\sigma$ )** – In this approach, for each measured parameter, the mean of multiple determinations from a questioned source is compared to the mean of multiple measurements of a known source using as a match criterion, a fixed number of standard deviations, calculated from the measurements (at least three measurements for each questioned and known fragment are required). The match criterion most often used is  $\pm 2\sigma$ , because that corresponds approximately to a 95% confidence level for errors of false exclusion. Rules based on  $1\sigma$ ,  $2.6\sigma$ , and  $3\sigma$  are also discussed in the literature and used by some examiners. A small match criterion, such as  $1\sigma$  has been advocated by some, because it is more conservative in the sense of minimizing the frequency of incorrect associations (although this significantly

increases the number of false exclusion to as much as 32% of the comparisons). Two assumptions that must be met in order to apply the fixed sigma approach are that the measured variables follow Normal distributions and that the limited number of measurements provides an accurate measure of the population variance. A statistical disadvantage of the fixed sigma approach is that the probability of an incorrect exclusion is also a function of the number of replicates. This drawback is addressed in the use of the Students t-test. One advantage of the fixed sigma approach is that it is easily performed by simple observation of the data, without the need for computer calculated statistics (not necessarily a major obstacle in the typical forensic laboratory). The fixed-sigma approach is also not recommended for the reasons mentioned above.

3. **Range overlap** – For each variable, the range of measured values for a questioned sample is compared to the range of measured values for a known sample. If the ranges overlap, the two samples are determined to be indistinguishable. Advantages of this approach are that no assumptions are required concerning the distribution of measured values over the sample population and that it is simple to compare ranges for multivariate data without complicated statistical calculations. For Normally distributed data, the range overlap method results in approximately the same match criteria as a  $2\sigma$  to  $3\sigma$  overlap, depending upon the number of replicates. Range overlap has been used most frequently for comparison of element concentrations [29-30]. The range overlap method can be used for one vs. many and for many vs. many comparisons. The disadvantages of range overlap include the same disadvantages as a fixed sigma approach and when a small number of measurements are reported, the range of values may not adequately characterize the population variance for a given glass object.

4. **Student's t-test** – The Student's t-test, with various modifications, is the most widely used statistical method for comparison of the means of two sample populations [31-33]. In this test, the match criteria is selected such that the probability of an error of false exclusions is fixed at an acceptable level, usually 5%. An assumption required for use of the t-test is that the measured variable has Normal distributions in the two populations being compared. Equal variances are also required for use of the simple t-test. If variances are not equal, then a pooled standard deviation can be used and the Welch's modification equation is used (Microsoft Excel has options for either of these cases and can easily calculate the t-test). Variations of the t-test can be used to test hypotheses of a single population in the cases of one measurement vs. many as well as many vs. many. Single-variate versions of the t-test treat each variable independently in making source comparisons. Independence of measured variables must then be assumed or shown in order to assess the significance of samples that are indistinguishable for multiple variables. The multivariate version of the t-test is known as Hotelling's  $T^2$  and this approach has been recommended by the National Academy of Sciences (NAS) committee on bullet lead analysis for the evaluation of multivariate data derived from the composition of bullet lead, for example [34]. Our research group (and other workers) have used both the successive use of the uni-variate t-test for multiple elements with much success [9,17-18] as well as the use of Hotelling's  $T^2$  for multivariate glass comparisons [33, 35].

The creation of a database of elemental data for over 700 different glass samples has provided background for the interpretation of the significance of trace elemental "matches" [36]. These databases, some of which were created at FIU over the course of the last 10 years can be used as the basis to determine the **significance of a match**, once a match has been determined. Forensic

examiners can then use this information to assist in the interpretation of the value of a “match” by trace elemental content for a given case so that the opinion is ***not overstated or understated***. Glass examiners have also reported the use of LA-ICP-MS in real cases [37-39] and testified in Frye hearings in Miami Florida in 2004 [40] whereby the judge allowed this method for use in a hit and run accident fatality case. While LA-ICP-MS has been accepted as a powerful technique to discriminate between different glass samples through the comparison of the elemental data generated, it is an expensive and sophisticated technique that is out of the reach of many forensic laboratories. This project included the use of a large number of experienced uXRF users in a series of inter-laboratory exercises for the evaluation of glass evidence (round-robin 1-4) and for the elemental analysis of soil samples (round robin 5). The results of the first 4 round robin exercises are summarized in the chapters 3 and 4 of this report and include the conclusions derived from performance tests for the different match criteria used in the round robin exercises. The results from the soil sample round robin exercise are summarized in chapter 5 of this report. Finally, the Elemental Analysis Working Group (EAWG) collaborated on an ASTM method of analysis for the use of uXRF for glass evidence analysis and for the use of LA-ICP-MS for glass evidence analysis. These ASTM methods have been submitted to the E30 Committee on Forensic Sciences for evaluation and voting and are included at the end of this report in the appendix.

## **Chapter 2. Forensic analysis of glass by $\mu$ -XRF, ICP-MS and LA-ICP-MS; Method Standardization**

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 [41]. 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 [42-59] 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 [60]. 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 development of the analytical protocols 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 [52].

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 [47]. This technique is nondestructive of the sample 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 % [60-62]. In addition, the precision is

generally poorer than other methods such as XRF and ICP-based methods [21]. 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 ( $\mu$ -XRF). Emitted x-rays are detected with an energy dispersive detector in  $\mu$ -XRF instruments. The advantages of  $\mu$ -XRF are similar to those of SEM-EDX: it is nondestructive, relatively easy to operate, and provides simultaneous multi-elemental information. However,  $\mu$ -XRF is more sensitive than SEM-EDX especially for elements of energy higher than 3keV providing better discrimination between glasses of the same type [42-44, 63, 64]. Advantages of  $\mu$ -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.

The main drawback to  $\mu$ -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 [52]. Also,  $\mu$ -XRF is not sensitive enough to measure several trace elements that have been shown to have good source discrimination capability [52]. 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 [63]. 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 [44-46,59,65,66].

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 [52]. 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 [45, 49, 67-71].

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 [70]. The resulting digest is aspirated into the plasma for analysis [45, 50, 57, 70]. The major drawbacks to these protocols are that they are rather time-consuming, require the use of hazardous reagents and can introduce contamination into the solution. ICP-MS instruments are normally 1-2 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 [52].

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  $\mu\text{g}$  per replicate) [68, 72-79]. 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.* [48], where the discrimination potential of different techniques such as SEM-EDX,  $\mu$ -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 [76]. 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 chapter, important considerations in analytical method validation for  $\mu$ -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 [80].

## **Materials and Methods**

Thirteen participants reported the first interlaboratory test results. One (1) participant performed acid digestion followed by ICP-MS, five (5) participants conducted the analysis using LA-ICP-MS and seven (7) participants used  $\mu$ -XRF. Sixteen participants reported the second test results. One (1) participant performed acid digestion followed by ICP-MS, six (6) participants conducted the analysis using LA-ICP-MS and nine (9) used  $\mu$ -XRF.

## 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:  $^7\text{Li}$ ,  $^{25}\text{Mg}$ ,  $^{27}\text{Al}$ ,  $^{29}\text{Si}$  (as internal standard),  $^{39}\text{K}$ ,  $^{42}\text{Ca}$ ,  $^{49}\text{Ti}$ ,  $^{55}\text{Mn}$ ,  $^{57}\text{Fe}$ ,  $^{85}\text{Rb}$ ,  $^{88}\text{Sr}$ ,  $^{90}\text{Zr}$ ,  $^{118}\text{Sn}$ ,  $^{137}\text{Ba}$ ,  $^{139}\text{La}$ ,  $^{140}\text{Ce}$ ,  $^{146}\text{Nd}$ ,  $^{178}\text{Hf}$ , and  $^{206,207,208}\text{Pb}$  (reported as total Pb). The participant that conducted acid digestion followed by ICP-MS used the same menu with the exception of  $^{29}\text{Si}$ ,  $^7\text{Li}$ ,  $^{139}\text{La}$ ,  $^{140}\text{Ce}$ ,  $^{146}\text{Nd}$ ,  $^{178}\text{Hf}$  and  $^{206,207,208}\text{Pb}$ . The digestion and ICP-MS method followed the ASTM method E2330 [80].

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

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 pill boxes 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  $\sim 2\text{-}3\text{ cm}^2$  were collected from an area of about  $30\text{cm}^2$  of the inside panel of the windshield. The glass samples were then washed with methanol, nitric acid (0.8M) 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  $\sim 3\text{-}5$  fragments of 3 to 7 mm length were placed in pillboxes and labeled as K1. About 7-10 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 (4) small fragments were also sampled from the bulk area. All fragments were analyzed in 6 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 x 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 and Q1 (August 12, 1998).

A total of three (3) fragments, all of them full thickness ranging from 2 to 7 mm across, were submitted as known samples (K1). Three (3) 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 (0.8M) 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**

Most ICP-participant laboratories processed their TRA signal from laser ablation with GLITTER<sup>TM</sup> 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 was 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 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 6-7 laboratories that performed elemental analysis by ICP-MS or LA-ICP-MS and (2) the “XRF Group” consisted of 7-9 laboratories that conducted elemental analysis by  $\mu$ -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.* [82]. The participant that conducted acid digestion followed the dissolution and calibration methods described in ASTM E2330 [81].

The glass reference materials NIST 1831, FGS 1 and FGS 2 were used to monitor the analytical performance of the methods. 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 to 5. Each of the ICP laboratories made seven replicate sample measurements each on SRM NIST 1831, FGS 1 and FGS 2. The repeatability and reproducibility are calculated as specified in ASTM Practice E 177 [83].

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.

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 to 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.

An exception 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  $^{40}\text{Ar}^{16}\text{O}^1\text{H}^+$ ,  $^{40}\text{Ca}^{16}\text{O}^1\text{H}^+$ ,  $^{41}\text{K}^{16}\text{O}$ ,  $^{40}\text{Ar}^{16}\text{O}^+$ ,  $^{40}\text{Ca}^{16}\text{O}^{1+}$  that compromise the analytical determination of  $^{56}\text{Fe}^+$  and  $^{57}\text{Fe}^+$ . Due to the nature and abundance of these interferences, standard unit resolution ICP-MS instruments cannot measure the most abundant iron isotope  $^{56}\text{Fe}^+$  (91.72 % abundant); therefore, limits of detection for the lower abundant isotope  $^{57}\text{Fe}^+$  (2.2 % abundant) are typically high ( $>10 \mu\text{g g}^{-1}$ ) [84]. 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 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 [85]. The acceptance range for the purposes of this interlaboratory study was defined as the study mean  $\pm$  three (3) times the study standard deviation [85].

Strontium results for FGS 1 are shown in Figure 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 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  $\mu\text{m}$ ) 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 $\mu$ -XRF methods

The  $\mu$ -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  $\mu$ -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:

$$E_{normalized} = \frac{\left[ \frac{1}{n} \sum_{i=1}^n E_i \right]_{sample}}{\left[ \frac{1}{n} \sum_{i=1}^n E_i \right]_{SRM1831}} \quad (\text{Equation 1})$$

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 random propagation of errors.

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

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 $\mu$ -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  $\mu$ -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 3-5) and  $\mu$ -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  $\mu$ -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 (LOD) and/or quantitation limits for some XRF systems, which affected the overall precision. However, most monitored elements in  $\mu$ -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 have 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  $\mu$ -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 [86,87].

The background count level in  $\mu$ -XRF is affected by the sample and uses counting statistics, therefore to estimate the signal-to-noise ratio (SNR), the noise in a  $\mu$ -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.* [88].

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 twenty-one instrumental replicates from the standards NIST 1831, FGS 1 and FGS 2.

ICP-based methods showed superior limits of detection than  $\mu$ -XRF (1-3 orders of magnitude) allowing the analysis of greater number of trace elements. As expected, the LODs for  $\mu$ -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 [52].

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 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 thirteen respondents correctly reported that item 1 (K1) was found to be indistinguishable from item 2 (Q1) based on LA-ICP-MS or  $\mu$ -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,  $\pm 2s$ ,  $\pm 3s$ ,  $\pm 4s$ , modified  $\pm 4s$ , 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  $\mu\text{g g}^{-1}$ , with exception of iron that was present at  $\sim 600 \mu\text{g g}^{-1}$ . 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  $\mu$ -XRF participants towards using spectral overlay and  $\pm 3s$  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  $\mu\text{g g}^{-1}$  (Ti, Mn and Fe).

The results of these two studies demonstrate that each of the evaluated methods (ICP-MS, LA-ICP-MS and  $\mu$ -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 are presented in a separate publication [49].

### **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 have not been reported, to the best of our knowledge, on SRM NIST 1831 [31,32].

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  $\pm 3SD$ ) (see Table 9).

Significant differences were also found between small and full thickness data collected by  $\mu$ -XRF. These differences were expected due to the well-known effects of the take-off angle and critical depth on XRF measurements [44]. 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 4 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 [90] 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 criterion, such as a modified  $\pm 4s$  approach, based on fixed relative standard deviations.

Due to the close precision obtained and reported by most of the ICP-based participants ( $\leq 1-5\%$  RSD), it was observed that 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  $\pm 4s$  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 is included in a separate publication [89].

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  $\pm 4s$  with minimum 3%RSD, reduced the number of false exclusions. Using ANOVA, 18 out of 28 possible comparison pairs were excluded (64%); using  $\pm 4s$  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 were limited to 7 out of 28 possible pairs (25 %).

Second, when using wider criteria (*i.e.*  $\pm 4s$  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 was further studied by the working group and is reported in a separate publication [89].

## 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  $\mu$ -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-ppm in the solid material. Advantages of these methods are that they are fairly standardized among participant laboratories, they are currently used in forensic laboratories and they have been accepted in court. A standardized ASTM method already exists for the digestion and analysis by ICP-MS (ASTM E2330) [34] and the EAWG is currently working on developing a standardized method for LA-ICP-MS. 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  $\mu$ m) 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 is also

using the experience gained from these interlaboratory tests to work towards the standardization of a  $\mu$ -XRF method for the elemental analysis of glass. Limits of detection are 2-3 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 provide 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 [89].

## **Chapter 3. Forensic analysis of glass by $\mu$ -XRF, ICP-MS, LA-ICP-MS and LA-ICP-OES; Evaluation of the performance of different criteria for comparing elemental composition**

### **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 [91-103]. 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 [104]. 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 [105]. Analytical protocols using  $\mu$ -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 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 source. Conversely, a lack of significant differences in the elemental compositions suggests that the samples could have originated from the same 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 been from the same broken glass source [102, 106-113]. The match criteria evaluated in this study are briefly described below.

Spectral overlay is typically one of the first steps used in comparison of  $\mu$ -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 or Er, 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. 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. 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 understand for a layperson and 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 [114].

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 [114-116]. 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.

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 a corresponding significance level can be assigned. Intervals of  $\pm 2$ , 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 [112].

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 [117] and for multi-elemental concentration data [112].

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 [105]. 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.

## **Materials and Methods**

### **Instrumentation and measurement parameters**

Several different instruments were used for the interlaboratory tests. The  $\mu$ XRF, ICP-MS and LA-ICP-MS instruments and analytical parameters used in the first two tests in this study were previously reported [105]. The instrumental parameters for LA-ICP-MS and LA-ICP-OES used by new participants that joined the group for the third and/or fourth tests are shown in Table 1 of this paper. The element list employed by the other participants was described in the first part of this paper. Typical intensity ratios reported by  $\mu$ -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.

The test samples provided to each participant were accompanied by instructions for analysis and reporting depending on their analytical method. The protocol of analysis was standardized for each analytical method to facilitate interlaboratory comparisons [105]. However, each laboratory was allowed some latitude in setting instrumental parameters according to their own optimized method. As the protocols were refined in later tests, several of the individual laboratories made minor changes to their protocols from those shown in the tables.

### **Standards and 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  $\mu$ -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. 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 dried, they 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.

The project consisted of four interlaboratory tests conducted by the EAWG members. The first and second tests were designed to develop and characterize the  $\mu$ -XRF and LA-ICP-MS analytical protocols. The results of these tests are detailed in the first part of this paper. 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. Previous analytical results obtained at FIU for the samples included in these tests are shown in Table 2. Due to space limitations, the data in Table 2 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 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 x 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 the same float line as those labeled sample K1, but they were manufactured 2 years and 8 months before glasses K1 and Q1 (August 12, 1998). 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.

The participants were not told of the sources of the samples for this blind test. They were given only the results of preliminary tests (color, microscopic examination and refractive index) that showed no significant differences between K1 and items Q1 and Q2. 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 ICP-MS, seven participants conducted the analysis using LA-ICP-MS, and seven laboratories used  $\mu$ -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 samples, labeled K1, K2, Q1, Q2, and Q3 were architectural float glass manufactured between April 15, 1998 and August 31, 2001 at the same Cardinal Glass Industries plant as used in the second test (see Table 2). They

were sampled from a 2 x 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.

Each participant was asked to conduct elemental analysis in order to compare K1 and K2 with each of the questioned items. The participants were informed that preliminary screening analysis (color and refractive index) showed no significant differences between K1 and K2 and any of the questioned items, Q1, Q2 and Q3. Participants received no other information concerning the sources of the samples. Participants were instructed to make three measurements on each of the 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 ICP-MS, six conducted the analysis using LA-ICP-MS, and seven used  $\mu$ -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 x 4.3 cm and the other 5.0 x 4.5 cm, which were once part of a single piece of glass. The glass was manufactured at a Pilkington plant (CA, USA) on 03/03/10. The glass items labeled as Q1 originated from glass manufactured in the same plant approximately two weeks before the other samples (02/18/10). 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.

Participants were told that preliminary analysis (color and refractive index) showed no significant differences between K1 and K2 and all of the questioned items (Q1, Q2 and Q3) and were given no other information about the sources of the samples. Each participant was instructed to conduct elemental analysis in order to compare each of the questioned items with K1 and K2 to determine if any of the questioned items could have originated from either K1 or K2. For this test, each Q fragment was to be considered separately, rather than grouping the three fragments as in the previous tests. Additionally, the Q fragments in this test were smaller than in previous tests, making this test more difficult. Seventeen participants submitted results for this test. One laboratory performed acid digestion followed by ICP-MS, eight conducted the analysis using LA-ICP-MS, one used LA-ICP-OES, and seven used  $\mu$ -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 [91-103]. 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  $\mu$ -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 samples. All 15 of the responding participants correctly reported that samples K1 and Q1 were indistinguishable and K1 and Q2 were distinguishable. Results for each analyst with corresponding match criteria and the number of element concentrations or intensity ratios measured are shown in Table 3. As indicated, the participants used several different match criteria 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 2, the concentrations of Ti, Mn, Fe, and Rb are quite different between K1 and Q2. With these results, the protocols for both  $\mu$ -XRF and LA-ICP-MS were considered robust and further, more difficult tests were designed.

### ***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. Samples with similar refractive indices but distinctive elemental compositions were selected for this exercise. Samples were manufactured at the same plant on dates that were weeks, months, and years apart from each other. The dates of manufacture and the elemental profile of each of the samples as recorded in the FIU glass database are shown in Table 2. The mean concentrations shown were obtained following the ASTM method for acid digestion and solution-based ICP-MS analysis (E330-04). The values shown are reported in parts per million ( $\mu\text{g g}^{-1}$ ). Samples manufactured only weeks or months apart have small, but significant differences in their elemental composition, e.g., K1 and Q1. However, most of the differentiating elements are present at low trace levels and it was therefore expected that only the more sensitive methods might detect some of the differences in compositions. The comparison results reported by the participants and the respective match criteria used to arrive at those conclusions are summarized in Table 4.

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 2, 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, manufactured 3 years and 4 months apart and items K1 and Q2, manufactured 3 years and 3 months 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 solution-based 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 2 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 that item K1 was distinguishable from item Q1. These four respondents used LA-ICP-MS methods to arrive at that conclusion. As reported in Table 3, these samples were manufactured at the same plant 2 weeks apart and therefore their elemental compositions are very similar. 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 2 shows that these elements were present in those samples at concentrations ranging from  $<2$  to  $30 \text{ ug g}^{-1}$  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  $2.5 \text{ ug g}^{-1}$ . 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  $\mu$ -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  $\mu$ -XRF measurements and between 6 to 12 elements for the ICP measurements.

Five of the fourteen respondents reported that item K2 was distinguishable from item Q3. None of the seven  $\mu$ -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 solution-based 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  $30 \text{ ug g}^{-1}$  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 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 03/03/10. Sample Q1 was manufactured at the Pilkington plant on 02/18/2010. As shown in Table 2, 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 solution-based 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. The results of comparisons made by the participants in the fourth interlaboratory test are shown in Table 5 for  $\mu$ -XRF methods and Table 6 for ICP methods.

Sixteen of seventeen respondents correctly reported that all of the Q1 fragments were distinguishable from items K1 and K2 (see Tables 5 and 6). 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  $\mu$ -XRF correctly reported that each fragment labeled as item Q2 or Q3 were indistinguishable from both K1 and K2. Therefore, all participants who used  $\mu$ -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  $\mu$ -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.

As shown in Table 6, 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 solution-based 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  $\mu$ -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.

### ***Error rates for $\mu$ -XRF data***

The summary results of error rate analyses obtained using  $\mu$ -XRF data for different match criteria expressed as the percentages of incorrect associations or exclusions are shown in Table 7. 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  $\mu$ -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 4s 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  $\mu$ -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 4s in the fourth test. The number of type 1 errors when using the

narrower match criteria of the t-tests and the 2s test are generally quite high. The range, 3s, 4s, 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  $\mu$ -XRF measurements have repeatability values of 10% or greater for elements present at lower concentrations, such as Ti, Sr, and Zr. The most 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 biases in measured intensities when compared to the data from the larger, multiple K fragments. 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. 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 3s, range overlap, and Hotellings  $T^2$ . As shown in Table 7, 3s 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 3s or Hotellings  $T^2$  criteria for the small irregular shaped fragments encountered in the fourth test. One advantage of  $\mu$ -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  $\mu$ -XRF that has been submitted to ASTM for consideration as a standard test method recommends the use of spectral overlay followed by either a 3s 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  $\mu$ -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.

### ***Error rates for ICP data***

The summary of error rates obtained for ICP data for different match criteria are shown in Table 8. Because of the good precision of most ICP data, additional broader match criteria were included in this study (5s, 6s and modified confidence intervals with minimum of 3% RSD)[15]. 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 5s 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 ( $> 4s$ ). 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 4s and modified 4s interval. Using these broader match criteria reduced significantly the Type 1 errors without sacrificing the capability to discriminate samples (type 2 errors). Hotelling's  $T^2$  is not as practical for ICP data as for the  $\mu$ -XRF data due to the larger number of variables measured (16-18 elements). However, Hotelling's  $T^2$  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.

#### **Homogeneity study: Pilkington and Cardinal plants**

The samples selected for the fourth interlaboratory test originated from a Pilkington glass manufacturing plant that experienced changes in the formulation of the glass due to market requirements. Figure 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 16, 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. 6-7 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***

Figure 2 shows the sampling scheme used to study spatial variation within a single fragment. Five replicate analyses 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 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 ( $<200\mu\text{m}$ ). 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. Figure 3 shows the variability of the iron content observed between different sampling areas of the fragments for one of the Pilkington samples (PK03010) and the Cardinal sample.

The results of this study demonstrate that the heterogeneity between fragments is 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 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  $\mu$ -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 solution 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**

Based on results obtained in the interlaboratory tests, it is concluded that the match criteria for comparison of elemental composition of glass fragments should be carefully selected based on the technique used for analysis as well as the number of replicates that are conducted to characterize the variability of the known and questioned samples.

For  $\mu$ -XRF analysis, the following observations are derived from the studies. Spectral overlay,  $\pm 3s$ , range overlap, and Hotelling's  $T^2$  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  $\mu$ -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  $\mu$ -XRF were able to detect differences in samples manufactured a month apart at the Cardinal plant. 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. Users of  $\mu$ -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 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 [34]. As with  $\mu$ -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  $\mu$ -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.

## Chapter 4. Overall Conclusions

Elemental analysis is a very powerful tool for the identification, characterization and/or differentiation of many man-made materials that could become a critical piece of information in a forensic investigation. The proper assessment of the value of an elemental profile depends on a) a full knowledge of the capabilities and limitations of the analytical technique(s) used for the acquisition of the measurements and b) an understanding of any effect that the nature of the material, its composition and/or its manufacture could have in the overall estimation of the discrimination potential.

In the case of glass, LA-ICP-MS and LIBS methods have been previously developed and optimized hence the main focus of this research was to evaluate the significance of its elemental composition in forensic comparisons. In order to accomplish this, both laser ablation methods were compared to other techniques such as digestion-ICP-MS and uXRF through a series of inter-laboratory studies conducted by 31 forensic examiners representing 23 different laboratories in the US and outside the US (Mexico, Canada and Germany).

The first part of this research describes a series of interlaboratory studies that allowed for a direct comparison between four of the most sensitive methods currently available for the forensic elemental analysis of glass samples (LA-ICP-MS, solution ICP-MS, LIBS and  $\mu$ -XRF). Members of the Elemental Analysis Working Group (EAWG) completed four (4) interlaboratory studies designed and administered by FIU that were specifically on the evaluation of glass evidence. One additional interlaboratory trial was designed to evaluate the bulk elemental analysis of soil evidence. This project enabled the standardization of two commonly used elemental analysis methods (using LA-ICP-MS and uXRF) for glass analysis, the evaluation of the performance of several forensic laboratories and the evaluation of the use of various different "match criteria" for the interpretation of the elemental data generated.

Some of the important findings from the use of the LA-ICP-MS methods for glass analysis include:

- a) The detection and reporting of heterogeneity of Ce and La close to the rim on FGS standards (< 250  $\mu\text{m}$ )
- b) The observation that possible differences between surface and bulk composition in compared glasses may lead to false exclusions if sampling and data interpretation are not carefully considered.
- c) As a result of 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.
- d) These wide 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.
- e) The time interval after 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.

Some important outcomes from the use of  $\mu$ -XRF methods include:

- a) The use of normalized data to a glass standard such as SRM NIST 1831 provide a means to account for differences among instrumental configurations and to permitted the comparison of interlaboratory data.
- b) The use of a glass standard as a “control” glass is recommended to check method performance prior to analysis.
- c) Spectral overlay,  $\pm 3s$ , range overlap, and Hotellings T2 performed well in terms of both false exclusions and false inclusions.
- d) Participants who used  $\mu$ -XRF methods were able to detect significant differences between fragments of glass that were manufactured in the same plant within short periods of time.
- e) The period of time after which samples cannot be distinguished is dependent on the variability of the formulation of the glass within a plant. Users of  $\mu$ -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

Although LIBS is not as mature an analytical method as the other techniques evaluated in this work, the results suggest that LIBS offers potential for the forensic analysis of glass samples. LIBS has analytical capabilities close to those obtained by LA-ICP-MS and better than uXRF. Further interlaboratory optimization and validation of the analytical protocols is believed to be key to improve the agreement of results between laboratories.

In summary, based on results obtained in the interlaboratory tests, it is concluded that the match criteria for comparison of elemental composition of glass fragments should be carefully selected based on the technique used for analysis as well as the number of replicates that are conducted to characterize the variability of the known and questioned samples.

It can be concluded from the glass 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  $\mu$ -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.

The inter-laboratory experience demonstrated to be a very efficient method to validate forensic methods and to assess the significance of the evidence and therefore they are recommended in the future for other matrices such as soil, paint, ink and paper.

## Appendixes

### Appendix A - References

1. Hickman, D, Glass types identified by chemical analysis, *Forensic Science International*, **1986**, 33(1), 23-46.
2. Ryland S: Sheet or container? Forensic glass comparisons with an emphasis on source classification; *J Forensic Sci.* **1986**, 31(4), 1314.
3. Koons, R; Fiedler, C; Rawalt, R, Classification and discrimination of sheet and container glasses by ICP-AES and pattern recognition, *Journal of Forensic Sciences*, **1988**, 33(1), 49.
4. J. Curran, C.Triggs, J.R. Almirall, J. Buckleton and K. Walsh, The interpretation of elemental composition measurements from forensic glass II, *Science and Justice; Journal of the Forensic Science Society*, **1997**, 37:(4) 245-249.
5. J. Curran, C. Triggs, J.R. Almirall, J. Buckleton and K. Walsh, The interpretation of elemental composition measurements from forensic glass , *Science and Justice; Journal of the Forensic Science Society*, **1997**, 37:(4) 241-245.
6. J.R. Almirall, M. Cole, KG Furton and G Gettinby, Discrimination of glass sources using elemental composition and refractive index: Development of predictive models, *Science and Justice; Journal of the Forensic Science Society*, **1998**, 38:(2) 93-101.
7. D.C. Duckworth, C.K. Bayne, S.J. Morton, J.R. Almirall, Analysis of variance in forensic glass analysis by ICP-MS: Variance within the method, *J. of Analytical and Atomic Spectrometry*, **2000**, 15:(7) 821-828.
8. J.R. Almirall, J. Buckleton J. Curran and T. Hicks, Examination of glass, Chapter in *Forensic Interpretation of Glass Evidence*, J. Curran, J. Buckleton and T. Hicks-Champond, Eds., CRC press, Boca Raton, **2000**, 1-24.
9. S Montero, A Hobbs, T. French and JR Almirall, Elemental profiling of glass fragments by ICP-MS as evidence of association: analysis of a case", *Journal of Forensic Sciences*, **2003**, 48(5) 1101-1107.
10. Becker, S; Gunaratnam, L; Hicks, T; Stoecklein, W. and Warman, G, The differentiation of float glass using refractive index and elemental analysis: Comparisons of techniques, *Problems of Forensic Science*, Vol. XLVII, **2001**, 80-92.
11. J.R. Almirall, Elemental analysis of glass fragments, Chapter in *Trace Evidence Analysis and Interpretation: Glass and Paint*, Edited by Brian Caddy, Taylor and Francis, **2001**, 65-83
12. Trejos, T and Almirall, J, Effect of fractionation on the elemental analysis of glass using LA-ICP-MS, *Analytical Chemistry*, **2004**, 76(5) 1236-1242.
13. Trejos, T and Almirall, J, Sampling strategies for the analysis of glass fragments by LA-ICP-MS. Part I: micro-homogeneity study of glass and its application to the interpretation of forensic evidence, *Talanta*, **2005**, 67(2) 388-395.
14. Trejos, T and Almirall, J, Sampling strategies for the analysis of glass fragments by LA-ICP-MS. Part II: sample size and sample shape considerations, *Talanta*, **2005**, 67(2) 396-401.
15. Latzchoczy,C; Dücking, M; Becker, S; Günther, D; Hoogewerff J; Almirall, J; Buscaglia, J; Dobney, A; Koons, R; Montero, S; van der Peyl, G; Stoecklein, W; Watling, J; Zdanowicz, V, Evaluation of a standard method for the quantitative elemental analysis of float glass samples by LA-ICP-MS, *J. of Forensic Sciences*, **2005**, 50 (6), 1327-1341.

16. A Hobbs and JR Almirall, Trace elemental analysis of automotive paints by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS), *Journal of Analytical and Bioanalytical Chemistry*, **2003**, 376: 1265-1271.
17. T Trejos, S Montero, and JR Almirall, Analysis and comparison of glass fragments by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS), *Journal of Analytical and Bioanalytical Chemistry*, **2003**, 376: 1255-1264.
18. B Naes, S Umpierrez, S Ryland, C Barnett and JR Almirall; A Comparison of Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS), Micro X-Ray Fluorescence ( $\mu$ XRF), and Laser Induced Breakdown Spectroscopy (LIBS) for the Discrimination of Automotive Glass, *Spectro. Acta Part B: Atomic Spectroscopy*, **2008**, 63 (10), 1145-1150.
19. Bridge CM, Powell LJ, Steele KL, Williams M, MacInnis JM, Sigman ME. Characterization of automobile float glass with laser-induced breakdown spectroscopy and laser ablation inductively coupled plasma mass spectrometry. *Applied Spectroscopy*, **2006**, 60(10), 1181-1187.
20. Bridge CM, Powell LJ, Steele KL, Sigman ME, *Spectrochim. Acta Part B*, **2007**, 62, 1419.
21. W Castro, T Trejos, B Naes and JR Almirall, Comparison of the Analytical Capabilities of High Resolution Sector Field Inductively Coupled Plasma Mass Spectrometry and Dynamic Reaction Cell Inductively Coupled Plasma Mass Spectrometry for the Forensic Analysis of Iron in Glass Samples, *Analytical and Bioanalytical Chem*, **2008**, 392 (4), 663-672.
22. S Umpierrez , T Trejos, K Neubauer , JR Almirall, Determination of Iron in Glass by Laser Ablation and Solution using a DRC-ICP-MS. *Atomic Spectroscopy*, 27 (3), **2006**, 76-79.
23. K Smith, T Trejos, J Watling, JR Almirall, A Guide for the quantitative elemental analysis of glass using LA-ICP-MS. *Atomic Spectroscopy*, 27 (3), **2006**, 69-75.
24. L Arroyo, T Trejos, P.R. Gardinali, and J.R. Almirall, Optimization and Validation of a LA-ICP-MS Method for the Quantitative Analysis of Soils and Sediments, *Spectrochimica Acta Part B: Atomic Spectroscopy*, **2009**, 64 (1), 14-25.
25. JR Almirall and T Trejos, Advances in the Forensic Analysis of Glass Fragments with a Focus on Refractive Index and Elemental Analysis, *Forensic Science Review*, **2006**, 18(2) 73-96.
26. Standard Test Method for Determination of Trace Elements in Glass Samples Using Inductively Coupled Plasma Mass Spectrometry, E2330-04, ASTM Annual Book of ASTM Standards (**2004**) American Society for Testing and Materials 14.02
27. ET Miller, "Forensic glass comparisons," In *Forensic Science Handbook*, ed. R Saferstein, Prentice-Hal, Englewood Cliffs, NJ, **1982**, pp.139-183.
28. RD Koons and J Buscaglia, "The forensic significance of glass composition and refractive index measurements," *J. Forensic Sciences*, **1999**, 44:496-503.
29. IW Evett and JA Lambert, "The interpretation of refractive index measurements III.," RD Koons, CA Peters, and PS Rebbert, "Comparison of refractive index, energy dispersive x-ray fluorescence and inductively coupled plasma atomic emission spectrometry for forensic characterization of sheet glass fragments," *J. Analytical Atomic Spectrometry*, **1991**, 6:451-456.
30. RD Koons, C Fiedler, and RC Rawalt, "Classification and discrimination of sheet and container glasses by inductively coupled plasma-atomic emission spectrometry and pattern recognition" *J. Forensic Sciences*, **1988**, 33:49-67.
31. JC Miller and JN Miller, *Statistics and Chemometrics for Analytical Chemistry*, 5<sup>th</sup> edition, Prentice Hall, **2005**.

32. KAJ Walsh, CM Triggs, and JS Buckleton, "A practical example of glass interpretation," *Science and Justice*, **1996**, 36: 213-218.
33. JM Curran, CM Triggs, JR Almirall, JS Buckleton, and KJ Walsh, "The interpretation of elemental composition measurements from forensic glass evidence: I," *Science and Justice*, **1997**, 37:241-244.
34. *Forensic Analysis: Weighing Bullet Lead Evidence*, The National Academies Press, **2004**. p.36-38.  
([http://books.nap.edu/openbook.php?record\\_id=10924&page=36](http://books.nap.edu/openbook.php?record_id=10924&page=36))
35. J Curran, C Triggs, JR Almirall, J Buckleton and K Walsh, The interpretation of elemental composition measurements from forensic glass evidence: II, *Science and Justice; Journal of the Forensic Science Society*, **1997**, 37:(4) 245-249.
36. Final Report to the Technical Support Working Group (TSWG) of the Dept. of Defense for grant: "Glass Evidence Database Development" (DAAD05-00-C-7128 FIU), **2003**.
37. J.R. Almirall, Manslaughter caused by a hit-and-run: Glass as evidence of association, Chapter in *Mute Witness; When Trace Evidence Makes the Case*, Max Houck, Ed. Academic Press, **2001**, 139.
38. Personal Communication, Chip Pollock, Sacramento County Forensic Laboratory, Dec. 12, **2008**.
39. Personal Communication, Stefan Becker, BKA, Germany, August 9, **2007**.
40. Personal Communication, John Granoff, Asst. State Attorney of Florida, Dade County, April **2004**.
41. 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, Florida.
42. 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
43. 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
44. 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.
45. Koons RD, Peters CA, Rebert 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
46. 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  $\mu$  fluorescence. *Forensic Sci Int* 137: 107-118
47. Andrasko J, Machly A (**1978**) The Discrimination Between Samples of Window Glass by Combining Physical and Chemical Techniques. *J Forensic Sci* 22: 250-262
48. 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
49. 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
50. 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
  51. 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
  52. Almirall JR, Trejos T (2006) Advances in forensic analysis of glass fragments with a focus on refractive index and elemental analysis. *Forensic Sci Review* 18, 2: 74-96
  53. 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
  54. Hughes JC, Catterick T, Southear G (1976) The quantitative analysis of glass by atomic absorption spectroscopy. *J Forensic Sci* 8: 217-227
  55. Coleman RF, Goode GC (1973) Comparison of glass fragments by neutron activation analysis. *J Radiol Chem* 15: 367- 372
  56. Koons RD, Buscaglia J (2001) Distribution of refractive index values in sheet glasses, *Forensic Science Commun* 3,1 <http://www.fbi.gov/hq/lab/fsc/current/index.htm>.
  57. Catterick THickman DA (1981) The quantitative analysis of glass by inductively coupled plasma-atomic-emission spectrometry: A five-element survey. *Forensic Sci Int* 17: 253-263
  58. Buscaglia J (1994) Elemental analysis of small glass fragments in forensic science. *Anal Chim Acta* 288: 17-24
  59. Ryland S (2011) Discrimination of flat (sheet) glass specimens having similar refractive indices using micro X-ray fluorescence spectrometry. *Journal of the American Society of Trace Evidence Examiners* 2:1:2-12
  60. 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
  61. 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
  62. Krüsemann H (2001) SEMs and Forensic Science. *Problems in Forensic Science XLVII*: 110-121
  63. 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
  64. Ryland S (1986) Sheet or container? Forensic glass comparisons with an emphasis on source classification. *J Forensic Sci* 31: 1314-1329
  65. 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
  66. 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

67. 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
68. 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
69. Zurhaar A, Mullings L (1990) Characterization of forensic glass samples using inductively coupled mass spectrometry. *J Anal At Spectrom* 5: 611-617
70. 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
71. 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
72. Imbert JL, Telouck P (1993) Application of laser ablation ICP-MS to elemental analysis of glasses. *Microchim Acta* 110: 151-160
73. 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
74. 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
75. 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
76. 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
77. 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
78. 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, Proceedings of the International Society for Optical Engineering (SPIE)*, 5071: 193-196
79. 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
80. Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council, (2009) *Strengthening Forensic Science in the United States: A Path Forward*
81. 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, PA
82. 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
83. 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, PA
  84. 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
  85. Miller JN, Miller JC (2000) *Statistics and Chemometrics for Analytical Chemistry*, 4<sup>th</sup> ed., Pearson Education Limited, U.K.
  86. ACS Committee on Environmental Improvement. Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry (1980) *Analytical Chemistry*, 52: 2242-2248
  87. International Union of Pure and Applied Chemistry (IUPAC) (1976) Commission on Spectrochemical and Other Optical Procedures for Analysis. *Pure & Applied Chemistry* 45:99-103.
  88. 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*
  89. Trejos T, *et al* (submitted) Forensic elemental analysis of glass by  $\mu$ -XRF, ICP-MS and LA-ICP-MS – Part II: Evaluation of the performance of different match criteria for elemental composition, *Anal. and Bioanal. Chem.*
  90. 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.
  91. Almirall JR, Buckleton J, Curran JM, Hicks TN (2000) Examination of Glass. In: Curran JM, Hicks TN, Buckleton JS (eds), *Forensic Interpretation of Glass Evidence*, CRC Press, Florida
  92. 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.
  93. 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.
  94. 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.
  95. 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.
  96. 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  $\mu$  fluorescence. *Forensic Sci Int* 137: 107-118.
  97. Andrasko J, Machly A (1978) The Discrimination Between Samples of Window Glass by Combining Physical and Chemical Techniques. *J Forensic Sci* 22: 250-262.

98. 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.
99. 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.
100. 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.
101. 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.
102. Almirall JR, Trejos T (July 2006) Advances in forensic analysis of glass fragments with a focus on refractive index and elemental analysis. *Forensic Sci Review* 18 ( 2): 74-96.
103. Ryland S (2011) Discrimination of flat (sheet) glass specimens having similar refractive indices using micro X-ray fluorescence spectrometry. *Journal of the American Society of Trace Evidence Examiners* 2 (1):2-12.
104. Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council, (2009) *Strengthening Forensic Science in the United States: A Path Forward*.
105. Trejos T, et al (submitted) Forensic elemental analysis of glass by  $\mu$ -XRF, ICP-MS and LA-ICP-MS – Part I: Method standardization, *Anal. and Bioanal. Chem.*
106. 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
107. 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
108. Curran JM, Hicks T, Buckleton JS (2000) *Forensic Interpretation of Glass Evidence*, CRC Press, Boca Raton
109. Curran JM, Triggs CM, Almirall JR, Buckleton JS and Walsh KAJ (1997) The Interpretation of Elemental Composition Measurements from Forensic Glass Evidence: I. *Sci Justice* 37: 241-244
110. Koons RD, Buscaglia J (2002) Interpretation of glass composition measurements: The effects of match criteria on discrimination capability. *J Forensic Sci Soc* 47: 505-512
111. 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
112. 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
113. 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 (11): 1185–1193.
114. Miller JN, Miller JC (2000) *Statistics and Chemometrics for Analytical Chemistry*, fourth ed., Pearson Education Limited, U.K.

115. Johnson RA, Wichern DW (1982) Applied Multivariate Statistical Analysis, third ed, Prentice Hall, New Jersey
116. Aitken CGG, Taroni F (2004) Statistics and the Evaluation of Evidence for Forensic Scientists, second ed., John Wiley & Sons Ltd., Chichester, England
117. Garvin EJ, Koons RDC (2011) Evaluation of match criteria used for the comparison of refractive index of glass fragments. J Foren Sci 56 (2): 491-500.