Evaluation of the Performance of Modern X-Ray Fluorescence Spectrometry Systems for the Forensic Analysis of Glass

Ruthmara Corzo[a], Troy Ernst[b], Joseph Insana[c], Claudia Martinez-Lopez[d], Jodi Webb[e], Emily Haase[e], Peter Weis[f], and Tatiana Trejos[d]*

[a] *National Institute of Standards and Technology, Gaithersburg MD, United States*
[b] *Michigan State Police, Grand Rapids Forensic Laboratory, Grand Rapids MI, United States*
[c] *Microtrace LLC, Elgin IL, United States*
[d] *West Virginia University, Morgantown WV, United States*
[e] *Federal Bureau of Investigation, Quantico VA, United States*
[f] *Bundeskriminalamt, Wiesbaden, Germany*

*Corresponding author: Tatiana Trejos, tatiana.trejos@mail.wvu.edu

Keywords: glass, interlaboratory study, XRF, silicon drift detector (SDD)

Abstract

Micro X-ray Fluorescence Spectrometry (μXRF) is a well-established technique for the elemental analysis of glass in forensic casework. The standard test method for the forensic analysis of glass using μXRF (ASTM E2926) provides recommendations for the number of replicate measurements that should be collected to characterize a known source and the criteria for the comparison between the known and questioned samples. However, these recommendations were based on interlaboratory data collected using μXRF instrumentation equipped with traditional lithium-doped silicon (SiLi) detectors. This interlaboratory study aimed to evaluate the performance of modern μXRF systems equipped with silicon drift detectors (SDDs) for the forensic comparison of glass. While the SDD-μXRF instruments resulted in improved precision and detection limits (1.4 $\mu g \cdot g^{-1}$ – 1386 $\mu g \cdot g^{-1}$) and excellent discrimination (> 99%) of different-source samples, the false exclusion rates for same-source samples were relatively high (> 20%). Two methods were evaluated to reduce the high false exclusion rates: increasing the number of fragments collected for the known source and modifying the recommended comparison criteria. To reduce the false exclusion rate to 5% or less, a minimum of five known fragments were needed. Alternatively, modifying the recommended comparison criterion reduced the false exclusion rate from 23% to 2%, while maintaining low false inclusions (< 1%). The findings in this study demonstrate the improved sensitivity and precision observed in glass measurements acquired with μXRF-SDD systems. However, these systems may require adjustments to sampling and the comparison criteria to minimize potential error rates in the forensic comparison of glass fragments.

1. Introduction

Glass is an important type of forensic trace evidence frequently encountered in crime scenes such as hit-and-runs and burglaries. Elemental analysis of glass has been shown to provide valuable information for the discrimination of glass originating from different sources or the association of glass originating from the same source [1-10]. The elemental analysis of glass in forensic casework is typically accomplished using laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS) or micro X-ray fluorescence spectrometry (μXRF). The former technique, LA-ICP-MS, has the advantage of superior sensitivity and the capability to produce quantitative data, which enables the establishment of shareable databases. However, LA-ICP-MS is costly, partially destructive, typically consuming 0.4 μg to 2 μg of material, and requires matrix-matched standards for quantitative analysis, which are available for glass but may not be available for other matrices [11]. Quantitative analysis is possible with XRF, but it requires extensive sample preparation, which may not be feasible with casework samples. Therefore, element intensity ratios, rather than quantitative data, are used for pairwise comparisons of the known and questioned sources. XRF measurements are also known to be affected by the sample shape (i.e., surface irregularity) and thickness; thus, it is important that the known and questioned samples be of similar shape and thickness for comparison, as described in ASTM E2926 [12]. Despite its lower sensitivity, μXRF provides excellent discrimination of different glass sources [2, 10, 13, 14]. Additionally, because of its non-destructive nature and its lower cost, μXRF is more widely used in forensic laboratories.

The interlaboratory studies conducted by the Elemental Analysis Working Group (EAWG) and Natural Isotopes and Trace Elements in Criminalistics and Environmental Forensics (NITECRIME) European Network resulted in standard test methods for the forensic analysis of glass using LA-ICP-MS (ASTM E2927) and μXRF (ASTM E2926) [1, 2, 11, 12, 15, 16]. ASTM E2926 provides the following recommendations for the analysis of glass using μXRF: a minimum of three replicate measurements should be collected for each questioned fragment; a minimum of nine replicates should be collected for the known source; and a range overlap or a ±3 standard deviation (±3s) interval should be used to compare chemical composition, as characterized by element intensity ratios, of the known and questioned samples. For range overlap, the range of measurements collected on the questioned fragment is compared to the range of measurements collected on the known fragments. For the ±3s interval, the average of the questioned sample measurements is compared to the average ±3 times the standard deviation of the known sample measurements. For either comparison criteria, if there is an overlap in the questioned and known measurements for all element ratios under consideration, the possibility that the known and questioned samples originated from the same source cannot be excluded. If there is no overlap for at least one element ratio, it is concluded that the known and questioned samples originated from different sources.

The current ASTM E2926 recommendations were based on data collected with μXRF systems equipped with a traditional lithium-doped silicon (SiLi) detector. However, recent advancements in μXRF systems have improved the sensitivity of modern instruments. These advancements include higher-intensity poly-capillary X-ray optics and the introduction of silicon drift detectors (SDDs), which offer better resolution and higher signal throughput. The

improved sensitivity of SDD-μXRF instruments merits a reevaluation of the ASTM E2926 recommendations to determine if the current sampling strategies and comparison criteria are fit-for-purpose when conducting glass analyses. Findings from a previous study by Martinez-Lopez et al. suggest that an update to the ASTM E2926 recommendations is warranted [17]. In the study, a set of 100 fragments taken from a single windshield and a set of 13 fragments taken from different vehicles were analyzed with a μXRF system equipped with an SDD. The study reported high false exclusion rates (8% to 17%) when using the ASTM E2926 comparison criteria: range overlap and ±3s interval. The high sensitivity of the μXRF system used in the study led to the detection of microheterogeneity within the same-source glass sample, which ultimately resulted in high false exclusion rates. To address the high false exclusion rate and compensate for the improved sensitivity of the μXRF system, the authors proposed a modification to the ±3s criterion in which *s* is either the measured standard deviation or 3% of the measurement average, whichever is greater. This modification applies a lower limit so that *s* is always at least equal to 3% of the average (equivalent to a minimum 3% relative standard deviation). Applying a minimum relative standard deviation (RSD) is an approach that is currently used for LA-ICP-MS measurements; ASTM E2927 recommends a ±4s interval with a minimum 3% RSD for LA-ICP-MS glass comparisons [11]. In the Martinez-Lopez et al. study, the modified μXRF criterion reduced the false exclusion rate to below 3% without increasing the false inclusion rate. The authors also evaluated the effect of the number of known fragments on the false exclusion rates and found that a minimum of 6 known fragments was needed to reduce the false exclusion rates to below 10% for range overlap and the ±3s interval.

This interlaboratory study expands upon the previous study by including data acquired using four μXRF systems and an expanded glass sample set, which included challenging cases (i.e., samples known to originate from different sources but with similar chemical composition). The interlaboratory study was organized by the glass task group members of the National Institute of Standards and Technology's (NIST) Organization of Scientific Area Committees for Forensic Science (OSAC). The overall aims of this study were to: 1) compare the performance (precision, detection limits, error rates) of SDD-μXRF systems to older systems equipped with SiLi detectors, and 2) determine whether the ASTM E2926 recommendations are fit-for-purpose when using SDD-μXRF systems. Although this study mainly focused on the evaluation of SDD-μXRF systems, two additional labs included measurements collected with LA-ICP-MS and a relatively new technique that has shown utility for glass comparisons, Laser Induced Breakdown Spectroscopy (LIBS) [14, 18-22]. These two techniques served as a cross-reference to better understand the sources of observed variations in elemental profiles.

2. Materials and Methods

2.1. Participating Laboratories and Instrumentation Specifications

Four laboratories provided XRF measurements for the glass collection set included in this interlaboratory study. The instrument specifications and analysis parameters for the four labs, labeled Lab A through D, are provided in Table 1. Each lab followed the recommendations given in ASTM E2926 [12]. One lab's instrument was equipped with a lithium-doped silicon (SiLi) detector. The remaining labs' instruments were equipped with one or two silicon drift detectors (SDD). A fifth lab (Lab E) analyzed the collection set using LA-ICP-MS and LIBS as a comparison

reference for the variation of elemental profiles observed between and within glass samples. Finally, a sixth lab (Lab F) provided LA-ICP-MS for a subset of the samples. Table 2 lists the instrument specifications and analysis parameters for the LA-ICP-MS systems included in this study. The LA-ICP-MS analyses were conducted following the recommendations in ASTM E2927 [11]. The following isotopes were measured: ^{7}Li, ^{23}Na, ^{24}Mg, ^{27}Al, ^{39}K, ^{42}Ca, ^{47}Ti, ^{55}Mn, ^{57}Fe, ^{85}Rb, ^{88}Sr, ^{90}Zr, ^{137}Ba, ^{139}La, ^{140}Ce, ^{146}Nd, ^{178}Hf, and ^{208}Pb. Quantitative analysis was accomplished using a single-point external calibration procedure with the German float glass standard, FGS 2, as the calibrant [16]. All LA-ICP-MS data is provided in the supplementary materials for this manuscript. The Lab E LIBS measurements were collected with the an Applied Spectra J200 equipped with a 266-nm Nd:YAG laser. The parameters for LIBS analysis were as follows: laser fluence 70 J/cm^2, 10 Hz laser repetition rate, single spot mode, 100 µm spot size, 1 µs gate delay, 10 s dwell time, and 1 L/min argon gas flow. The following elements and emission lines were detected (SNR ≥ 3) in the sample collection sets: Fe 274.6 nm, Mg 279.6 nm, Al 309.3 nm, Cu 324.8 nm, Sn 326.2 nm, Ti 334.9, Si 390.6 nm, Ca 393.4 nm, Sr 407.8 nm, Ba 493.4 nm, Li 670.8, K 766.5 nm, and Na 819.5 nm. An element was considered detectable if the signal-to-noise ratio (SNR) was at least 3.

Table 1 *– Specifications for µXRF instrumentation used in this interlaboratory study.*

	Lab A	**Lab B**	**Lab C**	**Lab D**
Manufacturer	Bruker	Bruker	EDAX	EDAX
Model	Tornado	Tornado	Orbis PC	Orbis PC
Capillary Type	Poly	Poly	Poly	Poly
Tube Material	Rh	Rh	Rh	Rh
Beam Power (kV)	50	50	50	50
Beam Current (µA)	600	600	55	500
Capillary Type	Poly	Poly	Poly	Poly
Spot Size (µm)*	20	20	30	30
Acquisition Time (live s)	300	300	1200	1000
Detector Type	SDD	SDD	SiLi	SDD
Number of Detectors	2	2	1	1

** Full Width at Half Maximum (FWHM) spot size at Mo Kα.*

Table 2 *– Specifications for LA-ICP-MS instrumentation used in this interlaboratory study. Note that Lab F acquired data using two different laser ablation and ICP-MS systems.*

	Lab E	**Lab F**
	Laser Ablation System	
Manufacturer	Applied Spectra	ESI
Model	J200	UP213 / NWR193
Laser	266 nm Nd:YAG	213 nm Nd:YAG / 193 nm Nd:YAG
Fluence (J/cm^2)	155	10 / 5
Repetition Rate (Hz)	10	10
Spot Size (µm)	50	80 / 75
Helium Gas Flow (L/min)	0.9	0.8
Ablation Dwell Time (s)	50	60

ICP-MS System		
Manufacturer	Agilent	Agilent
Model	7800	7500cs / 7700x
RF Power (W)	1550	1540
Argon Makeup Gas Flow (L/min)	0.9	0.8
Dwell Time (ms)	10	10

2.2. Sample Set, Preparation, and Analysis

The samples included in this study were vehicle glass provided by the Federal Bureau of Investigation (FBI). A full description of all samples included in the study is provided in Supplementary Table S1. The first lab to receive the samples identified the float side of each fragment using a narrow-band UV lamp and cleaned the sample surfaces using methanol. Using double-sided tape, the fragments were then adhered, float-side down, to a thin film window (Chemplex® Industries Inc.). The thin film window with the loaded fragments was placed directly on the instrument stage for µXRF analysis. Once the first lab completed the analyses, the samples were then sent to another participant. Thus, each lab analyzed the same glass fragments in a Round Robin design. After µXRF, samples were analyzed by LA-ICP-MS and LIBS. The fragments were removed from the thin film window and adhered to a plastic acetate disk using double-sided tape; this was necessary to accommodate the fragments in the laser ablation sample chamber.

The sample set included full-thickness fragments known to originate from the same source and full-thickness fragments known to originate from different sources. The same-source set included 20 fragments from a single vehicle side window (a 2011 Infiniti EX35). Participants were instructed to collect five replicate measurements on each of the 20 fragments, for a total of 100 replicate measurements. The same-source set served to evaluate the within-sample variability and to estimate the false exclusion rates using various comparison criteria. The different-source set included fragments from 25 sources; 5 fragments were provided per source, for a total of 125 fragments. This collection set included side and rear windows from vehicles manufactured between 2009 and 2019 and comprising 16 different vehicle makes and 23 different models. Participants were instructed to collect two replicate measurements per fragment, for a total of 10 replicate measurements for each of the 25 sources and 250 measurements for the entire different-source set overall. The different-source set served to evaluate the between-sample variability and to estimate the discrimination power and false exclusion rates using various comparison criteria. Once the µXRF analyses of the different-source set were completed, it was found that the samples were easily distinguishable because of the substantial differences in their chemical composition. Therefore, an additional set of 17 different-source samples, which were selected based on their similarity of chemical composition, was further analyzed by Lab B. This set had previously been analyzed by Lab F using LA-ICP-MS. The additional different-source set served to evaluate the false exclusion rates for challenging cases. This sub-set included five fragments per source (85 fragments). Two measurements were collected per fragment, for a total of 10 replicates for each of the 17 sources and 170 measurements for the entire sub-set.

One fragment of the National Institute of Standards and Technology (NIST) glass Standard Reference Material®, SRM 1831 was also included in the sample set to serve as a quality control standard. For each analysis day, μXRF participants collected one measurement at the beginning of the analysis, one measurement after every 50 spectra collected on the same-source or different-source fragments, and one measurement at the end of the analysis. The limit of detection (LOD) for each element of interest was calculated using the SRM 1831 data provided by each participant. Thus, the SRM 1831 data set provided a comparison of LODs between instruments equipped with a SiLi detector or an SDD. For LA-ICP-MS and LIBS analyses, four measurements were collected for each quality control (SRM 1831 and SRM 612) after every 50 measurements collected on either the different-source or same-source sample set.

2.3. Data Processing and Analysis

For μXRF and LIBS spectra, background subtraction and peak fitting were implemented using the manufacturer's software (Bruker M4, EDAX Vision, and Applied Spectra Aurora). For LA-ICP-MS, data reduction and quantitation were performed with an in-house R script using RStudio (version 1.0.143) or with GLITTER (version 4, New Wave Research Inc., Fremont CA). The SNR calculations and pairwise comparisons using various comparison criteria were also conducted using RStudio. For μXRF, only elements with an SNR greater than 10 (i.e., above the limit of quantitation, LOQ) were included in pairwise comparisons, as recommended in ASTM E2926 [12]. Plots were generated in Microsoft Excel (version 16.35) or Plot (version 2.6.17). Appendix A provides a worked example for the modified 3s interval in which a minimum 3% RSD is applied.

3. Results and Discussion

3.1. Limits of Detection

To compare the performance of instruments equipped with a SiLi detector versus an SDD, the LOD was calculated for each lab using the SRM 1831 data collected by each lab. The LOD was calculated as follows:

$$LOD = \frac{3 \times c}{SNR}$$

In the equation above, *c* is the concentration of the element of interest. A more detailed description of SNR and LOD calculations for XRF spectra is provided in Ernst et al. [23] and in ASTM E2926 [12]. Briefly, the signal is defined as the sum of the counts for the full width of the element peak and the noise is defined as the square root of the background. The background is estimated using the counts from two regions, one on either side of the element peak, that are free of spectral interferences (i.e., other element peaks or artifact peaks). If available, the NIST certified concentration for SRM 1831 was used for LOD calculations. Otherwise, the consensus concentration reported in ASTM E2927 was used [11]. Table 3 shows the SRM 1831 element concentration used to calculate LOD. For elements with an SNR greater than 10 (above the LOQ), the calculated LOD for each of the four participating labs is shown. Since only Lab C's instrument is equipped with a SiLi detector, Table 3 also includes the average LODs reported in

ASTM E2926 (for mono-capillary or poly-capillary instruments equipped with a SiLi detector). The LODs reported in ASTM E2926 were calculated using SRM 1831 and two German float glass standards (FGS 1 and FGS 2) [16]. As shown in Table 3, the two instruments equipped with two SDDs (Labs A and B) provided the best LODs for most elements, followed by the instrument equipped with one SDD (Lab D), and finally the instruments with a SiLi detector (Lab C and the labs reported in ASTM E2926). Some systems performed better for certain energy regions. For example, Lab D (with one SDD) offered superior sensitivity for elements at the low-energy region such as sodium and magnesium. On the other hand, Labs A and B (with two SDDs) provided superior sensitivity for elements at the high-energy region such as strontium and zirconium. Table 3 shows that the Lab D system was approximately twice as sensitive than the Labs A and B systems for sodium, but approximately half as sensitive for strontium and zirconium. The differences in sensitivity across the energy spectrum can lead to a large variation in LODs for different systems. These differences in LODs, along with the lack of quantitative data, can make it challenging to compare and share data across different systems. However, there are methods to increase the total counts, thereby improving LODs, such as increasing the acquisition time and/or increasing the X-ray beam current. There are also methods to enhance sensitivity in a particular energy region of the spectrum such as adjusting the X-ray beam voltage or applying primary beam filters. Finally, there are normalization methods that may enable the sharing of non-quantitative XRF data across different forensic laboratories.

Table 3 *– Limit of detection (μg·g^{-1}) ± one standard deviation for each participating lab using SRM 1831 (n replicate measurements collected from different locations of the fragment). Bracketed values indicate an SNR < 10 (below the limit of quantitation). The average limit of detection ± one standard deviation (N laboratories) reported in ASTM E2926 [12] for mono-capillary and poly-capillary instruments is also included.*

		Limit of Detection (μg·g^{-1})					
Element	**SRM 1831 Concentration (μg·g^{-1})**	**Lab A (n = 16)**	**Lab B (n = 14)**	**Lab D (n = 22)**	**Lab C (n = 20)**	**ASTM E2926 Mono-Capillary (N = 5)**	**ASTM E2926 Poly-Capillary (N = 4)**
		SDD			**SiLi Detectors**		
Na	98816[a]	1027 ± 17	1386 ± 37	639 ± 10	1411 ± 27	6820 ± 2855	3275 ± 1861
Mg	21200[a]	266 ± 4.7	319 ± 8.1	178 ± 3.3	762 ± 27	1654 ± 662	870 ± 421
Al	6380[a]	118 ± 2.0	155 ± 3.4	122 ± 5.0	SNR < 3	1108 ± 711	517 ± 350
K	2740[a]	15.2 ± 0.13	14.3 ± 0.16	23.7 ± 0.31	56.4 ± 1.45	131 ± 33	40.5 ± 18
Ca	58600[a]	7.1 ± 0.04	7.3 ± 0.06	8.6 ± 0.09	22.2 ± 0.20	56.0 ± 17.0	17.0 ± 0.82
Ti	114[a]	3.0 ± 0.08	3.4 ± 0.06	3.6 ± 0.08	9.6 ± 0.48	23.0 ± 8.2	9.9 ± 0.87
Mn	15.00[b]	1.4 ± 0.08	1.6 ± 0.08	[11.0 ± 4.13]	SNR < 3	14.5 ± 4.5	7.5 ± 0.54
Fe	608[b]	2.1 ± 0.01	2.2 ± 0.02	3.1 ± 0.04	7.4 ± 0.07	11.8 ± 4.0	6.7 ± 0.75
Rb	6.11[b]	1.4 ± 0.12	[2.1 ± 0.79]	[4.9 ± 2.54]	SNR < 3	5.6 ± 2.1	7.6 ± 1.1
Sr	89.12[b]	1.8 ± 0.03	2.0 ± 0.05	4.5 ± 0.23	9.7 ± 0.66	5.5 ± 2.2	8.9 ± 2.0
Zr	43.36[b]	1.5 ± 0.04	1.6 ± 0.05	3.4 ± 0.15	6.4 ± 0.68	4.2 ± 1.6	6.6 ± 1.6

[a] NIST certified concentration

[b] Consensus concentration reported in ASTM E2927 [11]

3.2. Same-Source Set

The relative standard deviation (RSD) was calculated for the following element ratios, selected from the elements detected in the single glass source: Na/Mg, Ca/Mg, Ca/K, Ca/Ti, Ca/Fe, Ti/Fe, Mn/Fe, Fe/Zr, Zn/Fe, Rb/Fe, Sr/Zr, and Ce/Fe. The most intense X-ray emission lines were used for each element (i.e., L-lines for Ce and K-lines for all other elements). The list of ratios includes the six ratios recommended in ASTM E2926: Ca/Mg, Ca/K, Ca/Ti, Ca/Fe, Fe/Zr, and Sr/Zr. For the remaining ratios, elements were paired based on similar X-ray energies to minimize take-off angle effects [12]. For each lab, the RSD was calculated for the 100 replicate measurements collected on the 20 fragments from the single source. Figure 1 shows the RSDs for each lab. The RSDs for most element ratios were below 3%; the exceptions included ratios with Rb, Sr, and/or Zr. Note that the RSD for Rb/Fe is excluded for Lab C because the SNR for Rb was below 10. The RSDs were similar for all SDD instruments, while slightly worse precision was observed for SiLi (notably Fe/Zr, Sr/Zr, Mn/Fe, and Zn/Fe). The higher RSDs for the Lab C Fe/Zr and Sr/Zr ratios resulted from the lower intensities for the Sr and Zr peaks. The Sr and Zr average SNR for Lab C (SNR < 20) were lower than the average SNR for the remaining three labs (SNR ranging from 35 to 110). To illustrate this difference in SNRs, Figure 2 shows an SRM 1831 µXRF spectrum for each lab, with a zoomed-in subset for Fe, Sr, and Zr. Overall, the findings demonstrate that the improved sensitivity of SDD systems resulted in improved precision for some ratios.

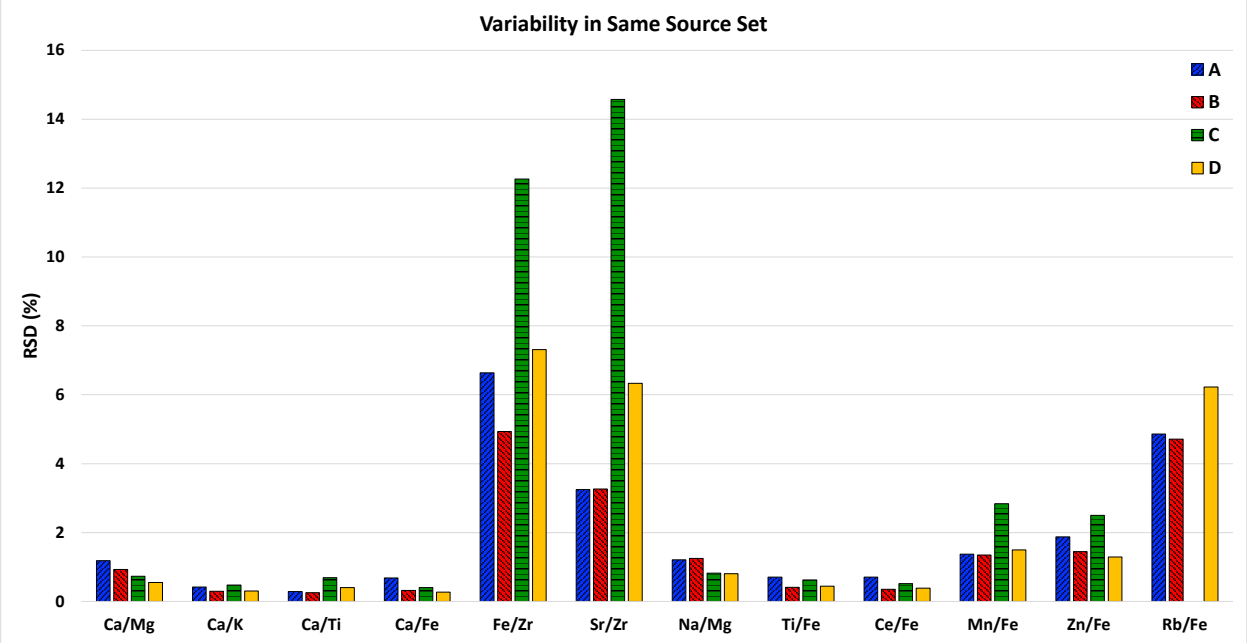

Figure 1 *– Same-source relative standard deviation (100 replicate measurements) for Lab A (blue), Lab B (red), Lab C (green), and Lab D (yellow).*

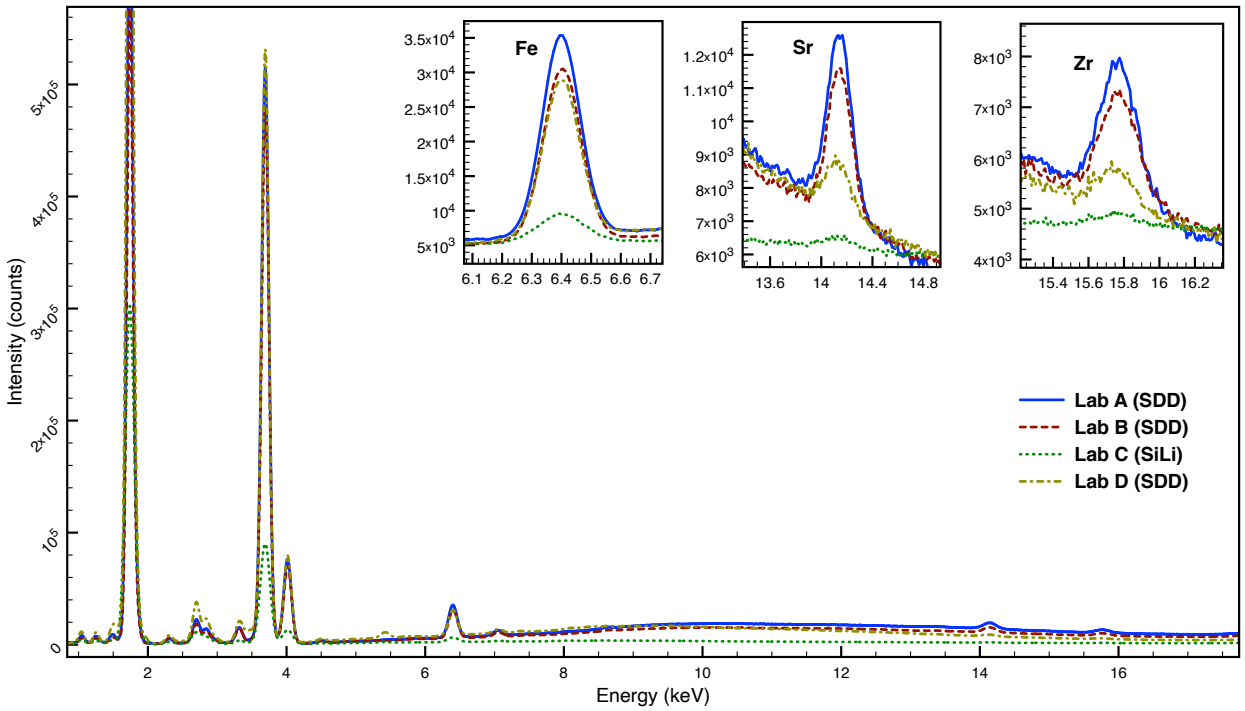

Figure 2 *– µXRF spectra for SRM 1831 with zoomed-in subset for Fe, Sr, and Zr. The spectra in the subset were shifted along the y-axis for better visualization. The detector type for each lab's instrument is indicated in the legend.*

To estimate the false exclusion rate, the following 12 ratios were used for the pairwise comparisons: Ca/Mg, Ca/K, Ca/Ti, Ca/Fe, Fe/Zr, Sr/Zr, Na/Mg, Ti/Fe, Ce/Fe, Mn/Fe, Zn/Fe, and Rb/Fe. As mentioned previously, the Rb/Fe for Lab C was excluded since the SNR for Rb was below 10. ASTM E2926 recommends that a minimum of nine replicates be collected for the known sample to properly characterize the heterogeneity of the known glass source. Since five replicates were collected for each fragment, the false exclusion rates were calculated by comparing the five replicates of one fragment (questioned) to the ten replicates of two randomly selected fragments (known). Each questioned fragment was compared to two randomly selected known fragments 19 times, each time selecting a different combination of two fragments for the known specimen, for a total of 380 comparison pairs. This procedure was implemented to have the same number of comparison pairs if comparing one questioned fragment to one known fragment: $n \times (n\text{-}1)$, where n is the total number of fragments (n = 20). Table 4 shows the false exclusion rate for each lab and overall (all four labs combined) using the ASTM E2926 comparison criteria, range overlap and ±3s. The false exclusions ranged from 16% to as high as 42%. The labs using systems with two SDDs, Labs A and B, produced the highest false exclusion rates.

As a result of the high false exclusion rates observed with the range overlap and the ±3s interval, the following modifications to the ±3s criterion were evaluated: increasing the standard deviation coefficient to ±4s and ±5s; and applying a minimum 3% RSD to the ±3s, ±4s, and ±5s criteria so that the standard deviation is at least equal to 3% of the mean. The latter

approach of applying a minimum RSD (RSD_{min}) is similar to the criterion recommended for LA-ICP-MS measurements in ASTM E2927. Increasing the standard deviation coefficient from 3s to 4s or 5s reduced the false exclusion rates to below 23%. Applying a 3% RSD_{min} to the standard deviation intervals reduced the false exclusion rates to below 5%. Thus, the comparison criteria that included a 3% RSD_{min} produced the lowest false exclusion rates. Although the labs with an SDD-µXRF showed the highest false exclusion rates, the lab with a SiLi-µXRF still showed an improvement in false exclusion rates when a 3% RSD_{min} criterion was applied. Therefore, a modified criterion with an RSD_{min} can be applied to SiLi-µXRF systems.

Table 4 *– False exclusion rate (%) for individual labs and overall (all labs combined) using various comparison criteria with a 1-to-2 fragment comparison scheme. RSD_{min} = minimum relative standard deviation. The number of pairwise comparisons, n, is indicated in parentheses.*

Comparison Criterion	Lab A (*n* = 380)	Lab B (*n* = 380)	Lab C (*n* = 380)	Lab D (*n* = 380)	Overall (*n* = 1520)
Range Overlap*	42.1	36.8	23.7	28.2	32.6
± 3s*	31.1	22.9	16.3	22.9	23.2
± 4s	22.9	15.0	4.7	15.3	14.4
± 5s	17.1	7.1	2.4	10.5	9.2
± 3s, 3% RSD_{min}	4.7	3.2	0.8	0.3	2.2
± 4s, 3% RSD_{min}	2.4	0.5	0.3	0	0.7
± 5s, 3% RSD_{min}	0.5	0.3	0.3	0	0.2

**Comparison criterion recommended in ASTM E2926 [12]*

To determine which element ratios contributed to the high false exclusion rate, the relative false exclusion rates (i.e., the false exclusion for each individual ratio) were also calculated. Of the 12 ratios included in the pairwise comparisons, Ca/Fe led to the highest number of false exclusions, resulting in a relative false exclusion of 28% and 19% for range overlap and the ±3s interval, respectively. Other ratios with false exclusion > 5% included Ce/Fe, Ca/K and Ti/Fe. Note that these ratios all had RSDs < 1% (Figure 1), which indicates a narrow range and standard deviation interval for the pairwise comparisons. Widening the comparison interval to ±4s or ±5s reduced the relative false exclusion to < 13% for all element ratios. Finally, applying a 3% RSD_{min} further reduced the relative false exclusion to < 3% for all ratios. Figure 3 shows the overall (all labs) relative false exclusion for each element ratio using range overlap, a ±3s interval, and a ±3s interval with a 3% RSD_{min}.

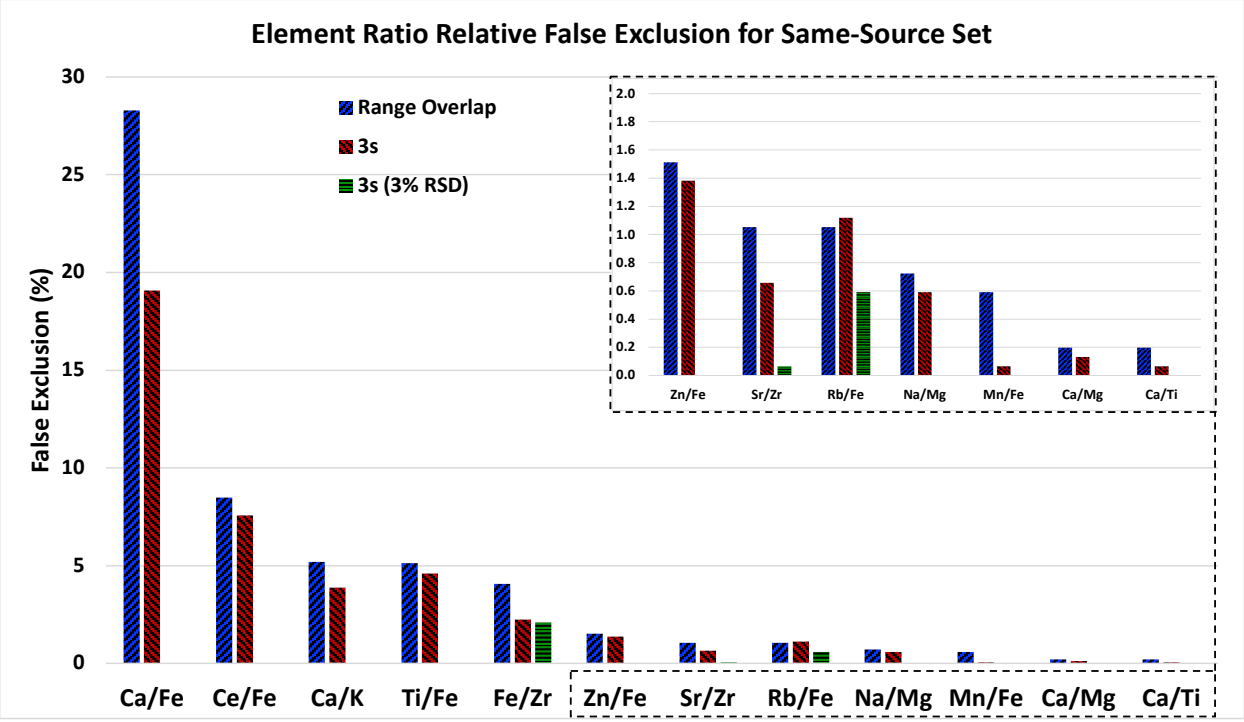

Figure 3 *– Overall relative false exclusion (%) for each element ratio using range overlap (blue), a ±3s interval (red), and a ±3s interval with a minimum 3% RSD (green) for the same-source set. The inset shows a close-up for ratios with relative false exclusions below 2%.*

To illustrate the effect of the number of known fragments analyzed on the false exclusion rate, the false exclusion rate was calculated using a varying number of known fragments (from two to ten fragments) [17]. Figure 4 shows the overall false exclusion rate with an increasing number of known fragments using range overlap, a ±3s interval, and a ±3s interval with a 3% RSD_{min}. As with the 1-to-2 comparison scheme discussed above, the known fragments were randomly selected, and each questioned fragment was compared to the group of known fragments 19 times to ensure the same number of comparison pairs. As illustrated in Figure 4, the false exclusion rate starts quite high (> 20%) for range overlap and the ±3s interval when two fragments (10 replicates) are used for the known sample. To decrease the false exclusion

rate to ≤ 5%, a minimum of five fragments for the ±3s interval and seven fragments for range overlap were needed. For both range overlap and the ±3s interval, increasing the number of known fragments resulted in improvements in the false exclusion rate for all four systems regardless of the detection system (i.e., SDD or SiLi detector). On the other hand, the false exclusion rate for the ±3s with a 3% RSD_{min} begins below 3% for two known fragments and does not change appreciably when additional fragments are grouped together for the known source. Thus, two methods can be used to improve the false exclusion rate: a larger number of known fragments can be sampled or an RSD_{min} can be applied to pairwise comparisons. Increasing the number of known fragments better accounts for the inherent heterogeneity of the known source but increases the analysis time in casework. Notably, applying a 3% RSD_{min} interval resulted in lower false exclusions compared to range overlap and ±3s, regardless of the number of known fragments.

As a cross-reference, the same-source set was also analyzed by another standard method, LA-ICP-MS (ASTM-E2927), and a relatively newer technique that has shown utility for glass comparisons (i.e., LIBS). LA-ICP-MS provides accurate quantitative analysis and, therefore, the concentrations of glass elemental signatures were used in this study to gain a deeper understanding of the observed variations in elemental compositions. For instance, larger measurement variations on μXRF data were observed for elements that LA-ICP-MS determined to be present at lower concentrations. For LA-ICP-MS, quantitative data collected for the 17 isotopes listed in ASTM E2927 was used for the pairwise comparisons. E2927 states that three fragments should be analyzed for the known source, with a minimum of three replicates per fragment. Therefore, a 1-to-3 fragment comparison scheme was used with the recommended E2927 comparison criterion, ±4s with a 3% RSD_{min}, which resulted in a false exclusion rate of 0.8%. Pairwise comparisons for LIBS were conducted using the SNR (element peak area divided by the noise) for the 13 elements detected: Fe, Mg, Al, Cu, Sn, Ti, Si, Ca, Sr, Ba, Li, K, and Na. There is currently no standard test method for the analysis of glass using LIBS. However, a previous publication reported that the ±4s with a 3% RSD_{min} provided the best compromise between false exclusion and inclusion rates [17]. With this comparison criterion and a 1-to-3 fragment comparison scheme, LIBS resulted in a false exclusion rate of 6.6%. Thus, when using the ASTM E2926 comparison criteria (range overlap or ±3s), μXRF resulted in relatively high false exclusion rates compared to either LIBS or LA-ICP-MS. Modifying the ±3s by applying a 3% RSD_{min} resulted in a more reasonable false exclusion rate that fell between LIBS and LA-ICP-MS.

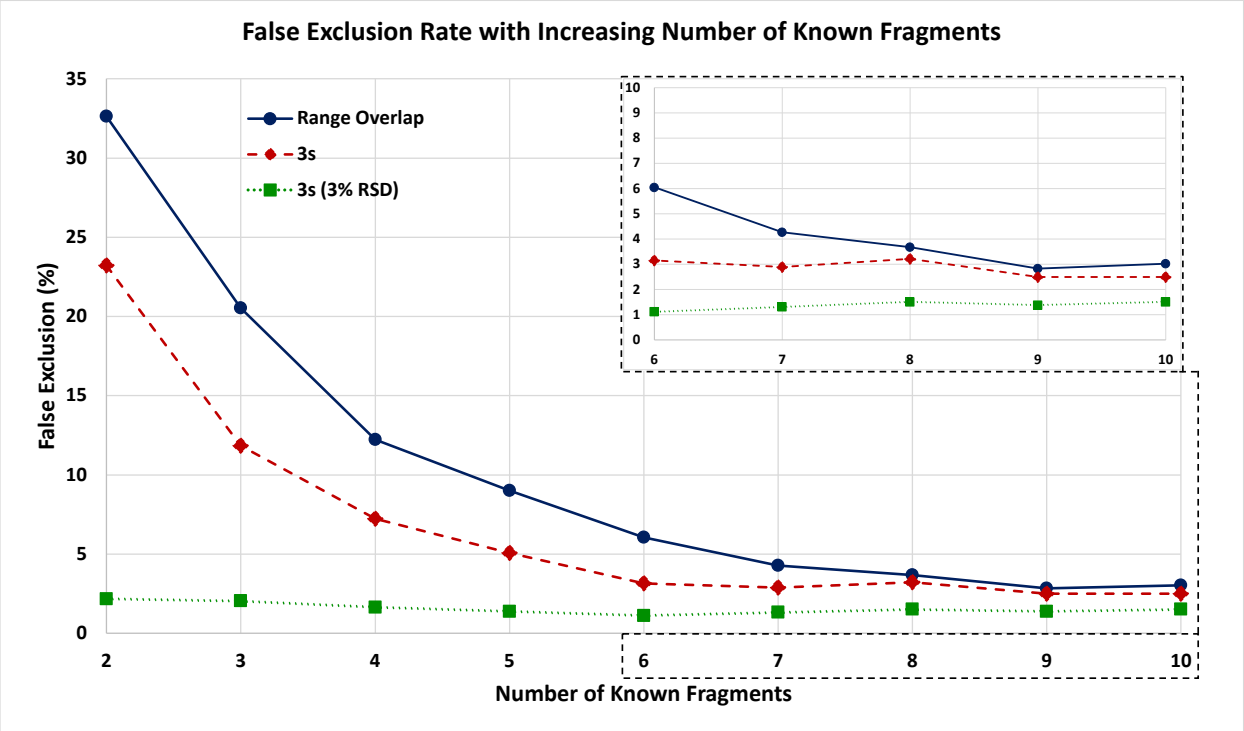

Figure 4 *– Overall false exclusion rate with an increasing number of known fragments using range overlap (blue circles), a ±3s interval (red diamonds), and a ±3s interval with a minimum 3% RSD (green squares). The inset shows a close-up of the false exclusion rate for six to ten known fragments.*

3.3. Different-Source Set

To estimate the false inclusion rate, the following 15 ratios were used for the pairwise comparisons: Na/Mg, Ca/Mg, Ca/K, Ca/Ti, Ca/Fe, Ti/Fe, Mn/Fe, Fe/Co, Fe/Zr, Ni/Fe, Zn/Fe, Se/Fe, Rb/Fe, Sr/Zr, and Ce/Fe. These ratios were selected based on the detected elements (SNR > 3) in the entire set of 25 different-source samples. As with the same-source set, the six ratios recommended in ASTM E2926 were included and additional ratios were selected based on similar X-ray energies to minimize take-off angle effects. The improved detection limits of instruments equipped with one or more SDDs (Labs A, B, and D) resulted in the detection of certain elements (i.e., Se and Rb) that were not detectable with Lab C's instrument, which is equipped with a SiLi detector. Only ratios with elements with SNR > 10 were included in the pairwise comparisons. All five fragments (10 replicate measurements) of each source were compared to the five fragments of every other source, for a total of 600 pairwise comparisons for each lab and 2400 pairwise comparisons overall. Unlike the other comparison criteria included in this study, range overlap is a symmetrical criterion; that is, the results of a pairwise comparison are the same regardless of which sample is treated as the known or questioned sample. Therefore, each comparison pair is evaluated once, resulting in half as many comparison pairs for range overlap (n = 300) as for the other criteria. Table 5 shows the false inclusion rate for each lab and overall using the various comparison criteria. As can be seen, there were no false inclusions for most criteria (i.e., 100% discrimination). Only the ±5s interval with a 3% RSD_{min} for Lab C resulted in a false inclusion for one sample pair (a 2014 BMW versus a 2014 Maserati), which led to an overall false inclusion rate of 0.08%. Therefore, widening the

ASTM E2926 ±3s comparison criterion by applying a 3% RSD_{min} did not adversely affect the false inclusion rate or the discrimination power for the SDD-µXRF systems or the SiLi-µXRF system.

Table 5 *– False inclusion rate (%) for individual labs and overall (all labs combined) using various comparison criteria. The false inclusion rate for the challenging set and combined set (analyzed by Lab B) are also provided. RSD_{min} = minimum relative standard deviation. The number of pairwise comparisons, n, is indicated in parentheses.*

Comparison Criterion	Lab A (*n* = 600)	Lab B (*n* = 600)	Lab C (*n* = 600)	Lab D (*n* = 600)	Overall (*n* = 2400)	Lab B Challenging Set[a] (*n* = 272)	Lab B All Sets Combined (*n* = 1722)
Range Overlap[*b]	0	0	0	0	0	1.5 %	0.2 %
± 3s[*]	0	0	0	0	0	0.4 %	0.06 %
± 4s	0	0	0	0	0	1.1 %	0.2 %
± 5s	0	0	0	0	0	1.5 %	0.3 %
± 3s, 3% RSD_{min}	0	0	0	0	0	2.6 %	0.5 %
± 4s, 3% RSD_{min}	0	0	0	0	0	2.9 %	0.6 %
± 5s, 3% RSD_{min}	0	0	0.3 %	0	0.08 %	3.7 %	0.8 %

[*]Comparison criterion recommended in ASTM E2926 [12]

[a]The challenging set included samples with similar chemical composition (determined with LA-ICP-MS)

[b]The number of pairwise comparisons, n, for range overlap is half that of the other comparison criteria

Table 5 also includes the false inclusion rates for the more challenging set of 17 different-source samples analyzed by Lab B. All five fragments (10 replicate measurements) of each source were compared to the five fragments of every other source, for a total of 136 comparisons for range overlap and 272 comparisons for all other criteria. The 16 ratios listed for the original different-source set and one additional ratio (Fe/Bi) were used for the pairwise comparisons. The Fe/Bi ratio was not included in the original different-source comparisons because Bi was not detected in any of the 25 samples. As expected, the false inclusion rates were higher for the challenging set, ranging from 0.4% to 3.7% depending on the comparison criterion applied. However, since this set included samples that were intentionally selected based on their similar chemical composition, these false inclusion rates would likely be lower in a larger population of glass. The original and challenging sample sets for Lab B were combined to provide a more realistic measure of the false inclusion rates. The combined set of 42 samples resulted in 1722 comparison pairs (861 pairs for range overlap). The false inclusion rates for the combined set were below 1% for all comparison criteria (Table 5). Most false inclusions were comparisons between glass panes from vehicles of the same make. It is worth noting that, in casework, multiple physicochemical properties (e.g., thickness, refractive index, and chemical composition) are used in conjunction for the comparison between a known and questioned sample. When including thickness and refractive index measurements, provided by the FBI, for comparisons within the combined set, only one pair was indistinguishable (both samples originated from a 2009 Honda Accord). This pair was indistinguishable by thickness, refractive index, range overlap, and the modified ±3s interval with an RSD_{min}; however, the pair was distinguishable using the ±3s interval without an RSD_{min} based on Ca/Ti, Ca/Fe, Sr/Zr, and Ti/Fe. Supplementary Table S2 lists all false inclusions for the combined different-source set; the table lists the pairwise results using chemical composition (i.e., µXRF element ratios), refractive index measurements, and thickness measurements. Again, LA-ICP-MS data was used to assist the understanding of quantitative similarities and differences in the datasets. False inclusions observed by µXRF or LIBS, but not by LA-ICP-MS, were caused by elements close to or below the LOQ or LOD for these methods, but quantifiable by LA-ICP-MS.

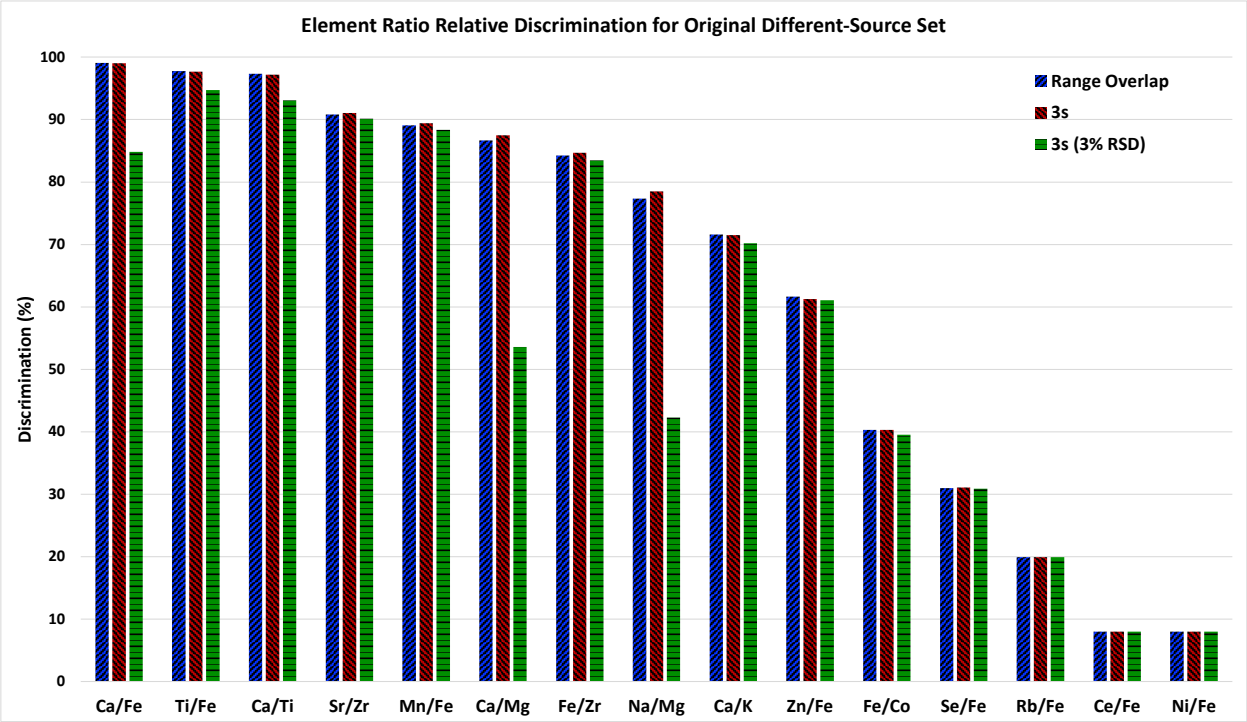

Figure 5 *– Overall relative discrimination (%) for each element ratio using range overlap (blue), a ±3s interval (red), and a ±3s interval with a minimum 3% RSD (green) for the original different-source set. Note that the challenging set is not included here.*

Figure 5 and Figure 6 show the relative discrimination of each individual element ratio for the original and the combined different-source set, respectively. The two figures show the results using range overlap, a ±3s interval, and a ±3s interval with a 3% RSD_{min}. For both sets, Ca/Fe, Ti/Fe, Ca/Ti, and Sr/Zr were the most discriminating ratios, each providing > 90% discrimination (except Ca/Fe when applying a 3% RSD_{min}). Ce/Fe, Ni/Fe, and Fe/Bi were the least discriminating ratios; however, Ni and Bi were above the LOQ in only one sample, while Ce was above the LOQ in four samples. It is worth noting that the low relative discrimination for ratios with elements that are typically below the LOQ (e.g., Ce, Ni, Bi) is not an indication that these ratios are not useful discriminators; since these elements are less commonly encountered, their presence above the LOQ is a discriminating feature. In fact, samples with Ce, Ni, or Bi above the LOQ were distinguishable from all other samples within the different-source set based on the ratios Ce/Fe, Ni/Fe, or Fe/Bi. For the original and combined different-source sets, the three comparison criteria provided similar relative discrimination for each element ratio. The notable exceptions included Ca/Fe, Ca/Mg, and Na/Mg, which showed a lower relative discrimination when a 3% RSD_{min} was applied to the ±3s interval. However, despite the lower relative discrimination with the 3% RSD_{min}, the modified ±3s interval still resulted in an excellent overall discrimination power: 100% discrimination for the original set, > 99% for the combined set, and > 97% for the challenging set (Table 5).

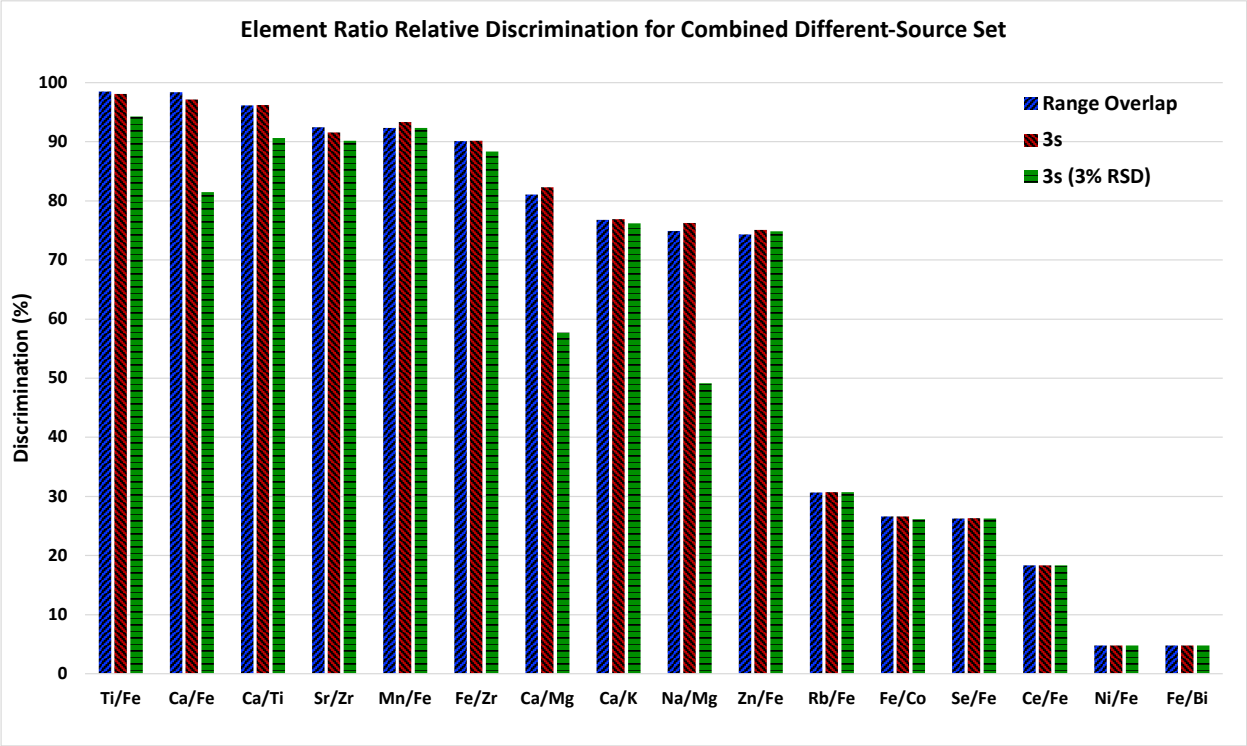

Figure 6 *– Overall relative discrimination (%) for each element ratio using range overlap (blue), a ±3s interval (red), and a ±3s interval with a minimum 3% RSD (green) for the Lab B combined different-source set. Note that this figure includes pairwise comparisons from the original set (shown in Figure 5) and the challenging set.*

The original different-source set was also analyzed with LA-ICP-MS and LIBS (Lab E), while the challenging set was analyzed with LA-ICP-MS (Lab F). Using the ±4s with a 3% RSD_{min} interval recommended in ASTM E2927, LA-ICP-MS resulted in 100% discrimination (no false inclusions) for both different-source sets. LIBS provided 98.8% discrimination (1.2% false inclusion rate) for the original different-source set using the same comparison interval. For the original different-source set, µXRF resulted in 100% discrimination using all comparison criteria except the ±5s with a 3% RSD_{min}. Thus, µXRF performed equally well to LA-ICP-MS and better than LIBS when analyzing this sample set. However, as the sources became more similar in elemental composition (combined different-source set), LA-ICP-MS provided superior performance: 100% discrimination using LA-ICPMS and > 99% using µXRF.

4. Conclusions

Four laboratories participated in an interlaboratory study to compare the performance (limits of detection and error rates) of µXRF instruments equipped with SDDs versus traditional SiLi detectors for the analysis of glass. Additionally, two labs provided LA-ICP-MS measurements and one lab provided LIBS measurements for the same glass sample sets. The LA-ICP-MS and LIBS data provided cross-verification and added other reference points to understand sources of variations in elemental profiles. LA-ICP-MS has the advantage of superior sensitivity and the capability of quantitative analysis, without the XRF disadvantage of sample shape effects. Therefore, the LA-ICP-MS quantitative information was used as the ground truth for the elemental composition of the datasets.

SRM 1831 was analyzed as a quality control sample and to estimate the limits of detection for each instrument. The instruments equipped with one or two SDDs showed improved detection limits, which increased the number of detectable elements. The two instruments with two SDDs provided the best LODs for most elements, with LODs ranging from 1.4 $\mu g \cdot g^{-1}$ to 1386 $\mu g \cdot g^{-1}$. The instrument with a single SDD provided LODs between 3.1 $\mu g \cdot g^{-1}$ and 631 $\mu g \cdot g^{-}1$. Finally, the instrument with a SiLi detector provided LODs between 6.4 $\mu g \cdot g^{-1}$ and 1411 $\mu g \cdot g^{-}1$. The precision, measured as the RSD for each element ratio of interest, was similar across all four labs, except for those ratios of elements where notably better sensitivity was obtained with SDD detectors (Fe, Sr, Zr).

Twenty fragments from a single glass source were analyzed to evaluate precision and the false exclusion rate (*n* = 380 pairwise comparisons). Five fragments from each of 25 different glass sources were analyzed to evaluate the discrimination and false inclusion rate (n = 136 to 1722 comparisons). This study found that the ASTM E2926 comparison criteria developed for µXRF systems are no longer appropriate for newer, more sensitive, µXRF systems equipped with SDDs and high-intensity X-ray optics. ASTM E2926 recommends that a minimum of nine replicate measurements be collected for the known source to properly characterize the microheterogeneity. However, the standard test method does not provide a recommendation for the minimum number of fragments that should be collected for the known source. When using a one questioned fragment to 2 known fragments (10 replicates for the known source) comparison scheme for the same-source set, the overall false exclusion rate was 33% for range overlap and 23% for the ±3s interval. To improve the false exclusion rate, two approaches were evaluated: applying a minimum relative standard deviation (RSD_{min}) to the ±3s interval or analyzing a greater number of known fragments. Applying a 3% RSD_{min} to the ±3s interval reduced the overall false exclusion rate from 23% to 2%. Alternatively, the overall false exclusion rate dropped to ≤ 5% when five or more known fragments were used for the ±3s interval comparisons and seven or more known fragments were used for the range overlap comparisons. Although increasing the number of known fragments better accounts for the inherent heterogeneity of the known source, it increases the analysis time in casework.

For the different-source set, > 99% discrimination (< 1% false inclusions) was observed, regardless of the criterion used, for the original different-source set and the combined set that included challenging cases (pairs with similar chemical composition). Thus, widening the ASTM E2926 ±3s interval by applying a 3% RSD_{min} did not significantly affect the discrimination power for these sample sets. Laser ablation methods allowed a reference point to assess their relative performance compared to µXRF measurements and evaluate the elemental compositions detected by each method.

ASTM E2926 currently recommends six element ratios for pairwise comparisons: Ca/Mg, Ca/K, Ca/Ti, Ca/Fe, Fe/Zr, and Sr/Zr. The six ratios were selected because they were found to be the best discriminators in a 2009 interlaboratory study conducted by the Elemental Analysis Working Group (EAWG). For the different-source sets evaluated in this study, an additional ten ratios were included: Na/Mg, Ti/Fe, Mn/Fe, Fe/Co, Ni/Fe, Zn/Fe, Se/Fe, Rb/Fe, Ce/Fe, and Fe/Bi.

Only elements that were above the LOQ (i.e., SNR > 10) were considered for pairwise comparisons. The elements in the numerator and denominator of the ratio were paired based on their similar X-ray energy to minimize take-off angle effects. Of the additional ten ratios, two provided good discrimination (> 80% relative discrimination): Ti/Fe and Mn/Fe. Two additional ratios provided moderate discrimination (60% to 80% relative discrimination): Na/Mg and Zn/Fe. Finally, the remaining six ratios provided < 50% relative discrimination: Fe/Co, Se/Fe, Rb/Fe, Ce/Fe, Ni/Fe, and Fe/Bi. The relatively low discrimination for these six ratios was because Co, Se, Rb, Ce, Ni, and Bi are not commonly present above the LOQ. However, these elements may still be useful discriminators. For example, the two samples with Ni or Bi above the LOQ were discriminated from all other samples based on the presence of either element alone, demonstrating the utility of less commonly encountered elements for discrimination.

Despite the improved detection limits for XRF instruments equipped with one more SDDs, the lab with a SiLi detector provided the same discrimination power as the three labs with SDDs. However, improved detection limits can be beneficial to better detect differences in samples with very similar compositions. The high overall false exclusions reported in this interlaboratory study indicate that an update to ASTM E2926 is warranted to account for the advancements in SDD-µXRF instrumentation. This can be accomplished by recommending an RSD_{min} for pairwise comparisons and/or recommending a greater number of known fragments be analyzed to better account for within-sample variation. Although the false exclusion rates were higher for the SDD-µXRF systems compared to the SiLi-µXRF system, the modified ±3s with a 3% RSD_{min} still provided an improvement in false exclusions without adversely affecting the false inclusion rate for the SiLi-µXRF system. Thus, the recommendations provided in this study can apply to polycapillary SDD- or SiLi-µXRF systems. Finally, the list of recommended ratios in ASTM E2926 can be expanded to include ratios that showed utility as discriminators (e.g., Ti/Fe, Mn/Fe, Na/Mg, Zn/Fe). However, as noted in ASTM E2926, the list of recommended ratios does not preclude the use of additional ratios for discrimination. As discussed, less commonly encountered elements (e.g., Co, Se, Rb, Ce, Ni, and Bi) can be useful discriminators. Expanding the list of recommended ratios can encourage practitioners to use more ratios for pairwise comparisons, which can ultimately improve discrimination.

Acknowledgements
The authors would like to thank Oriana Ovide and Lauryn Alexander for their assistance in LIBS data acquisition. Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose. This is publication 22.20 of the FBI Laboratory Division. 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 the FBI. The views expressed are those of the author(s) and do not necessarily reflect the official policy or position of the FBI or the U.S. Government.

References

[1] T. Trejos, R. Koons, S. Becker, T. Berman, J. Buscaglia, M. Duecking, T. Eckert-Lumsdon, T. Ernst, C. Hanlon, A. Heydon, K. Mooney, R. Nelson, K. Olsson, C. Palenik, E. C. Pollock, D. Rudell, S. Ryland, A. Tarifa, M. Valadez, P. Weis and J. Almirall. Cross-validation and evaluation of the performance of methods for the elemental analysis of forensic glass by μ-XRF, ICP-MS, and LA-ICP-MS, Anal. Bioanal. Chem. 405 (2013) 5393-409.

[2] T. Trejos, R. Koons, P. Weis, S. Becker, T. Berman, C. Dalpe, M. Duecking, J. Buscaglia, T. Eckert-Lumsdon, T. Ernst, C. Hanlon, A. Heydon, K. Mooney, R. Nelson, K. Olsson, E. Schenk, C. Palenik, E. C. Pollock, D. Rudell, S. Ryland, A. Tarifa, M. Valadez, A. van Es, V. Zdanowicz and J. Almirall. Forensic analysis of glass by μ-XRF, SN-ICP-MS, LA-ICP-MS and LA-ICP-OES: evaluation of the performance of different criteria for comparing elemental composition, J. Anal. At. Spectrom. 28 (2013) 1270-1282.

[3] T. Trejos, S. Montero and J. R. Almirall. Analysis and comparison of glass fragments by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) and ICP-MS, Anal. Bioanal. Chem. 376 (2003) 1255-64.

[4] R. D. Koons and J. Buscaglia. Interpretation of Glass Composition Measurements: The Effects of Match Criteria on Discrimination Capability, J. Forensic Sci. 47 (2002) 505-512.

[5] R. D. Koons and J. Buscaglia. The Forensic Significance of Glass Composition and Refractive Index Measurements, J. Forensic Sci. 44 (1999) 496-503.

[6] J. Buscaglia. Elemental analysis of small glass fragments in forensic science, Anal. Chim. Acta 288 (1994) 17-24.

[7] R. Corzo, T. Hoffman, P. Weis, J. Franco-Pedroso, D. Ramos and J. Almirall. The Use of LA-ICP-MS Databases to Estimate Likelihood Ratios for the Forensic Analysis of Glass Evidence, Talanta 186 (2018) 655-661.

[8] T. Hoffman, R. Corzo, P. Weis, E. Pollock, A. van Es, W. Wiarda, A. Stryjnik, H. Dorn, A. Heydon, E. Hoise, S. Le Franc, X. Huifang, B. Pena, T. Scholz, J. Gonzalez and J. Almirall. An Inter-Laboratory Evaluation of LA-ICP-MS Analysis of Glass and the Use of a Database for the Interpretation of Glass Evidence, Forensic Chemistry 11 (2018) 65-76.

[9] A. Akmeemana, P. Weis, R. Corzo, D. Ramos, P. Zoon, T. Trejos, T. Ernst, C. Pollock, E. Bakowska, C. Neumann and J. Almirall. Interpretation of chemical data from glass analysis for forensic purposes, Journal of Chemometrics 35 (2021)

[10] R. D. Koons, C. A. Peters and P. S. 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. Anal. At. Spectrom. 6 (1991)

[11] ASTM E2927-16e1, ASTM International, West Conshohocken, PA (2016).

[12] ASTM E2926-17, ASTM International, West Conshohocken, PA (2017).

[13] S. Ryland. Discrimination of Flat (Sheet) Glass Specimens Having Similar Refractive Indices Using Micro X-Ray Fluorescence Spectrometry, JASTEE 2 (2011)

[14] B. E. Naes, S. Umpierrez, S. Ryland, C. Barnett and J. R. Almirall. 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 B 63 (2008) 1145-1150.

[15] P. Weis, M. Dücking, P. Watzke, S. Menges and S. Becker. Establishing a match criterion in forensic comparison analysis of float glass using laser ablation inductively coupled plasma mass spectrometry, J. Anal. At. Spectrom. 26 (2011) 1273-1284.
[16] C. Latkoczy, S. Becker, M. Ducking, D. Gunther, J. A. Hoogewerff, J. R. Almirall, J. Buscaglia, A. Dobney, R. D. Koons, S. Montero, G. J. Q. v. d. Peijl, W. R. S. Stoecklein, T. Trejos, J. R. Watling and V. S. Zdanowicz. Development and Evaluation of a Standard Method for the Quantitative Determination of Elements in Float Glass Samples by LA-ICP-MS, J. Forensic Sci. 50 (2005) 1327-1341.
[17] C. Martinez-Lopez, O. Ovide, R. Corzo, Z. Andrews, J. R. Almirall and T. Trejos. Homogeneity assessment of the elemental composition of windshield glass by μ-XRF, LIBS and LA-ICP-MS analysis, Forensic Chemistry 27 (2022)
[18] R. Corzo, T. Hoffman, T. Ernst, T. Trejos, T. Berman, S. Coulson, P. Weis, A. Stryjnik, H. Dorn, E. C. Pollock, M. S. Workman, P. Jones, B. Nytes, T. Scholz, H. Xie, K. Igowsky, R. Nelson, Kris Gates, h. Gonzalez, L.-M. Voss and J. Almirall. An Interlaboratory Study Evaluating the Interpretation of Forensic Glass Evidence Using Refractive Index Measurements and Elemental Composition, Forensic Chemistry (2021) 100307.
[19] C. M. Bridge, J. Powell, K. L. Steele and M. E. Sigman. Forensic comparative glass analysis by laser-induced breakdown spectroscopy, Spectrochim. Acta B 62 (2007) 1419-1425.
[20] C. M. Bridge, J. Powell, K. L. Steele, M. Williams, J. M. MacInnis and M. E. Sigman. Characterization of Automobile Float Glass with Laser-Induced Breakdown Spectroscopy and Laser Ablation Inductively Coupled Plasma Mass Spectrometry, Appl. Spectrosc. 60 (2006) 1181-1187.
[21] M. M. El-Deftar, N. Speers, E. Stephen, S. Foster, J. Robertson and C. Lennard. Assessment and forensic application of laser-induced breakdown spectroscopy (LIBS) for the discrimination of Australian window glass, Forensic Sci. Int. 241 (2014) 46-54.
[22] E. M. Rodriguez-Celis, I. B. Gornushkin, U. M. Heitmann, J. R. Almirall, B. W. Smith, J. D. Winefordner and N. Omenetto. Laser induced breakdown spectroscopy as a tool for discrimination of glass for forensic applications, Anal. Bioanal. Chem. 391 (2008) 1961-8.
[23] T. Ernst, T. Berman, J. Buscaglia, T. Eckert-Lumsdon, C. Hanlon, K. Olsson, C. Palenik, S. Ryland, T. Trejos, M. Valadez and J. R. Almirall. Signal-to-noise ratios in forensic glass analysis by micro X-ray fluorescence spectrometry, X-Ray Spectrom. 43 (2014) 13-21.
[24] J. N. Miller and J. C. Miller. Statistics and Chemometrics for Analytical Chemistry, Pearson, England, UK (2010).

Supplementary Information

Table S1 – *Vehicle information for sample set.*

Sample ID	Sample Set	Window	Make	Model	Year	VIN
D_01	Original Different Source	Driver's rear window	Dodge	Journey	2010	3D4PH5FV7AT107304
D_02	Original Different Source	Rear window	Ford	Mustang	2010	1ZVBP8CH1A5104406
D_03	Original Different Source	Driver's front window	Mitsubishi	Galant	2009	4A3AB36F49E011396
D_04	Original Different Source	Driver's front window	Volkswagen	Passat	2009	WVWJK73C99P049019
D_05	Original Different Source	Driver's front window	Volvo	XC90	2010	YV4852CZ8A1541591
D_06	Original Different Source	Driver's front window	Mazda	3	2011	JM1BL1K53B1445800
D_07	Original Different Source	Driver's rear window	Ford	Edge	2011	2FMDK4JC0BBA95054
D_08	Original Different Source	Passenger's front window	Honda	Civic	2012	2HGFB2F52CH300384
D_09	Original Different Source	Driver's rear window	Infiniti	G37	2011	JN1CV6AP0BM506147
D_10	Original Different Source	Front sunroof	Cadillac	XTS	2013	2G61P5S31D9123096
D_11	Original Different Source	Passenger's rear window	Ford	Fusion	2013	3FA6P0HR1DR131460
D_12	Original Different Source	Rear window	Nissan	Altima	2013	1N4AL3AP5DC153542
D_13	Original Different Source	Passenger's front window	Volkswagen	Beetle	2012	3VWJP7AT3CM601337
D_14	Original Different Source	Passenger's rear window	Chevrolet	Spark	2013	KL8CB6S95DC618031
D_15	Original Different Source	Passenger's front window	Toyota	Corolla	2014	2T1BURHE4EC001487
D_16	Original Different Source	Driver's front window	BMW	228i	2014	WBA1F5C58EVV98871
D_17	Original Different Source	Driver's front window	Nissan	Murano	2015	5N1AZ2MH3FN202392
D_18	Original Different Source	Driver's rear window	Cadillac	CTS	2015	1G6AR5538F0107436
D_19	Original Different Source	Driver's rear window	Dodge	Ram 1500	2019	1C6RREJT5KN506746
D_20	Original Different Source	Driver's front window	Hyundai	Veloster	2019	KMHTH6AB4KU002950
D_21	Original Different Source	Driver's front window	Chevrolet	Silverado 1500	2019	1GCUWEED5KZ114566
D_22	Original Different Source	Passenger's rear window	Kia	Soul	2014	KNDJP3A54E7000982
D_23	Original Different Source	Passenger's rear vent window	Kia	Soul	2014	KNDJP3A54E7000982
D_24	Original Different Source	Passenger's front window	Maserati	Ghibli	2014	ZAM57RTA5E1077101
D_25	Original Different Source	Rear window	Maserati	Ghibli	2014	ZAM57RTA5E1077101
S_26	Same Source	Driver's rear window	Infiniti	EX35	2011	JN1AJ0HRXBM852178
D_27	Challenging Different Source	Driver's front window	Chevrolet	HHR	2008	3GNDA23D68S569536
D_28	Challenging Different Source	Driver's rear window	Ford/Lincoln	MKS	2009	1LNHM93R69G603573

Sample ID	Sample Set	Window	Make	Model	Year	VIN
D_29	Challenging Different Source	Windshield inner pane	Dodge	Journey	2010	3D4PH5FV7AT107304
D_30	Challenging Different Source	Windshield inner pane	Dodge	Journey	2009	3D9GH57V49T587571
D_31	Challenging Different Source	Driver's front window	Ford	Taurus	2010	1FAHP2EW7AG108539
D_32	Challenging Different Source	Windshield inner pane	Honda	Accord	2009	1HGCP26349A018144
D_33	Challenging Different Source	Windshield outer pane	Honda	Accord	2009	1HGCP26349A018144
D_34	Challenging Different Source	Windshield inner pane	Honda	Accord	2009	1HGCP26400A095510
D_35	Challenging Different Source	Driver's front vent window	Honda	Civic	2010	2HGFA1F5XAH501525
D_36	Challenging Different Source	Rear window	Honda	Civic	2010	2HGFA1F5XAH501525
D_37	Challenging Different Source	Driver's rear window	Honda	Fit	2009	JHMGE882095008412
D_38	Challenging Different Source	Driver's front window	Toyota	Yaris	2009	JTDBT903X91307021
D_39	Challenging Different Source	Driver's front window	Honda	Accord	2009	1HGCP26349A018144
D_40	Challenging Different Source	Driver's rear window	Hyundai	Genesis	2009	KMHGC46E79U043154
D_41	Challenging Different Source	Driver's front window	Chevrolet	Malibu	2009	1G1ZH57B494224763
D_42	Challenging Different Source	Windshield inner pane	Hyundai	Sonata	2009	5NPET46C59H513668
D_43	Challenging Different Source	Rear window	Hyundai	Sonata	2009	5NPET46C59H513668

Table S2 *– False inclusions for Lab B combined different-source set using chemical composition (µXRF element ratios) for each comparison criterion. Comparison results from refractive index and thickness measurements, provided by the FBI, are also included. The last column indicates whether the sample pair was indistinguishable by all three characteristics: chemical composition (for at least one comparison criterion), refractive index, and thickness. IN = indistinguishable, DS = distinguishable.*

Sample Pair ID	Sample Pair Vehicle Information	Chemical Composition							Refractive Index	Thickness	Overall
		Range Overlap	3s	4s	5s	3s 3% RSD_{min}	4s 3% RSD_{min}	5s 3% RSD_{min}			
D_08 vs. D_34	2012 Honda Civic vs. 2009 Honda Accord	DS	DS	DS	DS	DS	IN	IN	DS	DS	DS
D_12 vs. D_34	2013 Nissan Altima vs. 2009 Honda Accord	DS	DS	DS	IN	DS	DS	IN	DS	DS	DS
D_12 vs. D_36	2013 Nissan Altima vs. 2010 Honda Civic	DS	DS	DS	DS	IN	IN	IN	DS	IN	DS
D_29 vs. D_30	2010 Dodge Journey vs. 2009 Dodge Journey	IN	IN	IN	IN	IN	IN	IN	DS	IN	DS

Sample Pair ID	Sample Pair Vehicle Information	Chemical Composition							Refractive Index	Thickness	Overall
		Range Overlap	3s	4s	5s	3s 3% RSD_{min}	4s 3% RSD_{min}	5s 3% RSD_{min}			
D_32 vs. D_34	2009 Honda Accord vs. 2009 Honda Accord	IN	DS	DS	DS	IN	IN	IN	IN	IN	IN
D_33 vs. D_34	2009 Honda Accord vs. 2009 Honda Accord	DS	DS	IN	IN	IN	IN	IN	DS	IN	DS
D_34 vs. D_35	2009 Honda Accord vs. 2010 Honda Civic	DS	DS	IN	IN	IN	IN	IN	DS	DS	DS
D_34 vs. D_36	2009 Honda Accord vs. 2010 Honda Civic	DS	DS	DS	IN	DS	DS	IN	IN	DS	DS
D_40 vs. D_41	2009 Hyundai Genesis vs. 2009 Chevrolet Malibu	DS	DS	DS	DS	IN	IN	IN	IN	DS	DS
D_42 vs. D_43	2009 Hyundai Sonata vs. 2009 Hyundai Sonata	DS	DS	DS	DS	IN	IN	IN	DS	DS	DS

APPENDIX A

This appendix provides a worked example applying the modified comparison criterion, ±3s with a 3% RSD_{min}, to a pairwise comparison using real data collected by Lab B for the same-source set (sample ID: S_26). To be consistent with the same-source pairwise comparisons described in the manuscript, a 1-to-2 fragment comparison scheme is used in this example. That is, the known sample data includes the ten replicate measurements from two randomly selected fragments and the questioned sample data includes the five replicate measurements from a single fragment. For simplicity, only the data for the six ratios recommended in ASTM E2926 is shown. The calculations are shown for Ca/Fe since this ratio led to the highest false exclusions. However, the same calculations can be applied to any ratio of interest.

Table A1 *– Element ratio data for known sample including ten replicate measurements from two randomly selected fragments. The average ($\bar{x}$), measured standard deviation (s_{meas}), and minimum standard deviation (s_{min}) are given. See Formulas 1, 2, and 3 for definition of $\bar{x}$, s_{meas}, and s_{min}.*

Known Sample Replicate	Ca/Mg	Ca/K	Ca/Ti	Ca/Fe	Fe/Zr	Sr/Zr
1	222.146	18.113	24.697	1.719	96.862	1.906
2	215.627	18.158	24.670	1.721	102.853	2.006
3	216.693	18.139	24.573	1.722	92.488	1.850
4	215.502	18.118	24.554	1.721	101.367	1.954
5	220.872	18.141	24.612	1.720	98.236	1.898
6	219.349	18.093	24.666	1.717	95.565	1.879
7	221.225	18.120	24.649	1.717	97.174	1.970
8	216.620	18.122	24.615	1.718	95.352	1.891
9	217.920	18.187	24.653	1.718	100.280	1.990
10	217.349	18.190	24.772	1.718	95.153	1.896
$\bar{x}$	218.330	18.138	24.646	1.719	97.533	1.924
s_{meas}	2.415	0.032	0.063	0.002	3.186	0.052
s_{min}	6.550	0.544	0.739	0.052	2.926	0.058

Table A2 *– Element ratio data for questioned sample including five replicate measurements from one fragment. The average ($\bar{x}$) is given (Formula 1).*

Questioned Sample Replicate	Ca/Mg	Ca/K	Ca/Ti	Ca/Fe	Fe/Zr	Sr/Zr
1	217.453	18.192	24.741	1.731	104.847	1.959
2	217.633	18.030	24.561	1.728	100.714	1.917
3	218.438	18.161	24.606	1.731	107.081	2.060
4	216.770	18.037	24.674	1.728	100.596	1.976
5	220.659	18.140	24.649	1.729	104.021	1.994
$\bar{x}$	218.191	18.112	24.646	1.729	103.452	1.981

Worked Example Using Ca/Fe Data:

First, the average, $\bar{x}$, of the known sample measurements is computed using Formula 1 [24], where x_i is the measurement value and n is the number of measurements (n = 10, in this example). For Ca/Fe, $\bar{x}$ = 1.719.

$$(1.) \qquad \bar{x} = \frac{1}{n} \times \sum_{i=1}^{n} x_i$$

Next, the *measured* standard deviation, s_{meas}, for the known sample is computed using Formula 2, where x_i is the measurement value, $\bar{x}$ is the average ($\bar{x}$ = 1.719), and n is the number of measurements (n = 10). Formula 2 is simply the standard deviation for a sample (i.e., n-1 degrees of freedom) [24]. For Ca/Fe, s_{meas} = 0.00179 (note that this value is rounded in Table A1).

$$(2.) \qquad s_{meas} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

Next, the *minimum* standard deviation, s_{min}, for the known sample is computed using Formula 3. The minimum standard deviation is defined as 3% of the average. Note that this is equivalent to a 3% relative standard deviation (RSD). For Ca/Fe, s_{min} = 0.0516 (rounded in Table A1).

$$(3.) \qquad s_{min} = 0.03 \times \bar{x}$$

The modified comparison interval is defined by Formula 4, in which the interval is calculated using either the measured standard deviation (s_{meas}) or the minimum standard deviation (s_{min}), whichever is greater.

$$(4.) \qquad Modified\ comparison\ interval\ = \bar{x} \pm 3\ \times \max(s_{meas}, s_{min})$$

In this example, s_{min} is greater than s_{meas} (0.0516 > 0.00179) so the comparison interval is defined as 1.719 ± 3 × 0.0516. Thus, the interval lower limit is 1.564 (1.719 – 3 × 0.0516) and the interval upper limit is 1.874 (1.719 + 3 × 0.0516).

Finally, the questioned sample average ($\bar{x}$ = 1.729), computed using Formula 1, is compared to the known sample interval, [1.564, 1.874]. Since 1.729 lies within the interval [1.564, 1.874], the known and questioned sample are indistinguishable by Ca/Fe. Note that if the s_{meas}, rather than s_{min}, was used to define the known sample interval (i.e., an interval of [1.714, 1.724]), the questioned sample average ($\bar{x}$ = 1.729) would lie outside the interval, resulting in a false exclusion.