Robert D. Koons,[1] Ph.D.; Christopher Fiedler,[2] B.A.; and Ronald C. Rawalt,[2] B.A.

Classification and Discrimination of Sheet and Container Glasses by Inductively Coupled Plasma-Atomic Emission Spectrometry and Pattern Recognition

ABSTRACT: Inductively coupled plasma-atomic emission spectrometric determination of the concentrations of aluminum, barium, calcium, iron, magnesium, manganese, sodium, strontium, and titanium in 184 colorless container and sheet glasses was used to evaluate classification procedures for these two classes of glass. Effective two-dimensional visualization of the classification results is afforded by principal components analysis of the data with correct classification of 180 of the 184 samples. Good discrimination among sources of glass within a class is provided using hierarchical, unsupervised clustering. Individual manufacturing plants of a common container glass can be distinguished in most instances.

KEYWORDS: forensic science, glass classifications, spectroscopic analysis, elemental analysis. discrimination, sheet glass, container glass, pattern recognition, cluster analysis

Forensic science examination of glass evidence typically involves the determination of optical and physical characteristics including refractive index, dispersion, color, transparency, fluorescence, surface features, and density. Determination of these characteristics to be indistinguishable between compared fragments indicates a common origin. Assignment of a significance measure to this indistinguishability requires comparison of the measured characteristics with those of many glasses of the same class as the one in question. Classification, defined as the placement of a sample into a product-use class, such as sheet or container glass, is required before any probability considerations to insure appropriate data base comparison for the questioned glass fragment. Glasses such as headlight. cookware, lightbulb, and optical glass are readily classified by their optical and physical properties. However, the ranges of density and refractive indices of sheets, containers, and tableware made of soda-lime-silicate glass overlap, making classification based on these parameters impossible.

Compositional analysis has frequently been considered as a method to provide classification of questioned soda-lime-silicate glass fragments. Composition has also proven useful for discrimination or differentiation between sources of glass within a class. Methods which have been used for determination of element concentrations in glass include neutron activation analysis, X-ray fluorescence, mass spectroscopy, atomic absorption spectrophotometry, spark source atomic emission, and inductively coupled plasma-atomic emission spectrometry (ICP-AES). Much of this research has been done over a period of years in the forensic science laboratories in Great Britain and is summarized in several recent articles by Hickman and coworkers [1-4]. These papers provide an excellent discussion of which elements are most useful for classification and discrimination of glass sources. An ICP-AES procedure developed in these studies [1,5] has proven effective for glass classification using the concentrations of six elements and has been used routinely in the Metropolitan Police Forensic Science Laboratory (MPFSL) for several years. More recently, a classification study of glass in the United States using a semiquantitative method has been presented by Ryland [6].

Hickman proposed a classification procedure using the concentrations of manganese, iron, magnesium, aluminum, and barium (Mn, Fe, Mg, Al, and Ba) plus refractive index (later adding strontium [Sr]) which produced a good success rate in classifying 349 samples of glass [1]. The samples used to construct and test this classification procedure were collected from a variety of clear and colored containers, sheets, and tableware. Most of the sheet glasses used in that and other studies conducted in Great Britain were produced by a single manufacturer, so the analytical results used to develop the classification criteria do not reflect the wider range of glass compositions found in the United States. Hickman and coworkers expanded the classification procedure by consideration of 22 element concentrations and found that Mg, lithium (Li), cobalt (Co), Sr, Fe, and arsenic (As) provide good classification of their glass samples [1,2]. Other studies have shown that in some instances As, cerium (Ce), Co, chromium (Cr), cesium (Cs), Li, lead (Pb), rubidium (Rb), antimony (Sb), Sr, and others offer good discrimination of glass sources [7-9]. The selection of the best elements for classification or discrimination must be based on both the variation of the elements among glass sources and the sensitivity and precision of the analytical method for each element. ICP-AES is a method that has recently been introduced into the forensic science laboratory as a means of quantitatively determining the concentrations of a wide range of major, minor, and trace elements in relatively small samples following their dissolution [5,10,11]. ICP-AES offers the advantages of long linear dynamic ranges, relatively small sample size requirements, multielement capability, and freedom from many of the interelement effects inherent in other spectroscopic methods, making it an attractive method for analysis of glass fragments.

In this paper we report the results of an ICP-AES based study of elemental composition of colorless container and sheet glass samples similar to those appearing as evidence in United States forensic science laboratories. In our previous experience and as noted variously in the British publications, the wide ranges of glass compositions arising from the many sources of glass products in this country exceeds those reported in the British studies. This can produce errors in classification when applying the MPFSL class criteria and other less complete studies reported previously. We selected nine elements for this study because they occur at levels that can readily be determined by ICP-AES in almost all glasses. Of course, in many cases, the presence of Ce or As as decolorizers or boron (B), Li, Co, Cr, Sb, Pb, or others in special-use or colored glasses will offer a very high degree of classification (and discrimination). The concentrations of these elements could readily be determined using the ICP-AES technique presented in this paper with the inclusion of suitable standards. In such cases, refractive index measurements would often be outside the 1.51 to 1.53 range of normal sheet, and container glasses or other physical properties would assist in classifying the sample. We considered several pattern recognition techniques for evaluation of the analytical data for classi-

fication of sheet and container glasses and discrimination of sources within a class. We have considered only sheet and container glasses as the two classes of concern. The sheet glasses include both automobile and building windows and display case glass because they have previously been shown to be compositionally similar [2]. In our use, the term "sheet" applies to flat glasses manufactured by sheet, float, or plate processes. We have ignored tableware in our study because much tableware is either compositionally similar to container glass or one of such a wide range of compositions that it should not be considered as a single class.

Materials and Methods

Glass Samples

The 135 container glass samples used in this study include 85 beverage bottles and 50 food jars. The beverage bottles were collected locally. The food containers consist of two samples from each glass manufacturer supplying 4 packing plants of a babyfood manufacturer. The container glass samples include products from about 20 glass manufacturers (several represented by more than one manufacturing plant) originally containing 50 different products. Of the 135 container glasses, 20 are of foreign manufacture. The containers are not meant to be a representative survey, but should reflect the range of compositions expected for containers in the United States. The 49 sheet glass samples in this study were collected by the FBI Laboratory from casework submissions between 1975 and 1985. The sheet glass samples represent both float and drawn glasses and both automobile and architectural sheets. No information was obtained concerning the age or manufacturer of these samples.

Reagents

Hydrofluoric acid was reagent grade, hydrochloric acid (HCl) was ultrapure, and a 1000-μg/mL scandium solution was made by dissolution of reagent grade scandium oxide (Sc_2O_3) in 5% HCl [8]. Water used for all dilutions was of 18 megohm (MΩ) purity. Multielement standard solutions were prepared by dilution of 1000-μg/mL atomic absorption reference standard solutions in 2% HCl solution to match the sample solutions.

Digestion Procedure

Glass samples were rinsed in deionized water, dried, and broken, and several grams of glass in approximately 50-mg sized fragments were removed. The fragments were cleaned by soaking in concentrated nitric acid for 30 min, followed by three rinses each with deionized water and ethanol. After drying, the fragments were crushed between sheets of polyethylene to submilligram size. Several fragments from each sample with a total weight of 2 to 8 mg were weighed into a 15-mL polypropylene centrifuge tube. Digestion of each sample was begun by the addition of 500 μL of a 1:1 HF:HCl solution and by placing the capped tube in a bath-type sonicator for 1 h. Following sonication, the tubes were placed in an 80°C oven until taken to dryness. The solid was redissolved by the addition of 100 μL of 50% HCl and mixing. The tubes were returned to the oven and again taken to dryness. The resulting solid material was dissolved by the addition of 400 μL of 50% HCl. An aliquot of 0.500 mL of a 1000-μg/mL scandium (Sc) solution was added to each tube as an internal standard and the solution was diluted to about 10 mL with deionized water. The final volume of the solutions is not important since the use of the internal standard corrects all calculated concentrations to an equal volume. A reagent blank sample was made with each set of samples by adding all reagents to an empty tube and carrying it through the heating procedure along with the samples. Accuracy of the analysis and completeness of the digestion were checked by includ-

ing a sample of National Bureau of Standards Standard Reference Material 1831 or 651 with each set of samples.

ICP-AES Analysis Procedure

Following dissolution, sample solutions were analyzed using a Plasma II Model ICP-AES (Perkin-Elmer). Pertinent instrument operating conditions are given in Table 1. The wavelengths [in nanometres (nm)] used for analysis are Al (396.2), Ba (455.4), Ca (393.4), Fe (238.2), Mg (279.6), Mn (257.6), Na (589.6), Sr (407.8), and Ti (334.9). Emission intensities were recorded by scanning the 3600 lines/mm monochromator sequentially over the Fe, Mn, Mg, Ti, and Ca lines and simultaneously scanning the 1800 lines/mm monochromator over the Al, Sr, Ba, and Na lines. An internal scandium standard was used to normalize all emission readings using the automatic standardization mode of the instrument for all elements except Na. For Na determination, manual correction for the Sc internal standard was made using the Sc emission line at 361.4 nm. After analysis of the first 116 samples, we decided that Na offered no classification assistance, so it was not determined for the remaining samples. Calibration curves of element emission intensity ratios in 5 multielement standard solutions were used to calculate element concentrations in the samples. The concentrations of analytes in the standard solutions are given in Table 2. Most of the HF and silicon from the samples is volatilized during the digestion, resulting in an adequate match of sample and standard matrices at 2% HCl plus the elements of interest. The standards were rerun as checks of the calibration after every 20 samples and recalibration was done if needed. Triplicate successive emission readings were taken on each sample and standard solution. The triplicate measurements were used to indicate instrument stability and corrective action was taken when relative standard deviations for several elements exceeded 5%.

The element concentrations in the standard solutions were selected to cover the expected range of sample compositions. Measurements lower than the S1 standard were possible for all elements except iron, and extension of the upper range of linear response could be made for all elements with the inclusion of more concentrated standards. Good accuracy was obtained using this procedure, as indicated by the close agreement between analytical results and certified values for National Bureau of Standards (NBS) Standard Reference Material (SRM) 1831, soda-lime sheet glass shown in Table 3. A bottle[3] was divided into ten 2.5-cm segments and analyzed to provide a measure of the precision of the analytical method. The results of analysis of two samples from each segment, given in Table 4, indicate the combined digestion and analysis precision and compositional variability of the bottle to be in the 1 to 4% relative standard deviation range for most elements. This is generally within the expected precision of ICP-AES for analysis of complex solutions. The results exhibiting the greatest variability are Fe and Ba. Most of the Fe variability occurs because the bottle selected contains levels of Fe close to the lower limit of detection of the method. Also, a small reagent blank correction was required for Fe and some contamination has occurred as evidenced by the two data points which were rejected prior to calculating the mean Fe value. The 7.2% relative standard deviation of Ba values about the mean probably reflects a real variation in the composition of the bottle, since the ICP-AES procedure is very sensitive for Ba response and the concentrations in this bottle are within the analytical range where results normally exhibit 1 to 4% relative deviations. In general, the standard deviation of element concentration measurements in a homogeneous glass is less than 4% for all elements when they are at concentrations greater than in the S1 standard (Tables 3 and 4). The analytical procedure could readily be modified for determination of other elements in glass with similar precision by the inclusion of appropriate standards.

[3]Manufactured by Diamond Glass Company, WV.

TABLE 1—*Instrumental conditions for ICP-AES analysis of glass fragment digest solutions.*

Condition	Value
Incident plasma power	1000 W
Plasma gas flow	15 L/min
Auxiliary gas flow	1.0 L/min
Nebulizer gas flow	1.0 L/min
Viewing height	15 mm above rf coil
Monochromator	1800 lines/mm (wavelength > 395 nm)
	3600 lines/mm (wavelength < 395 nm)
Signal compensation	yes, for all elements except Na
Background correction	off
Integration time	100 ms
Sample uptake rate	1.0 mL/min

TABLE 2—*Concentrations of analyte elements in ICP-AES standard solutions for glass analysis. All concentrations are in μg/mL.*

Element	S0	S1	S2	S3	S4
Al	0	0.200	0.50	1.00	4.00
Ba	0	0.005	0.01	0.02	0.05
Ca	0	0.500	2.00	5.00	20.00
Fe	0	0.010	0.04	0.10	0.40
Mg	0	0.040	0.20	1.00	8.00
Mn	0	0.005	0.01	0.02	0.05
Na	0	0.500	2.00	5.00	20.00
Sr	0	0.005	0.02	0.05	0.10
Ti	0	0.020	0.05	0.10	0.20

TABLE 3—*Results of ICP-AES analysis of NBS SRM 1831 soda-lime sheet glass. Values given are element percent by weight.*

Element	Mean Concentration of Five Replicates	Standard Deviation	NBS Certified Value
Fe	0.067	0.006	0.061
Al	0.644	0.007	0.640
Mn	0.0015	0.0003	...
Sr	0.0089	0.0002	...
Mg	2.18	0.02	2.120
Ba	0.0031	0.0001	...
Ti	0.0125	0.0008	0.011
Ca	6.00	0.06	5.860

TABLE 4—*Results of analysis of two samples of glass fragments from each of ten sections, numbered from top to bottom, of a bottle manufactured by Diamond Glass Co., WV. Results are element percent by weight in the glass except relative standard deviation, which is in percent.*

Sample	Al	Fe	Mn	Sr	Mg	Ba	Ti	Ca
C71–1a	0.802	0.025	0.0102	0.1441	0.172	0.0147	0.0361	7.76
C71–1b	0.775	0.026	0.0105	0.1436	0.172	0.0159	0.0362	7.58
C71–2a	0.787	0.024	0.0101	0.1439	0.170	0.0147	0.0361	7.66
C71–2b	0.773	0.026	0.0103	0.1439	0.170	0.0145	0.0360	7.57
C71–3a	0.800	0.026	0.0107	0.1456	0.176	0.0146	0.0354	7.74
C71–3b	0.775	0.051[a]	0.0103	0.1439	0.170	0.0146	0.0363	7.74
C71–4a	0.784	0.027	0.0103	0.1417	0.171	0.0144	0.0352	7.65
C71–4b	0.777	0.027	0.0100	0.1438	0.169	0.0144	0.0354	7.64
C71–5a	0.790	0.023	0.0104	0.1418	0.171	0.0149	0.0351	7.70
C71–5b	0.784	0.031	0.0106	0.1423	0.173	0.0177	0.0363	7.74
C71–6a	0.800	0.028	0.0100	0.1440	0.172	0.0150	0.0360	7.81
C71–6b	0.778	0.041[a]	0.0112	0.1419	0.175	0.0182	0.0365	7.70
C71–7a	0.793	0.024	0.0101	0.1427	0.171	0.0147	0.0355	7.73
C71–7b	0.794	0.027	0.0105	0.1435	0.172	0.0148	0.0352	7.80
C71–8a	0.809	0.025	0.0103	0.1468	0.176	0.0151	0.0364	7.90
C71–8b	0.772	0.313[a]	0.0117	0.1410	0.172	0.0147	0.0358	7.57
C71–9a	0.788	0.025	0.0105	0.1429	0.170	0.0146	0.0357	7.76
C71–9b	0.753	0.026	0.0102	0.1428	0.167	0.0147	0.0360	7.53
C71–10a	0.796	0.031	0.0100	0.1423	0.172	0.0172	0.0366	7.66
C71–10b	0.780	0.030	0.0109	0.1430	0.173	0.0162	0.0368	7.75
Mean	0.786	0.027	0.0104	0.1433	0.172	0.0153	0.0359	7.70
S.D.[b]	0.013	0.002	0.0004	0.0014	0.002	0.0011	0.0005	0.09
R.S.D.[c]	1.7	8.9	3.8	1.0	1.3	7.2	1.4	1.2

[a] Rejected as an outlier by the method of Dixon [12].

[b] Standard deviation.

[c] Relative standard deviation.

Discussion of Results

The results of analysis of the glass samples in this study can be expressed as a matrix consisting of 9 columns, 1 for each of the 9 elements determined and 184 rows, 1 for each sample. Two immediate problems facing the examiner of such a multidimensional data set are the recognition of patterns within it and visual presentation of these results. We used a statistical software package (SYSTAT, ver 3.0) operating on a desktop personal computer to evaluate several approaches to solve these problems. The following discussion details some of the conclusions that were drawn from the evaluation of this data. The attraction of an interactive statistical approach operating on a personal computer is the flexibility it affords the examiner of data in observing patterns and making conclusions based on valid statistical comparisons. However, in evidentiary situations, pattern recognition procedures should only be used to assist the examiner in observing and displaying patterns in the data rather than as a direct statistical test of the significance of a hypothesis of equality of two samples. The two procedures reported in this study, cluster analysis and principal components analysis, should be considered as data analysis and display methods. Classification or discrimination protocols resulting from the application of these procedures can be evaluated only through the use of blind tests. Once a procedure has been shown to produce accurate classification results consistently in blind trials, then a high degree of confidence can be placed in the classification of unknown source samples using this procedure.

Data Presentation

The first step in data analysis is the determination of which elements are useful classifiers of sheet and container glasses. Histograms displaying the concentration distributions of the nine elements in sheet and container glasses are shown in Fig. 1. Elements with good classifying power can be identified in Fig. 1 as those which separate container and sheet glass samples into unique compositional groups. It is evident from the distributions that no single element affords complete separation of the two classes of glass. However, low concentrations of Mg or high concentrations of Ba or Sr uniquely characterize many of the container glasses as distinct from the sheets and the remaining few containers. High concentrations of Fe, or to a lesser extent Mn or low concentrations of Al, characterize some of the sheet glasses. The concentrations of Ca, Na, and Ti indicate no ranges that uniquely characterize either of the two classes of glass. One disadvantage of histogram representations of data is that multiple element comparisons of individual samples cannot be made. For example, of the 38 container glass samples containing high concentrations of Ba, 13 also contain high concentrations of Sr. There are 9 additional container glasses which contain high concentrations of Sr, but not of Ba. More sophisticated approaches than histograms are required to display multidimensional relationships of this type.

The element concentration data can be further evaluated by the comparison of the sample statistics for each element in the two classes of glass. A common method of comparing the means of two sample populations is by the use of Student's t-test. Use of the t-test in its simplest form requires measured values for random samples of both populations to exhibit a normal distribution about the population mean with equal variances, assumptions which clearly are not met for several of the elements shown in Fig. 1. The lack of normal distributions can be ignored since it is relatively unimportant in applying the t-test [13]. Equality of variances can be evaluated by the F test [13]. Results of F test calculations indicate, for our data, that only Na and Ti have equal variances in container and sheet samples, and all other elements have variances which are significantly different between the two classes of glass. We have selected the option of ignoring the violations of equality of variance assumption and have calculated t-values for our data rather than using more complicated testing procedures not requiring equality of variances. The t-values thus calculated can be used as a relative indication of the classifying power of the individual elements. The mean concentrations, standard deviations, and t-values for each element in the two classes of glass are given in Table 5. The t-values decrease in the order Mg > Al > Ca > Fe > Mn > Ba > Sr > Ti > Na. The low values of t for Ti and Na indicate that there is no significant difference between the mean concentrations of these elements in sheet and container glasses. Hence, their inclusion in the data set offers no additional classifying power over the data set with their omission. The t-values for the remaining seven elements are all high, indicating that there is a significant difference between the mean element concentrations between the two classes. The relative values of t agree well with the observations made from the histograms shown in Fig. 1, although differences in the order of t-values and our observations of the classifying power of the elements from the histograms occur in several instances. The t-value for Ca is quite high even though no unique ranges of concentrations exist for either class of glass. This reflects the good approximation to a normal distribution for both sheet and container Ca concentrations and the fact that the mean Ca concentration of 6.64% for containers is significantly higher than the 4.71% for sheets. In comparison, the Ba histogram uniquely classifies 38 of the 135 container glass samples by having values greater than 0.1%, and the mean Ba concentration for container glass is about 5 times the concentration in sheet glass, yet the t-value for Ba is lower than that for Ca. This results because the distribution of container Ba concentrations is skewed, resulting in a large relative standard deviation. Using both the t-values and direct visual observation of the histograms, we conclude that the elements Al, Ba, Ca, Fe, Mg, Mn, and Sr provide good classification power and will be used to build a
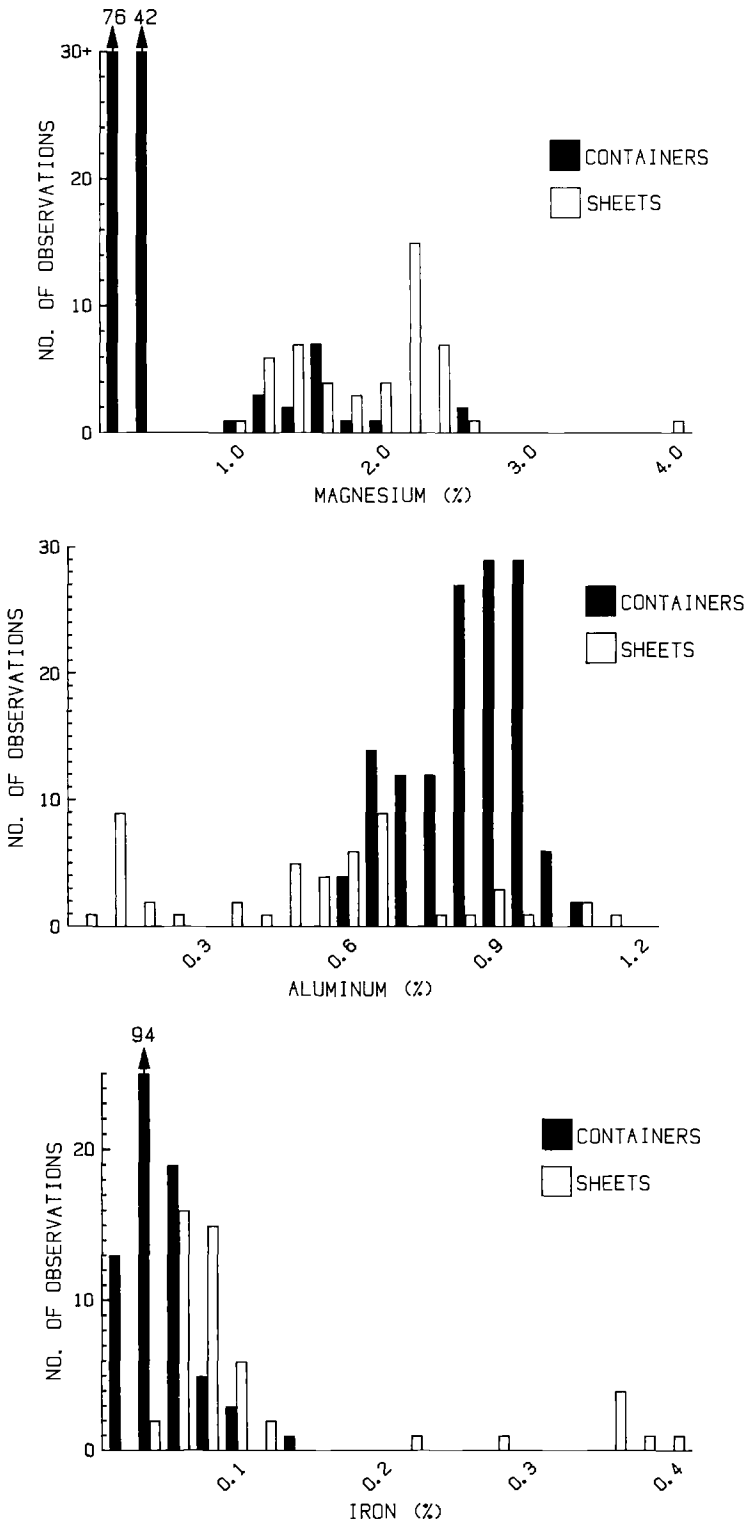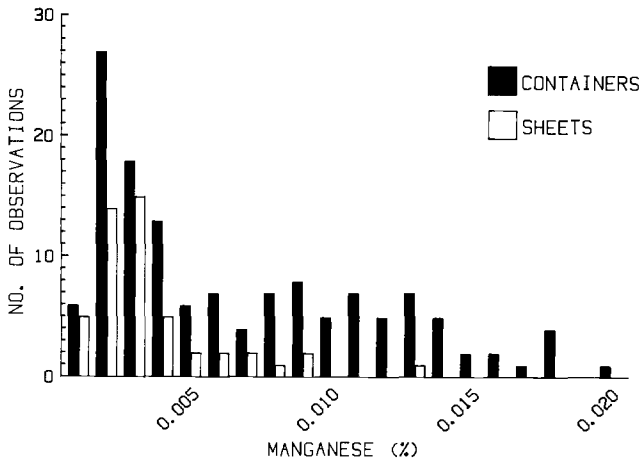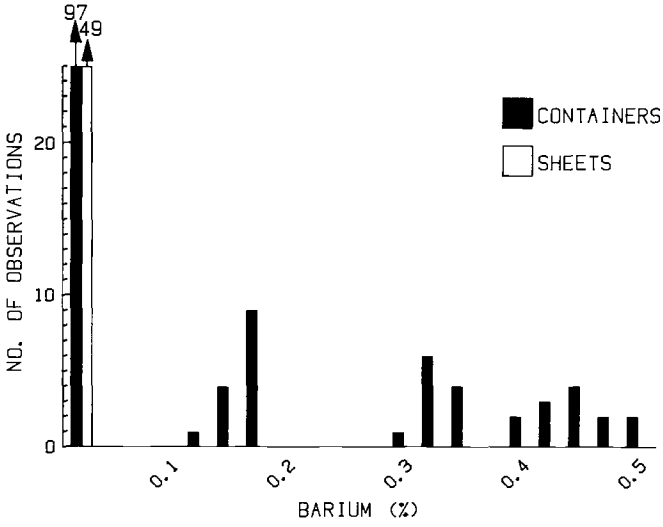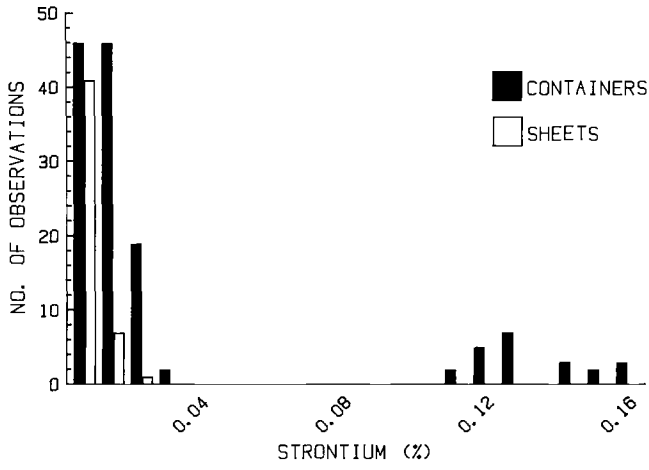
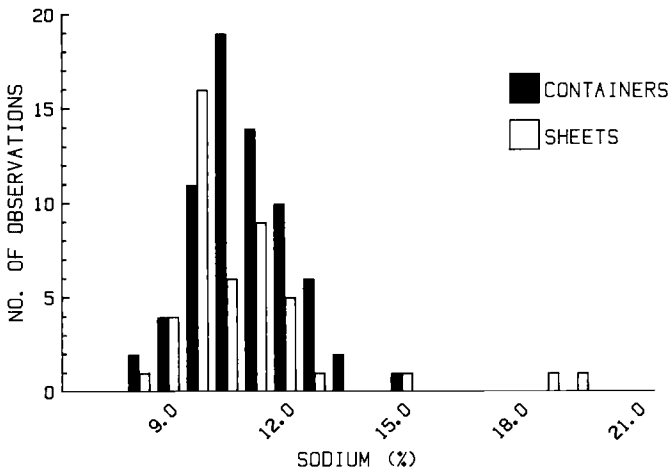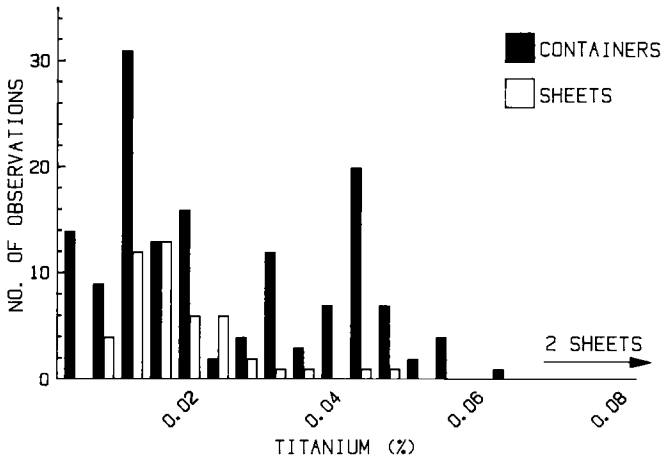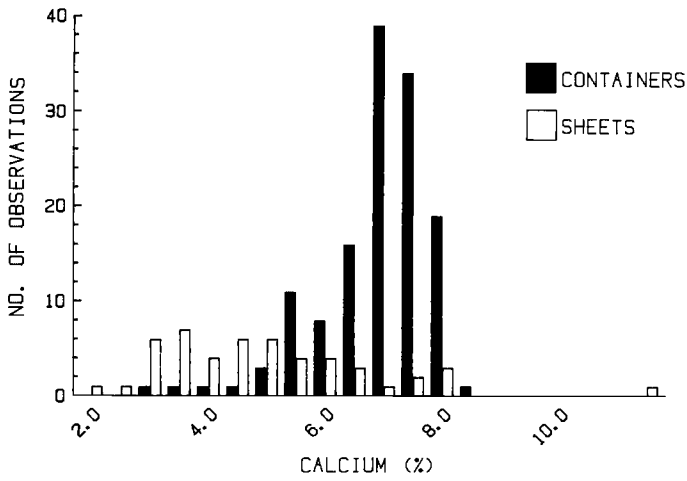FIG. 1—*Element concentrations in glasses.*

FIG. 1—*Continued.*

FIG. 1—*Continued.*

TABLE 5—*Sample statistics for sheet and container glass compositions. Mean and standard deviation results are expressed as percent by weight of element in glass.*

Variable	Fe	Ca	Mn	Al	Mg	Sr	Ti	Ba	Na
Container mean	0.034	6.64	0.0062	0.82	0.32	0.03	0.024	0.088	10.6
Container S.D.	0.017	0.95	0.0049	0.11	0.49	0.045	0.015	0.145	1.3
Sheet mean	0.107	4.72	0.003	0.49	1.82	0.0046	0.021	0.0017	10.5
Sheet S.D.	0.104	1.82	0.0024	0.29	0.53	0.0033	0.025	0.0039	2.1
Overall mean	0.054	6.13	0.0054	0.73	0.72	0.023	0.023	0.065	10.6
Overall S.D.	0.064	1.50	0.0046	0.23	0.83	0.040	0.018	0.130	1.6
Pooled class S.D.	0.055	1.24	0.0044	0.18	0.5	0.038	0.018	0.124	1.7
t-statistic	7.890	9.33	4.35	11.12	17.88	3.92	1.06	4.16	0.3

classification system. This result agrees well with the MPFSL procedure which is based on ICP-AES determination of Al, Ba, Fe, Mg, Mn, and Sr. The concentrations of Fe, Al, Mg, Mn, and Ba in each class of glass in our study are similar to those reported by Hickman (for example, compare our Fig. 1 with Fig. 7 of Ref 4), but the ranges of concentrations are wider in our study, particularly for Ba. This reflects the wider range of glass sources available in the United States than in Great Britain. Comparison of our data with the British data indicates that the compositions of our glass samples commonly fall outside the limits of their class ranges in the MPFSL classification system. Although it is not useful for classification, Ti is a good element to use for discrimination of similar sources, since its concentration range is more than 100 times greater than its analytical uncertainty shown in Table 4.

Nonstatistical Approach to Classification

A simple, nonstatistical approach to classification of the glass samples used in this study can be derived by the application of a series of decisions to the data using those elements offering strong classification power. Using this approach, we observe Fig. 1 and select ranges of an element which uniquely define one product-use class and delete those samples from the data set. For example, from the histogram for Mg, we can see that if the Mg content of an unknown glass fragment is less than 0.5%, then there is a high degree of probability that the glass fragment came from a container. By deletion of the samples meeting this criterion, we produce a smaller data set. The process of selection of classifying criteria can be successively repeated for the remaining samples and elements until all samples are correctly classified. For this approach to be most effective, no samples should have element concentrations close to the cutoff values used in placing samples into classes.

One possible classification procedure based on such a sequence of decisions is shown in Fig. 2. In this procedure, high concentrations of Ba and Sr and low concentrations of Mg are used to classify container glasses with a relatively high degree of certainty. Concentrations of Fe, Mn, and Al characterize the sheet and remaining container glass samples with progressively decreasing degrees of certainty. This procedure is attractive for courtroom presentation because it is not based upon statistical considerations and can be applied to an unknown sample without any computation. It also gives some degree of confidence to the placement of a sample into its resulting class. Disadvantages of this classification procedure are that it does not provide any quantitative information about the interrelationships between samples within a class or between classes and that the predictive value for those samples near the bottom of Fig. 2 is not very good. In fact, a more conservative classification would replace the last decision with "Is the concentration of Fe > 0.05?" with positive answers (13) being classified as sheets and negative answers (4) being classified as "unclassifiable." This classi-
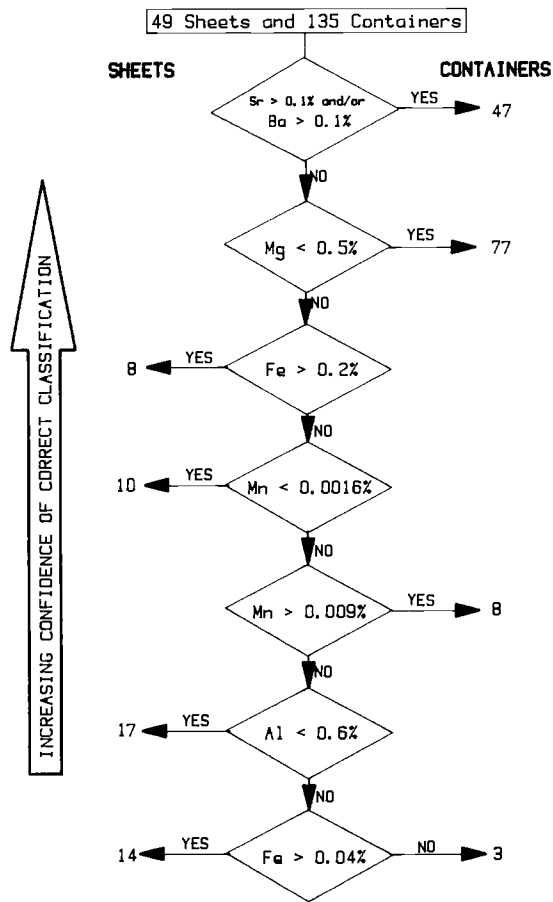
FIG. 2—*Flow chart for nonstatistical decision approach to classification of a glass fragment into sheet and container classes.*

fication procedure also uses only six elements, and several of those to define only one class; thus it may not be the best means of using the data to perform the classification.

For comparison, we applied Ryland's approach based on selected ratios between Mg, Ca, and Fe [6] to the 184 samples in our data base after adjusting his selection criteria to reflect differences in concentration and X-ray fluorescence peak intensities. Ryland's procedure correctly identifies 128 of the 135 containers and 25 of the 49 sheets with the remaining 7 container and 24 sheet glasses classified as "unclassifiable." From this limited comparison, it appears that the use of 6 elements rather than 3 increases the probability of correctly classifying glass fragments, particularly for sheet glass samples.

Principal Components Analysis

The pattern recognition procedures used in our study are designed to provide insight into the classification of glass, that is, to determine the natural clustering of data points in the nine-dimensional space of our analytical results. "Cluster analysis" is the name given to several methods by which data points lying close together can be grouped and distances between clusters of points can be determined.

Before discussion of clustering methods, however, we will discuss reducing the dimensionality of the data set by principal components analysis (or eigenanalysis). If we represent each sample in our study as a vector in the multidimensional space comprising the measured variables or features, then the goal of principal component analysis is to reduce the variables to a number where patterns can be seen in conventional plots (ideally two or three features) while retaining as much variance of the data as possible. In this context, principal component analysis can be considered a display method in which data points are projected onto new sets of axes formed by linear combinations of the variables (or features) in the original data set. This approach is similar to factor analysis, which relies upon using a known set of samples to form a combination of factors such that the original separation of true clusters of samples is retained in the minimum number of factors. In chemical systems, the minimum number of factors that retain the separation of different clusters of samples represents the number of chemical processes that control the distribution of measured features in the training set. Principal components and factor analysis methods are well developed in many applications, particularly in the geology, chemistry, and social science disciplines, and the interested reader should refer to the literature for more information about the derivation and use of these methods [14,15].

We have used unsupervised principal components analysis to reduce our data to two components to see whether containers and sheets form compositionally distinct natural groupings. This principal components analysis algorithm uses the Pearson Product-Moment correlation coefficients between all combinations of the measured variables as a starting point for combining features to reduce the number of components. This approach reveals the natural variations of the composition data, as distinct from a learning set approach, such as used by Hickman [2], where separation of the two classes is made based upon a training algorithm derived from known source glass measurements. A plot of the first 2 principal components for the concentrations of Al, Ba, Ca, Fe, Mg, Mn and Sr (Na and Ti have been omitted because of their lack of classifying power) in the 184 container and sheet glasses is shown in Fig. 3. The solid line indicated in Fig. 3 separates the container and sheet glass composition domains. The separation is complete except for 2 container glass samples (both from the same manufacturing plant) and 2 sheet glass samples which are classified incorrectly. Component 1 is a linear combination of the variables, dominated by Al, Ca, Fe, and
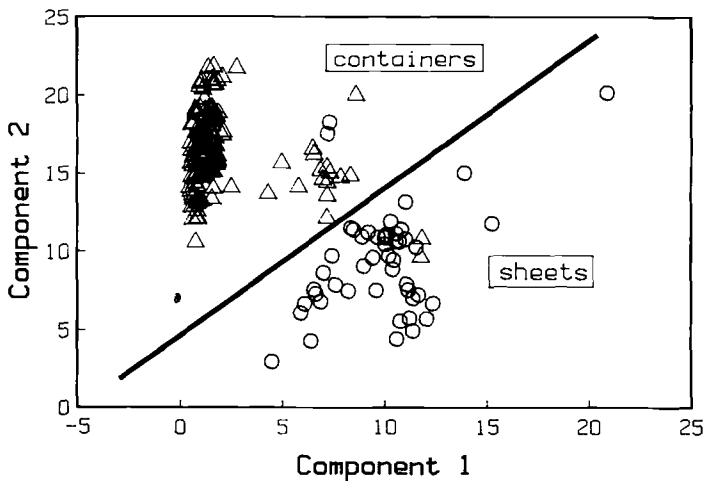

FIG. 3—*Unsupervised principal components analysis results for Al, Ba, Ca, Fe, Mg, Mn, and Sr. First two principal components are shown.*

Mg, and Component 2 is composed predominantly of Sr and Ba. The loadings have been rotated to maximize variable contributions to either component (VARIMAX rotation). About 51% of the original variability of the data is retained in the first 2 components. Although inclusion of more components would retain more of the original sample variability, 2 were selected for ease of plotting and inclusion of a third component does not improve classification of the 4 incorrectly placed samples. The good separation of classes using only 2 components is an interesting observation. One explanation for this is that contamination in the sand, lime, and alumina used in the glass manufacture is the major factor controlling elements like Ba, Sr, Mn, and Fe (Component 1) and manufacturing processes control elements such as Al, Mg, Ca, and to some extent Fe (Component 2).

Cluster Analysis

Another approach for classification and also discrimination among glass sources is the use of cluster analysis. There are many excellent summaries of the use of clustering for glass analysis [1-3] and general applications [16], so we will limit discussion to the procedures used in our study. The goal of cluster analysis of multivariate data is to arrange all of the samples in multidimensional space and group those samples together that have short intragroup distances compared to intergroup distances. Cluster analysis is analogous to biological classification systems based on grouping samples together that display common features. The use of a distance measure and quantitative feature descriptions for classification of samples places several requirements on the data.

First, scaling is required so that different weights are not given to elements occurring at widely different concentrations in the samples. For example, if we used our unscaled data, where all concentrations are expressed as weight percent, the differences in Ca concentrations that are in the percent range would dominate over the differences in Ba that are in the 0.001% range. The most common forms of data scaling for clustering are logarithmic and Z-score transformations. A logarithmic transform, as used in the British glass studies [1-4], replaces each element concentration by the logarithm of its value. The disadvantage of this approach is that it overemphasizes the differences in small concentration values, which may have large relative analytical uncertainties. The Z-score transformation replaces each analytical value with the difference between that value and the mean divided by the standard deviation of all values for that element. Disadvantages of Z-score scaling are the requirements of a normal distribution of data and the fact that the variable scores must be changed every time an additional sample is added to the data set. To avoid the drawbacks of these methods, we used range scaling in which each measured element concentration is replaced by a scaled value according to the formula

$$\text{Scaled element value} = \frac{\text{measured value} - \text{lowest value}}{\text{highest value} - \text{lowest value}}$$

For each element, the highest and lowest values are those weight percents obtained for the entire data set. Thus, each element concentration is transformed in the scaled data set to the fraction of its distance between the lowest and highest value of that element concentration (a value between 0 and 1). This approach offers the advantages of placing all variables on scales of nearly identical weight and retaining low levels of significance for concentrations close to the detection limit and for small differences between samples. For this application, we used the measured range for each element for scaling. For wider application, as in building an archived data base, ranges of expected values can be used so that addition of a new sample having a concentration outside of the previously determined range will not cause a change of all prior scaled results. We have found the method of range scaling to be very useful, not only for glass analysis, but for compositional comparisons of several types of evidence, where

clustering is to be used to display the similarities and relationships among a group of up to several hundred samples. In most evidentiary situations, we recommend using range scaling based on the range of results for those samples only. This approach generally provides the greatest discrimination within the comparison set of samples.

The second restriction on the data is the requirement of independence of variables if a Euclidean distance measure is to be used for clustering. The Pearson Product-Moment correlation coefficients for all combinations of the 9 elements in the 184 samples are shown in Table 6. It is generally accepted that variables can be considered independent for clustering purposes if the correlation coefficient is less than 0.5. For the combined sheet and container glass data set, all element pairs are correlated at a level significantly below this value except for Mg and Ca which are near it. This inverse relationship, which we will consider insignificant for purposes of clustering, is to be expected from consideration of glass manufacturing methods. Magnesium is the only element in our study which manufacturers intentionally maintain at a different concentration in sheet and container glass. Magnesium is added to a level of about 2% in most sheet glass to prevent devitrification and improve its flow properties. For glasses having Mg contents below 0.5%, none has intentionally been added by the manufacturer. Magnesium occurs in these glasses adventitiously from its presence in the lime or cull glass used as raw materials in glass making. The slight negative correlation between Mg and Ca concentrations arises in part because manufacturers of container glass attempt to keep the sum of MgO plus CaO constant at about 7% by weight. The addition of Mg to sheet glasses lessens the inverse correlation between the two elements when sheet and container glasses are mixed in a single data base. As the dolomite content of lime increases, the Mg increases and Ca decreases in an approximately proportional manner, further contributing to the observed negative correlation. The consequence of the lack of linear correlations between variables in our data set is that Euclidean distance measures can be used in clustering.

The unsupervised hierarchical clustering algorithm we used begins by finding the pair of data points connected by the shortest Euclidean distance in the multidimensional space. These points are replaced by a single point at their mean location and the process repeated. Results are printed out in a dendrogram or "tree diagram" in which the lengths of branches connecting samples reflect the distance between them and their relationship to other samples. The cluster diagram for 20 sheet and 20 container glass samples randomly selected from our data set is shown in Fig. 4. The scaled values of Al, Ba, Ca, Fe, Mg, Mn, and Sr were used to construct this dendrogram. For classification purposes, we would like to be able to cut the tree at the point where there are only 2 branches and have the 2 resulting groups be

TABLE 6—*Pearson Product-Moment correlation coefficients for 184 sheet and container glasses.*

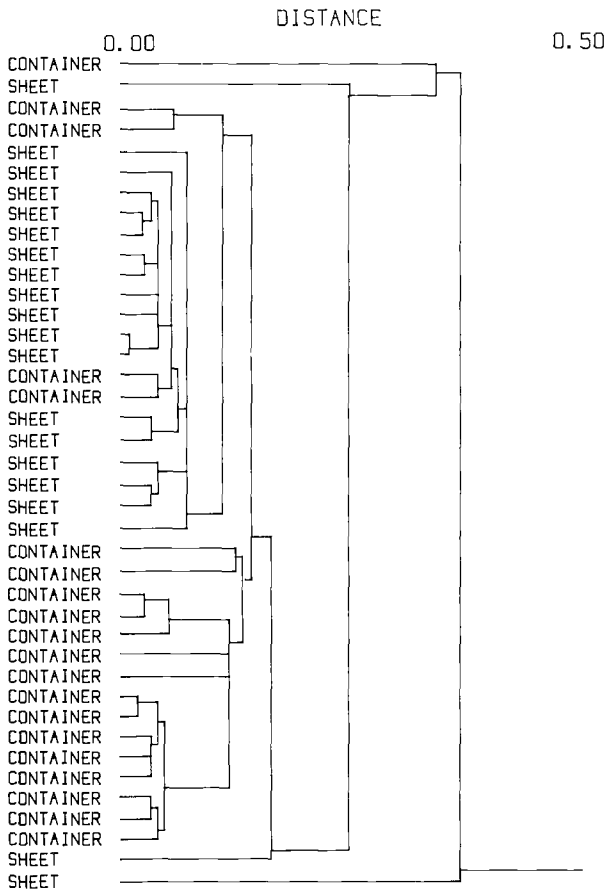|							
Ba	0.163						
Ca	0.407	0.113					
Fe	−0.361	−0.120	−0.263				
Mg	−0.493	−0.292	−0.533	0.370			
Mn	0.260	−0.002	0.122	−0.119	−0.128		
Sr	0.220	0.317	0.214	−0.168	−0.288	0.196	
Ti	0.163	−0.001	0.129	0.316	−0.141	0.284	0.359

FIG. 4—*Dendrogram of 40 container and sheet glass samples using Al, Ba, Fe, Mg, Mn, Sr, and Ti for classification.*

containers and sheets. However, as shown in Fig. 4, grouping to the level of two clusters does not produce good classification characteristics. Instead, containers from some manufacturing plants are more like some sheets than they are like containers from a different source. Discrimination between samples as made by a one-to-one comparison of element concentrations indicates that samples separated by a clustering distance of about 0.1 units are significantly different. Cutting the branches in Fig. 4 at this distance results in formation of about 14 discriminated clusters within these 40 glasses. Similar results occur when other glasses in the data set are selected for clustering. We noted during attempts to cluster a wide variety of glasses that samples made in the same manufacturing plant tend to cluster together. Thus, clustering offers an effective means to answer questions of discrimination or differentiation among glass sources within the same class. An example follows illustrating the use of cluster analysis for a data set typical of those which might be encountered in evidentiary situations.

The dendrogram resulting from unsupervised hierarchical clustering of the 46 food container samples in our study using all elements except Na and Ca is shown in Fig. 5. Each sample is identified by the glass manufacturer and location of the production facility where it is known. In all cases, short branch lengths connect jars from within each manufacturing plant. Cutting the branches at a length of about 0.025 distance units divides the data into clusters with complete discrimination of each of the manufacturing plants, with the excep-
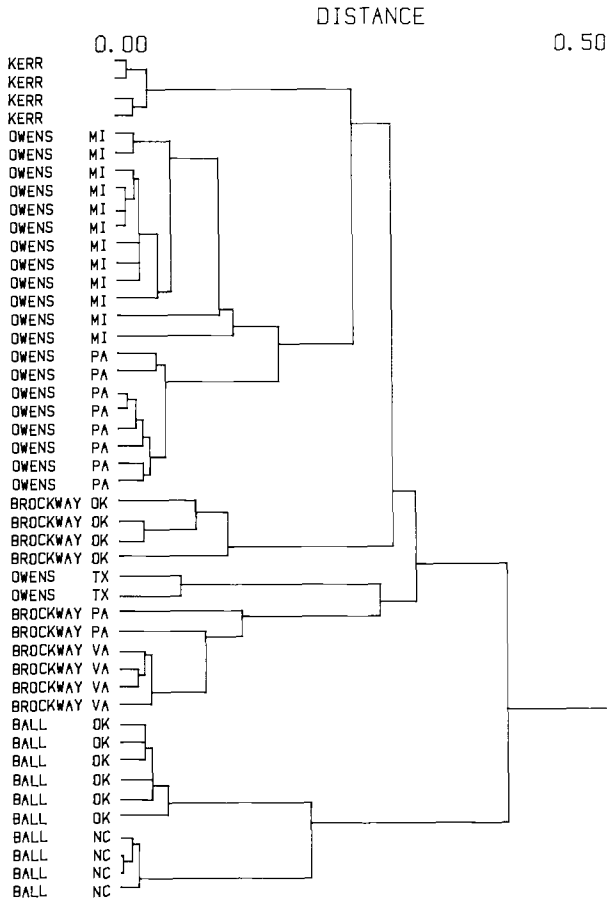
DISTANCE

FIG. 5—*Dendrogram of 46 food container samples using Al, Ba, Fe, Mg, Mn, Sr, and Ti. Samples are identified by manufacturer and plant location.*

tion of the Brockway Glass Company plants in Virginia and Pennsylvania. The interrelationships between different sources of glass jars are indicated by the lengths and ordering of branches connecting different manufacturing plants. In several instances, two plants of different companies are more similar to each other than they are to other plants of the same company. Some of the differences in the glass appear to relate to geographical location, an observation that supports the hypothesis that some of the measured element concentrations derive from the local sources of raw materials used in glass manufacture. Also note that the glass samples for the Owens-Illinois plants in Clarion, Pennsylvania, and Charlotte, Michigan, were produced over a 3-year period. Considering this, the clustering of these samples together is remarkably good.

Conclusions

Accurate concentration measurements are essential for the use of classification procedures, since the placement of an unknown sample into a class is based on comparison with previously defined parameters. Discrimination between glasses from different sources hav-

ing similar refractive indices requires precise analysis to differentiate between the similar compositions. The procedure using ICP-AES for analysis offers the precision and accuracy necessary for such comparisons for glass fragments larger than about 500 μg. Other procedures such as electron beam-induced X-ray analysis or spark source emission spectrometry allow use of smaller samples, but generally do not provide the necessary accuracy and precision.

Pattern recognition techniques are useful for determining sample groupings that are difficult to recognize when considering multidimensional data sets and can provide effective visual display of complex compositional relationships. Principal components analysis provides a means of displaying the natural separation of sheet and container glasses into two groups with little overlap. Extension of this approach to previously unclassified samples provides a means of classification into the two groups. Cluster analysis is an effective means of displaying similarity of composition of glass fragments and discrimination among sources. It is particularly useful when several sources of glass are present in a collection to be evaluated together.

In an evidentiary situation, where known and questioned source fragments are found to be indistinguishable in their physical and optical properties, classification by compositional analysis can be used to exclude alternate sources. Additionally, from the data gathered in this study, there is an indication that when the composition of known and questioned glass fragments are found to be indistinguishable, greater assurance can be placed in the correctness of a common source conclusion than when optical and physical parameters are used alone. Alternatively, it has been reported and we have also observed that instances occur where glass fragments from different sources have indistinguishable refractive indices but readily distinguishable compositions. Thus, in comparing known and questioned sources of glass fragments, the significance of a "common source" conclusion appears to be improved when composition is considered in addition to optical and physical parameters.

A general procedure to be followed for handling casework situations where two glass samples are to be compared to see if a common source is indicated is to look for a one-to-one comparison of each element after finding indistinguishable physical and optical parameters. Pattern recognition procedures may help display relationships among several sources of glass, but will not, in general, provide the criteria for considering two samples to be a compositional "match." As indicated in other studies on this subject, our classification system has not yet reached a point of optimum classification of all classes of glass and it would be helpful to acquire more data on additional elements. The ICP-AES method is particularly advantageous in this regard. because additional elements could be determined with only the inclusion of more standard solutions. To answer questions of discrimination of similar sources, more information is needed concerning the homogeneity of glass objects submitted as evidence. From preliminary data, it appears that compositional analysis offers additional discrimination capability over physical and optical characterization of glass fragments. Additional data can and are, in fact, being acquired in several laboratories in the United States and Great Britain concerning the variation in composition of evidential glasses. Compilation of these data will provide useful information concerning the reliability of compositional analysis for the discrimination of glass sources.

References

[1] Hickman, D. A., "A Classification Scheme for Glass." *Forensic Science International.* Vol. 17, 1981, pp. 265–281.
[2] Hickman, D. A., Harbottle, G., and Sayre, E. V., "The Selection of the Best Elemental Variables for the Classification of Glass Samples," *Forensic Science International.* Vol. 23, 1983, pp. 189–212.
[3] Hickman, D. A., "Elemental Analysis and the Discrimination of Sheet Glass Samples," *Forensic Science International,* Vol. 23, 1983, pp. 213–223.

[4] Hickman, D. A., "Linking Criminals to the Scene with Glass," *Analytical Chemistry*. Vol. 56, No. 7, June 1984, pp. 844A-852A.

[5] Catterick, T. and Hickman, D. A., "The Quantitative Analysis of Glass by Inductively Coupled Plasma—Atomic-Emission Spectrometry: a Five-Element Survey," *Forensic Science International*, Vol. 17, 1981, pp. 253-263.

[6] Ryland, S. G., "Sheet or Container?—Forensic Glass Comparisons with an Emphasis on Source Classification." *Journal of Forensic Sciences*, Vol. 31, No. 4, Oct. 1986, pp. 1314-1329.

[7] Andrasko, J. and Maehly, A. C., "The Discrimination between Samples of Window Glass by Combining Physical and Chemical Techniques," *Journal of Forensic Sciences*, Vol. 23. No. 2, April 1978, pp. 250-262.

[8] Goode. G. C., Wood, G., Brooke. N., and Coleman, R. F., "Multi-element Analysis of Glass Fragments by Neutron Activation and the Application to Forensic Science," Report 024/71, Weapons Research Establishment, Aldermaston, U.K., 1971.

[9] Haney, M. A., "Comparison of Window Glasses by Isotope Dilution Spark Source Mass Spectrometry," *Journal of Forensic Sciences*. Vol. 22, No. 3, July 1977, pp. 534-544.

[10] Carpenter, R. C. and Till, C., "Analysis of Small Samples of Brasses by ICP-OES and their Classification by Two Pattern-Recognition Techniques," *Analyst*, Vol. 109, July 1984, pp. 881-884.

[11] Locke, J., "The Application of Plasma Source Atomic Emission Spectroscopy in Forensic Science." *Analytica Chimica Acta*. Vol. 113, No. 1, Jan. 1980, pp. 3-12.

[12] Youden, W. J. and Steiner, E. H., *Statistical Manual of the Association of Official Analytical Chemists*, Association of Official Analytical Chemists, Arlington, VA, 1975, pp. 30-31, 75-76.

[13] Steel, R. G. D. and Torrie, J. H., *Introduction to Statistics*, McGraw-Hill, New York, 1976, pp. 245, 252-254.

[14] Massart, D. L. and Kaufman, L., *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, Wiley, Chemical Analysis Monograph Series, Vol. 65, 1983.

[15] Davis, J. C., *Statistics and Data Analysis in Geology*, Wiley, New York, 1973.

[16] Sneath, P. H. A. and Sokal, R. R., *Numerical Taxonomy: The Principles and Practice of Numerical Classification*, W. H. Freeman, San Francisco, 1973.

Address requests for reprints or additional information to
Robert D. Koons, Ph.D.
Forensic Science Research, FBI Lab
FSRTC, FBI Academy
Quantico, VA 22135