

The Use and Limits of Compositional Analysis for Discrimination and Classification of Samples

Michelle M. Barton and Gordon M. Miskelly*

Forensic Science Programme, Department of Chemistry, University of Auckland,
Private Bag 92019, Auckland (e-mail: g.miskelly@auckland.ac.nz).

The recent decision by the FBI to cease using compositional analysis of bullet lead (CABL)¹ has prompted those of us interested in compositional fingerprinting of materials to re-evaluate our practices, and ensure that we are aware of the boundaries of such techniques. Our group has been particularly interested in whether it is possible to identify the region of origin of wines, or the source of a diesel spill, while others in Australasia have been interested in identifying the origin of items as diverse as cannabis² and gold.³

The FBI used compositional analysis of bullet lead (CABL) in cases where bullets associated with a crime scene were so badly damaged that they could not be compared by the more traditional method of *e.g.* comparison microscopy. The FBI estimated that it had used CABL for ca. 2500 cases, with ~20% of these proceeding to trial¹ whereas CABL analysis has been used only once in New Zealand, during the third trial⁴ of John Barlow in 1995. In CABL, replicate bullet fragments are digested, and then the resulting solutions analysed for Ag, As, Bi, Cd, Cu, Sb, and Sn using inductively-coupled plasma-optical emission spectroscopy (ICP-OES). The data are then compared with analyses from bullets associated with a suspect, with similar results being used to support the hypothesis that the bullets came from the same source.^{5,6} Following criticism of CABL,⁷ the FBI requested that the US National Research Council evaluate the CABL technique, including the sampling, chemical analysis, statistical analysis, interpretation, and presentation of results. This *Forensic Analysis: Weighing Bullet Lead Evidence* evaluation⁵ noted that the sampling, digestion procedure, and ICP-OES compositional analysis of bullets were appropriate. However, concern was expressed over the methods used for the comparison of compositional data from different bullets, and also the interpretation to be placed on such comparisons. The first involved consideration of the appropriate statistical methods to be used in comparisons of multivariate data, and included a recommendation to use a multivariate comparison such as Hotelling's T^2 . The second involved the lack of knowledge about the source populations: how bullet lead varied in composition, and how likely it was that two random bullets might be compositionally indistinguishable, together with how conclusions were presented in court.

Compositional Fingerprinting

Compositional fingerprinting aims to characterize a sample in such a way that its similarity or dissimilarity to other samples can be evaluated. The characterization typically includes elemental or GC analysis of organic content, but could also include measurements such as isotope ratios.

Commonly, many components are analyzed, with the naïve expectation⁸ that discrimination will improve with an increased number of variables. The elemental analyses are frequently performed with instruments such as ICP-OES or inductively-coupled plasma-mass spectrometry (ICP-MS) that are capable of analyzing many elements in a single sample within a short timeframe. Alternatively, organic compounds can be characterized by an information-rich technique such as GC-MS. It is also possible to include measures other than chemical composition into a classification scheme, *e.g.* the patterns of minor peaks in the GC trace of an organic substance can be used for discrimination or classification of oil samples, even if it is not known what compound causes a given peak.⁹ It is also possible to have categorical variables, colour, presence or absence of a component, *etc.*, included in a classification or discrimination scheme.

Questions addressed by compositional analysis

It is important to identify the question being asked of a given analysis. A common aim is to determine whether two samples are sufficiently similar that they could have come from the same source - the alternatives being that they are sufficiently dissimilar that they are unlikely to come from the same source, or that there is insufficient information available to make a decision. Such a decision can be referred to as *discrimination*. The identification of similar samples could be expanded to consideration of whether several samples should be placed in a given class - a *classification* decision. An alternative classification decision is to determine whether a sample came from a given source - region/manufacturer/lot, *etc.* There are two key questions that are implicit in such considerations: *What do we mean by source?* and *How do any classes identified within the samples relate to what one might regard as sources?* It is also important to consider the type and strength of conclusions that should be drawn from the observation that two samples are compositionally indistinguishable. Each of these issues is addressed below.

Measurement of elemental (or other) composition

Since fingerprinting analysis usually uses at least five elements, and may involve elements with widely differing concentrations, the most common techniques are ICP-OES and ICP-MS, with neutron activation analysis also having been used. For ICP-OES and quantitative ICP-MS this requires standards for all elements of interest, but if lower accuracy is acceptable ICP-MS in semi-quantitative mode (which does not require standards for all elements)

can be used. With many elements being analyzed over a wide range of concentrations, contamination and interferences are of serious concern, especially when the samples have complex matrices. For example, in a recent ICP-MS analysis of wine, we found that scandium ($^{45}\text{Sc}^+$) had false high readings due to interference from $^{12}\text{C}^{16}\text{O}_2^+\text{H}^+$ (from the high organic content of the samples),¹⁰ while ^{153}Eu showed interference from $^{137}\text{Ba}^{16}\text{O}$ when the rare earth element concentrations were very low. Interference can also be caused by multiply charged ions, thus $^{86}\text{Sr}^{2+}$ can interfere with $^{43}\text{Ca}^+$ (^{86}Sr : 9.86%; Ca : 0.135% natural abundance), the latter ion having been used in ICP-MS because of the interference by $^{40}\text{Ar}^+$ and $^{12}\text{C}^{16}\text{O}_2^+$ on the more plentiful isotopes of Ca . In typical samples, the Sr concentration is so much lower than that of Ca that interference is not of concern. However, if a commercial standard with equal concentrations of these elements (provided so that users do not have to purchase many elemental standards separately) is used to standardize the ICP-MS it can lead to a systematic underreporting of the Ca concentrations. These particular interferences are problems in using the most common, and lowest-cost, ICP-MS instruments with quadrupole mass spectrometers – high-resolution ICP-MS instruments, such as that recently acquired at Otago University, are not subject to these interferences at high resolution. ICP-OES also has interferences, including spectral interferences, and this technique is also less sensitive than ICP-MS. However, it is suitable for analyses such as the composition of elements in lead bullets, where the high lead content makes routine ICP-MS analysis difficult.

Obtaining appropriate standards can be difficult also for other compositional analyses, e.g. standards for all biomarkers in crude oils can be very expensive. In such cases, it is sometimes possible to rely on GC-MS libraries, or to use a sample of crude oil for which peak identities are known to identify the peaks of interest, as is done for the Eurocrude crude oil identification protocol.¹¹

Data Analysis

As noted earlier, component concentrations can vary widely. In general, it is important to ensure that those components with low concentrations receive adequate weight in the analysis (without unduly biasing the results) since trace components may be discriminating. This may require some form of data transformation prior to statistical analysis. It is also important to clearly define the limits of detection, and to decide how to deal with elements that lie below the limit of detection. This is especially true if a logarithmic transformation is to be used on the data prior to statistical analysis (a common practice for natural samples where concentrations are often log-normally distributed).¹²

There are two potential errors that can be made during comparisons: falsely matching samples (*false match*), or incorrectly stating that samples are different when, in fact, they are not (*false exclusion*). Any analysis needs to protect against both of these errors but, depending on the purpose of the analysis, one or other of these errors may be weighted more heavily.

A final consideration is whether or not the composition

of the sample could have been altered prior to analysis. This is well known for oil-spill samples, where weathering processes such as evaporation, dissolution, and biodegradation can alter the composition observed by gas chromatography.¹³ Compositional changes can also occur due to processing of a foodstuff, e.g. the fining of wine¹⁴ or the recovery or smelting of metals.¹⁵ Such changes may require the analyst to place less weight on (or even disregard) some elements or components.

The source

One of the key considerations in classification for provenance is to define and characterize the *source*. It may be a single, invariant material or there may be many sources, or the output from any source may vary with time. An obvious example of multiple invariant sources is human DNA, and a second is the biomarker profiles of crude oils that remain sufficiently constant so that spilled oil can be traced to the field of origin by use of a systematic comparison to a comprehensive database, e.g. that in Eurocrude.¹¹ Temporal variation is shown in examples such as biomarkers in diesel from the NZ refinery at Marsden Point (which uses many different crude oils as feedstock) and gold ore from South African mines (where the exact composition can vary within the veins being extracted from a given mine¹⁵).

Some source variation is useful since it provides increased potential for discrimination between different samples. However, it can make classification more difficult if the variability is not adequately characterized (through samples of known provenance) or modeled. The Marsden Point example can be expanded to show an additional problem of source identification. Thus it is possible to define many *sources* within the diesel supply chain: the oil fields, the tankers, the refinery, the batch, the service station, and the vehicle. Some of these sources may be useful in a given context whereas others may not. Once a source(s) and any temporal changes have been identified, the mean and spread of each variable characterizing that source (and time period) need to be obtained. Such studies also need to take into account other sources of variability (such as analytical measurement) and confounding factors such as alteration of the profile of the sample, e.g. *weathering* in oil spill analysis, and contamination. It is also necessary to know the overall spread of all populations, and how likely it is for two populations have similar compositions.

The NRC evaluation of the FBI CABL method used the term *compositionally indistinguishable volume* to denote the origin of a population of lead samples that would be analytically indistinguishable using the CABL technique.⁵ It also noted that it would be possible for two *different* compositionally indistinguishable volumes to be indistinguishable using CABL, although there were insufficient data to estimate the frequency of this occurrence. The NRC committee estimated that the compositionally indistinguishable volumes of lead could range from ca. 30 to 100,000 kg in size, resulting in estimates of ca 12,000 to 35 million bullets coming from a given compositionally indistinguishable volume (for comparison, the report also

estimated that about 10 billion bullets are produced annually in the US).⁵ Based on this high variability, combined with variability in the bullet manufacturing and packaging process, and lack of knowledge of the geographic distribution of bullets produced, the NRC stated that it was not possible to model the bullet manufacturing process in such a way as to predict the compositional distribution of bullets, and that it was not possible to make statements about either the specific times of manufacture of a given bullet or the probability of finding a bullet in a particular geographical area.⁵

The NRC evaluation would not have greatly impacted on decisions that bullets were dissimilar. However, it has been established that it is possible for a single box of bullets to contain samples from two or more different compositionally indistinguishable volumes,⁵ so that the CABL analysis could not in general be used as exclusionary evidence between a crime scene and a suspect. In contrast, in many other applications of forensic compositional fingerprinting dissimilarity can result in a decision to exclude a match.

Linking observed classes to specific sources

Often it is only necessary to determine whether or not it is likely that two samples originated from the same *source*. However, sometimes it is desirable to go further than this and link samples to a particular *source*. The observation of clustering in a statistical evaluation of multivariate composition data does not necessarily allow identification with a meaningful source. The apparent clustering could be due to random variation, or it could reflect an effect associated with the processing of the item examined – examples include wines that have been treated with bentonite and show elevated rare earth element content¹⁴ and the effects of metal recovery processes¹⁵ on trace metal profiles in gold. Alternatively, clustering could be due to contamination or other environmental alteration of a subset of samples. If the observed clustering of samples correlates with known origins it may be possible to develop a rule linking samples to a particular origin. Alternatively, supervised classification methods could be used where an adequate number of samples with known provenance are used to develop the classification rules. In this latter case it is important that sufficient samples are used to derive the rules, and that they are representative of all the possible sources from which samples could originate. This requires setting up large inclusive databases that can involve large amounts of time and resources.

Conclusion

The manufacturing, processing, and distribution of bullets in the US are sufficiently variable that it is very difficult to assign a particular meaningful *source*, and differentiate that source from other potential sources. Without detailed distribution knowledge it is also difficult to know the significance of finding two bullets that come from the same source. Our reading of the recent decision to stop CABL analysis is that the FBI felt that it would take too many resources to gain adequate bullet population information compared to the number of times that the information would be used in court. Presumably the FBI decided

that the NRC recommendation to present the comparison of two or more bullet analyses in court with the expert witness only stating that the result *is consistent with having come from the same compositionally indistinguishable volume* did not provide sufficient weight to the evidence to justify use of the technique. It should be noted that there are parallel limitations to data interpretations in other fields, e.g. the ASTM standard D5739-00 for comparison of spills of petroleum-based products states the oil samples should be interpreted as *similar, dissimilar or inconclusive*.¹⁶ Such apparently limited statements can still be useful, e.g. if there are four ships in port, spilled oil may be described as similar to that obtained from one but dissimilar to the other three. This limitation in data interpretation contrasts to the situation for crude oils, where there is a limited, known population so that comparison of GC profiles with a comprehensive database can identify a match with greater certainty.

In general, discrimination or classification based on compositional analysis requires a large database that is representative of the complete population from which a sample may be drawn. Such a database needs to show the variability within samples, the variability between samples with the same source, and variability between samples with different sources. Such databases can represent substantial investments of time, effort, and money. Many research papers on provenance are based on relatively small databases that can allow discrimination between a few significantly different classes but are not suited for more subtle discrimination; they may not provide true estimates of classification error rates, if the samples are not truly representative of the source populations.

Many of these issues are also being addressed in forensic multi-element glass analysis, traditionally performed using refractive index comparisons. There is a reasonable body of knowledge of glass fragment occurrences, including surveys to estimate the extent and variability of glass fragments being found in the environment or on people, and there is good knowledge of the variability of refractive index across glass populations. However, glass manufacture is now so standardized that the refractive indices of samples of a given type of glass are very similar. Multi-element analysis has shown an improved ability to discriminate between different pieces of glass, and the development of analysis techniques and methods for data interpretation are active areas of research. Given its recent implementation, the databases for multi-element analysis of glass are smaller than those for refractive index, which again places limits on the strength of any conclusion based on glass fragment similarity. The statistical methods to be used in multielement glass discrimination are still under discussion, with recommendations by some authors that Hotelling's T^2 or similar be used,¹⁷ or that a Bayesian approach is more appropriate.^{18,19}

Multi-element analysis of gold from mining operations has shown similar needs for extensive databases. Differentiation of gold from significantly different ore deposits is possible with a relatively low number of reference samples.³ However, differentiation of gold from mines within the relatively homogeneous ores of South Africa

has been much more difficult, and required precise measurements of many elements, combined with selected isotope ratios.^{15,20} These latter studies also noted that processing of the gold ore could change selected elemental compositions, and that gold obtained from a given mine could change in profile as different veins were accessed or if a given vein had a composition gradient.

In conclusion, while many of the criticisms of the CABL analysis can be applied to other compositional analyses, there are also some unique features. The issues of statistical data analysis certainly apply to all multivariate comparisons, and there is still controversy over the best approaches to use. The issue that posed greatest difficulty for CABL, however, was the variability inherent in the source bullet populations, and lack of knowledge of the geographical variability of bullet distributions. Clearly, for any provenance study based on compositional analysis, it is imperative that an adequate number of samples be collected to define all possible sources. If the aim is to test for discrimination between samples then it is sufficient to collect enough samples to make good estimates of between-sample variability and within-sample variability. Finally, it is necessary to take note of the comments made in the NRC report on the language used in reporting results, so that the strength of any association is not overstated.

References

1. FBI National Press Office, Washington, D.C, press release, 1 Sept. 2005.
2. Watling, R.J., *J. Anal. Atomic Spectr.*, **1998**, *13*, 917-926.
3. Watling, J.R., Herbert, H.K., Delev, D., and I.D. Abell, *Spectrochim. Acta B*, **1994**, *49*, 205-219.
4. *Barlow seeks royal pardon after FBI flaw*. In *The New Zealand Herald*, 8 Sept. 2005.
5. *Forensic Analysis: Weighing bullet lead evidence*, NRC, The National Academy of Science, Washington, DC, 2004.
6. Koons, R.D. and Buscaglia, J. *J. Foren. Sci.*, **2005**, *50*, 341-351.
7. Randich, E., Duerfeldt, W., McLendon, W., and Tobin, W., *Foren. Sci. Int.*, **2002**, *127*, 174-191.
8. Hughes, G.F., *IEEE Trans. Information Theory*, **1968**, *IT-14*, 55-63.
9. Lavine, B.K., Brzozowski, D., Moores, A.J., Davidson, C.E., and Mayfield, H.T., *Anal. Chim. Acta*, **2001**, *437*, 233-246.
10. Reed, N.M., Cairns, R.O., Hutton, R.C., and Takau, Y., *J. Anal. Atomic Spectrom.*, **1994**, *9*, 881-896.
11. *European crude oil identification system (project EURO-CRUDE)*, Denmark National Environmental Research Institute 1995.
12. Limpert, E., Stahel, W.A., and Abbt, M., *BioScience*, **2001**, *51*, 341-352.
13. Wang, Z. and Fingas, M.F., *Marine Pollution Bull.*, **2003**, *47*, 423-452.
14. Jakubowski, N., Brandt, B., Stuewer, D., Eschnauer, H.R., and Görtges, S., *Fresenius' J. Anal. Chem.*, **1999**, *364*, 424-428.
15. Merkle, R.K.W., Dixon, R.D., and Kijko, A., *Trace element characterization of gold and platinum-group element deposits and products in South Africa*. In *Applied Mineralogy: Developments in Science and Technology, ICAM 2004*, Sao Paulo, Brazil, 2004.
16. *D5739-00 Standard Practice for Oil Spill Source Identification by Gas Chromatography and Positive Ion Electron Impact Low Resolution Mass Spectrometry*, ASTM, 2000.
17. Curran, J.M., Triggs, C.M., Almirall, J.R., Buckleton, J.S., and Walsh, K.A.J., *Sci. Justice*, **1997**, *37*, 241-244.
18. Curran, J.M., Triggs, C.M., Almirall, J.R., Buckleton, J.S., and Walsh, K.A.J., *Sci. Justice*, **1997**, *37*, 245-249.
19. Curran, J.M., *Int. Stat. Rev.*, **2003**, *71*, 497-520.
20. Dixon, R.D., personal communication, NITEcrime workshop, Wellington, May 2004.