Data interpretation in forensic sediment and soil geochemistry

Ruth M. Morgan and Peter A. Bull

Oxford University Centre for the Environment, University of Oxford, South Parks Road, Oxford, OX1 3QY, UK

Communicating author: Ruth Morgan

Ruth.morgan@ucl.ac.uk
Tel: +44 20 3108 3037
Fax: +44 20 3108 3088
Data interpretation in forensic sediment and soil geochemistry

Ruth M. Morgan and Peter A. Bull

Oxford University Centre for the Environment, University of Oxford, South Parks Road, Oxford, OX1 3QY, UK

Abstract
Automated geochemical techniques enable reproducible elemental assays of small quantity samples and have been used in recent years in many forensic criminal investigations in England and Wales. Two case studies are presented which highlight the problems of testing the presence of pre-, syn- or post-crime event sample mixing. The number of elements or compounds analysed can have a bearing upon statistical discriminant techniques which may provide false-positive or false-negative associations or exclusions. Chemical analyses of soils and sediments using both Atomic Absorption Spectrometry (AAS) and Dionex, and Inductively Coupled Plasma Mass Spectrometry (ICP-MS) enabled the identification and classification of discrete groups by hierarchical cluster analysis and Canonical Discriminant Function Analysis. These groupings however, prove fragile to small variations within samples of even the most common minerals.

Keywords: Forensic science, geochemical sediment analysis, ICP-MS, AAS/Dionex, Canonical Discriminant Function Analysis.
**Introduction**

The forensic investigation of any sediment or soil usually involves the consideration of the physical, chemical and biological components of a particular sample compared with the same characteristics of other samples taken at a forensic scene. These samples are analysed in conjunction with comparator samples, which may be taken from the clothing or possessions of a suspect. Comparator samples can also be used to predict the provenance of a sample (Brown 2002, Lombardi 1999) and so, such forensic sediment analysis utilises techniques of description or comparison so that exclusion can be identified (Bull et al 2004, Bull et al 2005). The fundamental difference between the approaches adopted by geological and forensic practitioners for the analysis of a sediment or soil is that in forensic analysis samples cannot ‘match’, whereas in geological analysis, attempts to identify provenance are common and the fundamental approach is to group sediments into units, strata or facies (Bull and Morgan 2005).

In order that the relative similarity or exclusion by comparison of two or more samples can be effected (usually including samples from the scene of the crime), some measure of the probability of similarity or dissimilarity is needed and this can be best achieved with reference to a pertinent database (Saks and Koeler 2005). Unfortunately in forensic geoscience there are few relevant databases; indeed in the geological and pedological sciences, databases with the accuracy of elemental distribution and frequency have never been needed and in consequence only the sketchiest databases exist. A database is, of course, useful for search and location investigations but they are also important in the establishment of the relative individuality of the samples in question. Without a database it is not possible to
ascertain whether the sample in question is likely to have derived from the same provenance as the comparator sample.

Given the philosophical differences between geological and forensic precepts, it is necessary to consider the representative nature of sample collection (Gilbert and Pulsipher 2005, Nocerino et al 2005 and Warren 2005). Trace analysis often involves micrograms rather than kilograms of material (Murray and Solebello 2002); the relevant analytical techniques therefore, will be determined by the size of the available sample (not withstanding the need to keep some sample back for re-analysis by other scientists). There exists also the problem of comparing trace samples to bulk samples and fine grain materials to coarse grain materials (Morgan and Bull in press). Comparison needs to be ‘like with like’ and will be determined by the smallest sample within the forensic sediment set.

In forensic sediment geochemistry, given the philosophical precepts and sample representativeness outlined above, further problems may be encountered. Decision must be made whether to search for the exotic or the ubiquitous component of a material. The exotic may well provide a marker compound or element but question may arise as to its relevant paucity. Such rarity, although used to highlight the similarity of two samples by one side of a legal argument, may well be utilised by the other side of that argument to cast doubt upon the probability of finding such a small amount of material either within a comparison sample (false-positive association) or, not finding such a small amount of material within a comparison sample (false-negative association). Whether it is possible to argue the contention that an association is false presents a major problematic interpretation encountered by
forensic scientists. A direct analogy in geological sciences is found when comparing the relevant abundances of heavy minerals within a sample (see for example Hubert 1971).

Automated analysis, potentially perhaps the most important development in forensic sediment geochemistry in recent years, is also fraught with problems. Often, such automated analysis requires the homogenisation of a sample either into a liquid or into a reconstituted sediment ‘pellet’ in order to standardise the elemental composition throughout a sample and its subsequent aliquots. The problem with such homogenisation is that it prevents the consideration of pre-, syn- or post-forensic event mixing. Of course, the experienced forensic analyst will attempt to forestall such a problem during sediment collection prior to preparation and analysis. Unfortunately, the reality check is that many of these analytical techniques (such as ICP) will routinely measure to parts per million or billion (or even trillion and quadrillion). The very nature of the collection of trace materials from the belongings and clothing of a victim or suspect almost inevitably involves mixing of materials of different provenances before the relevant forensic event (by walking in shoes, or wearing of clothing in an open environment). It is just not feasible to rely upon the careful collection of the small amounts of material that is needed for analysis so that any mixing can be considered to have been avoided. Mixing during the forensic event is, of course, what is sought to be compared between the forensic scene and its comparator samples, however, these materials will have incorporated themselves with the pre-event deposition materials creating a classic admixture. Unless the suspect is apprehended at the scene, post-event transfer and thus, mixing, will occur and given
the persistent nature of trace materials this melange may continue for some considerable time.

In the last decade or more, relatively cheap automated analytical machines for elements and compounds have become readily available and thus have been used routinely for the analysis of physical evidence in the British courtroom. Most techniques utilise spectrometry in one form or another, progressing from the atomic absorption spectrometer (AAS) to the more recent mass spectrometers (ICP-AES, MS etc.). Initially, analysis was restricted to a limited number of cations and anions and, although any one element or compound could be analysed in batches, it was not possible to undertake multiple and rapid cation/anion analysis which can be routinely performed by the more recent generation of analytical machines. These machines provide multiple samples (with drift corrections) and multiple element analyses, so much so that they generate enormous amounts of data in a relatively short time span (and hence are considered cost effective). The problem with this plethora of information is that the interpretation and presentation of results is extremely problematic. Little account can be made of dependent elemental correlations, let alone the production of known, or more troublesome unknown positive and negative associations thrown up by the analytical machinery utilised. ‘The data do not lie, but the devil is in the interpretative detail’ (Ioannidis 2005, Kirk 1974).

There are specific problems in the analysis, interpretation and presentation of results from automated forensic sediment geochemical analysis. These problems are universal for any technique which involves the homogenisation of a sample for analysis. The interpretation of the results involves the thorny issue of identifying
false-positive or false-negative associations between samples which may lead to totally erroneous data interpretation. Below we provide two examples of geochemical analysis utilising simple techniques of elemental analysis to highlight the problems of data interpretation and ways to get it wrong.

**Case Study 1**

In the case of R vs. Flavious (2005) a guilty plea was made on the second day of the trial, subsequent to the completion of all forensic reports. The case involved the transportation of a body from London to a countryside location in the Midlands of England. During the subsequent forensic investigation, soil/sediment samples were taken around the close vicinity of the body deposition site and also from the entry and exit points to that site. A motor vehicle, used by the prime suspect, was found to have a muddy interior, particularly in the drivers footwell. Within this generally distributed mud, a superimposed footprint was found. Three groups of samples were taken; the footprint, the loose dried mud in the footwell, and the general mud debris adhering to the footwell floor. These three soil/sediment samples were compared to the soil/sediment samples taken from the body deposition site in order to assess whether the mud in the car could be excluded from having derived from the crime scene. To effect this comparison, one of the tests to which the samples were subjected involved AAS and Dionex analysis (following the usual sample preparation techniques outlined in Anderson 1976) to determine the elemental composition of a selection of cations and anions in each sample (see figure 1). The selection of the cations and anions was governed by the practicality of availability in the laboratory as has been the case for the majority of analyses undertaken in forensic enquiry. These results were then
computed, using SPSS, to produce a hierarchical cluster analysis to provide a visual representation of the results for the jury (Figure 2).

The cluster analysis used was that of average linkage, a technique well tried and tested in biological and geological studies (Davis 1973). The hierarchical clustering shows that the five samples taken from the body deposition site group well together in their relative abundances of the tested cations and anions with that of the superimposed footprint found in the car footwell. Further, figure 2 shows that the other mud found in the car footwell was not of the same chemical composition as that found at the body deposition site and the loose mud debris found in the driver footwell of the motor vehicle showed very great dissimilarity to any other samples analysed.

This association portrayed in figure 2 suggests that two of the samples found within the footwell of the car can be excluded from having derived from the body deposition site with reference to the gross chemical analyses, although the sample taken from the superimposed footprint can not be excluded from having derived from the body deposition site. The relative ‘success’ of this analysis must be tempered with reference to results obtained by other independent means of analysis (and in this case since the chemical composition of the soil has been analysed it should not include mineralogical identification, nor colour considerations as both can be argued to be dependent upon the chemical composition of the soil). Full details of other techniques employed can be found in Morgan (2006) but lie outside the remit of this paper. It is important to note however, whether other lines of evidence concur with or contradict those results obtained by chemical analysis, it is not possible to determine whether the
associations highlighted by chemical analysis are influenced by either one or more of pre- syn- or post-forensic ‘event’ mixing. The technique is correct, it is perhaps the failure to be able to identify mixing of sediment from different sources that is the problem. Specifically in this case, other non-dependent, physical-based methods of forensic analysis (such as scanning electron microscopy of quartz grain surface textures) show that the mud found in the motor vehicle footwell and the loose debris found in the same footwell indicate that some components of those samples may well have contained material from the body deposition site (for full details see Morgan 2006). However, the tests employed for chemical analysis were unable to discriminate the pre- syn- and post-forensic event mixing phases that were so evident from visual inspection of the samples. Furthermore, it may be argued that the hierarchical cluster technique employed is dependent on the number of samples being compared. Indeed, the use of any multivariate statistical technique may incur criticism, however, some simplified means of agglomerating the results in such a way for the jury to follow the case has been needed. Cluster analysis (and other multivariate statistical techniques) have been used in many trials for communicating the findings of geoforensic analysis to judge and jury in the English court system. These points are addressed further and more pertinently in the second case example presented in this paper.

**Case Study 2**

The body of a young woman was found in a wooded area in the English Midlands. The accused was alleged to have killed the woman and then to have driven a vehicle containing the body from a town centre to the body deposition site in the woods approximately four miles away. CCTV footage provided a 22 minute window for the
vehicle to make the round trip and for the body to be deposited up a steep and muddy wooded slope (in the middle of the night). The weather at the time was very wet and overcast and there was no street lighting whatsoever in the area. Given this scenario it would be reasonable to expect that the person or persons transporting the body up the slope should be covered in mud and other debris picked up at the site.

A whole array of chemical and physical tests (SEM of quartz and other minerals, mineralogy, grain size analysis, pH, conductivity, colour, AAS/Dionex and ICP-MS (see Morgan 2006 for full details)) were undertaken on samples taken about the deposition site (18 samples), from the vehicle (6 samples), from the victim and from various items of clothing from the suspect. Many of the samples collected contained only a trace amount of material (0.5g) although other samples contained considerably more material (see the problems of comparing trace and bulk samples analysed by chemical means, highlighted in Morgan et al. 2006). It was thought that it might be constructive to compare the chemical analysis results of materials taken from the various sites associated with this case from both AAS/Dionex (utilising 8 cation and 3 anion analyses; Ca, Mg, Fe, Mn, Cu, Pb, Na, K, Cl, SO$_4$ and NO$_3$) with ICP-MS analysis of 13 compounds and elements taken from subsamples of the same samples (Al$_2$O$_3$, Fe$_2$O$_3$, MgO, CaO, Na$_2$O, K$_2$O, TiO$_2$, P$_2$O$_5$, MnO, Ba, Cu, Ni and Zn). The aim of this test was to determine whether the number of cation/anions or element/compounds analysed would yield different results, identifying exclusions of samples.

Initial chemical analysis involved establishing cation and anion composition by AAS and Dionex (figure 3) and the results were arranged in a dissimilarity index by
average linkage hierarchical cluster analysis (figure 4). On initial inspection of the between group clustering it is very clear that the deposition site materials group separately from the vehicle sample analyses. Two samples from the area of the deposition site designated samples 7 and 11 (see figure 4) do however group independently of the others from the body deposition site location. Visual inspection of figure 3 data charts identifies that deposition site 11 contained an anomalous iron content (75ppm in contrast to the body deposition site mean value of 30ppm). Since some small iron nodules, common in British lowland soils, were found in low concentrations in each of the samples and combined with the fact that only 0.15g of sample was used in the analysis, it was felt that this minor constituent may have skewed the resultant clustering. Other authors have identified the problems of analysing very small sub-samples by chemical analysis and have debated the representative nature of the results with bulk samples and their relevance therefore, in forensic enquiry (Jarvis et al 2004 and Rawlins and Cave 2004). By replacing the concentration value of 75ppm found in the sample from deposition site 11 with the mean value for the body deposition site samples, the recalculation of the hierarchical cluster moved the sample into the middle of the body deposition site grouping (see figure 4 arrow: no respected forensic scientist would manipulate the data in this way for court but it was undertaken for this paper to illustrate the fragility of any such derived sample groupings based on the presence or absence of perhaps one iron nodule which may not be picked up in every subsample of 0.15g used for analysis).

Similar anomalies were identified with deposition site 7 (figure 3) where the substitution of a reading of 10ppm for calcium was replaced with the mean value for the deposition site (2ppm). This resulted in the reclassification of the sample into the
grouping which contained all the other deposition site samples (see figure 4 arrow). Similar exercises could be undertaken with many other elements and samples, but this example simply serves to illustrate the fragility of the groupings and indeed the exclusionary conclusions inherent more in the analysis of small sub-samples (homogenised or not) than in any vagaries of the particulate multivariate statistical analysis employed. Using only chemical analysis, the hierarchical clustering of the vehicle samples suggests that the vehicle would be unlikely to have been involved with transporting any persons who had clambered up a muddy slope and returned in pitch black, in haste, into the van.

Contained also within figure 4 are a number of samples provided for analysis in the laboratory taken from the clothing of the victim (three in number) and from a pair of shoes belonging to the suspect (two in number). Were we to rely entirely upon the chemical analysis we would conclude that the shoes of the suspect were grouped within the same clustering as all of the deposition site samples, whilst the clothing of the victim was found to contain materials grouped both with the deposition site and the motor vehicle. The former scenario would be expected, the latter would prove somewhat problematic to explain, given the variance between the groupings from the motor vehicle and from the body deposition site. Further manipulation of the dendrogram (figure 4) was undertaken, whereby the shoes belonging to the suspect and the clothes from the victim were removed from the data set and then the cluster was recalculated. The resultant groupings (figure 5) show that four main groups of samples are identified in contrast to the groupings presented in figure 4. Figure 5 shows two groups of samples taken from the deposition site and two groups of samples taken from the vehicle. Indeed, it suggests that one of the groups of
sediments taken from the vehicle is less dissimilar to the first group of samples from the deposition site than from the second clustering of samples taken from the deposition site. This cluster analysis presents a less clear distinction between the groups of samples than those calculated by the hierarchical clustering technique presented in figure 4. This statistical grouping technique, intended as a relevant summary for a jury, is extremely dependent upon the number and type of samples added or removed from the clustering exercise. Whether or not the clustering is also dependent on the number of cations and anions analysed is tested with reference to the ICP-MS data utilising the analysis of the composition of 13 elements in soil taken from the same sample locations. The same hierarchical cluster technique was employed for the ICP-MS data and the results are presented in figure 6. This dendrogram clusters different samples together compared with that achieved by AAS/Dionex analysis (figure 4). Broadly, a large cluster of deposition site material is interspersed with samples taken from the clothing of the victim and the vehicle thought to have been used to transport the victim. A second cluster group contains a mixture of materials taken from the suspect, the deposition site and the victim. Four other samples do not cluster with any others or each other and these are taken from the victim, the deposition site, the suspect and the vehicle. Clearly, these results presented in figure 6 are at variance with those obtained from analysis of figure 4.

In order that statistical appraisal can be made of the visual variance in both figure 4 and figure 6, canonical discriminant function analysis (CDFA) was undertaken on the data used to compute the dendrograms. This CDFA is presented in figure 7, where discrimination is identifiable between the victim/site, van and suspect samples utilising AAS/Dionex results and similar discrimination is suggested by ICP-MS data
analysis. Complementing the graphical representation presented in figure 7 is the tabulation of derived statistical data presented in figures 8 and 9. Briefly, for the AAS/Dionex analysis, the eigenvalues demonstrate that the first variate accounts for 88.9% of the variance compared to the second and third variates which account for 9.6% and 1.6% (figure 8). The low Wilks’ Lambda values (less than 0.02) indicate that all three variates are significant in accounting for the groupings of samples (identified in figure 7A) at the 95% significance level. For the ICP-MS analysis (figure 9) the eigenvalues demonstrate that the three variates account for 53.0%, 33.1% and 14% of the variance respectively. The Wilks’ Lambda values are similarly low (less than 0.01) indicating that all three variates are significant in accounting for the groupings of samples (identified in figure 7B) at the 95% significance level.

The association of victim and site is understandable since the mud taken from the victim would appear to have derived from the muddy area in which she lay.

Thus, the CDFA plot for chemical analysis by AAS/Dionex (figure 7A) show that function 1 discriminates between the samples taken from the suspect with those taken from the victim, site and vehicle, whilst function 2 discriminates between the samples taken from the vehicle, with those taken from the site, victim and from the suspect. The CDFA plot for ICP-MS analysis shows that function 1 discriminates the site and victim samples from the vehicle and suspect samples, whilst function 2 discriminates between the samples taken from the vehicle with those taken from the site, victim and from the suspect. (It should be noted that there is a scale difference between figure 7A and figure 7B). Whether the discriminated samples have been identified as a result of their lack of similar provenance to the site and victim sample analysis suite is
difficult to ascertain. This statistically strong discrimination of vehicle and suspect is clear, but whether this is a false discrimination is a contention which can not be tested using these chemical data. Independent means of investigating the causes of the dissimilarity between samples could be attempted using other independent techniques which in this case would need to be either physical identification of sample characteristics or biological analysis (for example total mineralogy, palynology, environmental profiling or quartz surface texture analysis using SEM – these techniques are visually based and act as good independent means of investigating sample associations, or more importantly, exclusions). If the discriminated samples contain materials of the same provenance as that of the victim or the site and also contain pre- and/or post-forensic event materials this may well act as the discriminating feature in the group chemical analysis. This can not be tested using the chemical analysis techniques used in this report. What is clear is that the AAS/Dionex analyses provide a very broad and ready comparison with the ICP-MS data analyses but neither can be used to provide detailed discrimination in the comparison of samples taken from potentially four different provenances.

We may add therefore, another way of misinterpreting the results of analysis when dealing with geochemical comparisons of forensic samples in such murder scenarios. The calculation of clustering and discriminate techniques is controlled by the number of samples added or removed for comparison. This is also compounded by the number of chemical features measured in each sample. Truly objective interpretation of data can not depend upon a priori knowledge of filtering procedures, of chemical components to be analysed in each sample and the graphic displays presented in the high powered black-box programs. Such tools must be used with caution before a
judge and jury, lest they provide biased scenarios and false relative discriminations based solely on one somewhat black-box effect analysis of a homogenised micro-sample.

**Conclusions**

The examination of the case studies provided in this paper present not a totally nihilistic end product. We show that different chemical analysis techniques can provide broadly similar results which show the discrimination between groups of samples. Lessons to be learned however, are many and pertinent. Analytical procedures require an understanding of the trace or bulk nature of the samples to be analysed, the representativeness of the laboratory sub-sampling methodology and the need to homogenise samples during preparation for analysis. A very important decision has to be made concerning the particular chemical characteristics of cations, anions and other compounds which are included in the automatic analysis sequencing.

In reality however, the analyst requires *a priori* knowledge in selecting particular cation and anion analyses and of course this is not available to the forensic geoscientist in the way that it would be to the conventional geoscientist.

It is critical that the analyst understands, identifies and accounts for any mixing that has taken place in the samples to be analysed and with those samples with which comparisons were made. Chemical analyses will be correct (the array and number of cations and anions to be analysed will make no difference to the accuracy of the bulk results) but the interpretation of these data, whether undertaken in a multivariate form of simple graphic representation will depend entirely upon whether there has been any pre-, syn- or post-forensic event mixing of the sample. Associations between
samples, whether positive or negative, can not be tested by reference to the results obtained from other independent techniques. Such comparison will be meaningless. Rather, the results of the other techniques of analysis, if complementary to each other, will be a better indicator of association.

The meaningful presentation of multivariate interpretations of chemical data is dependent upon accurate sample and subsample collection. No amount of summary statistics, whether they be Principal Components Analysis, Canonical Discriminant Function analysis, or hierarchical clustering techniques will salvage the interpretation of the association between samples. Indeed, it can be shown that (even ‘logical’) manipulation of the raw data or the number of samples included in a cluster analysis can alter the subsequent groupings presented by multivariate analysis. Whilst we agree that no self respecting geoscientist would attempt to manipulate data in this way nor deliberately exclude samples from analysis, the fragility of the derived groupings and associations are plain for all to see. In nature, soil and sediment bodies, whether they be micrograms or kilograms in size are not homogeneous, careful sampling will still provide variation.

The interpretation of results will normally have to be undertaken with reference only to those samples provided in the case study investigated. Geochemical databases as exclusion samples simply do not exist at the level of chemical finesse required for chemical interpretation. Even given a constructed geochemical database it can be shown that the inclusion or exclusion of one or more unusual elements or compounds, of the exclusion of one or more samples from multivariate analysis can dramatically alter the results obtained. Whilst this may provide for a more logical statistical
grouping, better for clearer forensic explanation in the courtroom, it is tantamount to introducing the ‘fiddle factor’ and should have no place in objective forensic explanation.

Techniques which involve both chemical analysis and crucially, visual appraisal by the operator, have significant potential application in geoforensic enquiry (e.g. QemSCAN, see Pirrie et al. 2004). The requirement for independent analytical methods in successful data analysis is not restricted however, merely to chemical analysis investigations. It is a necessity to provide corroboration in forensic analysis lest interpretation be taken as estimation.

Acknowledgements

The authors would like to acknowledge discussion and comments with Christopher B Jackson, Dr Andy Field and Matt Morgan. RM acknowledges financial support from the Bruce, Julia and Mortimer May Senior Scholarship at Hertford College Oxford. The comments made by three anonymous referees proved extremely useful in helping us to clarify the presentation of this paper.
References


Ioannidis, J. P. A. 2005 Why most published research findings are false. Public Library of Science Medicine August 2005 2 (8): 0696-0701 (available at www.plosmedicine.org)


Lombardi, G. 1999 The Contribution of Forensic geology and Other Trace Evidence Analysis to the Investigation of the Killing of Italian Prime Minister Aldo Moro Journal of Forensic Sciences 44(3):634-642


List of Figures

Figure 1: Chemical analysis of the soil/sediment samples in Case study 1

Figure 2: Hierarchical cluster analysis of AAS/Dionex analysis of samples taken from Case Study 1.

Figure 3: Chemical analysis of the soil/sediment samples in Case study 2.

Figure 4: Hierarchical cluster analysis of AAS/Dionex analysis of samples taken from Case Study 2.

Figure 5: Hierarchical cluster analysis of AAS/Dionex analysis of deposition site and vehicle samples (without victim and suspect samples) taken from Case Study 2.

Figure 6: Hierarchical cluster analysis of ICP-MS analysis of samples taken from Case Study 1.

Figure 7: Canonical Discriminant Function plots of chemical analysis by AAS/Dionex (7A) and ICP-MS (7B).

Figure 8: Statistical output from Canonical Discriminant Function analysis of AAS/Dionex data

Figure 9: Statistical output from Canonical Discriminant Function analysis of ICP-MS data.