SUMMARY REPORT

SEPARATING ANTHROPOGENIC METALS CONTAMINATION FROM BACKGROUND: A CRITICAL REVIEW OF GEOCHEMICAL EVALUATIONS AND PROPOSAL OF ALTERNATIVE METHODOLOGY

Ecological Risk Assessment Support Center
Office of Research and Development
U.S. Environmental Protection Agency
Cincinnati, OH

NOTICE

Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Al   aluminum

CA   cluster analysis

Can 1   first canonical variable

Can 2   second canonical variable

Cd   cadmium

COPCs   Contaminants of Potential Concern

Cr   chromium

Cu   copper

DA   discriminant analysis

EPA   U.S. Environmental Protection Agency

Fe   iron

Mn   manganese

NRCS   Natural Resource Conservation Service

Pb   lead

QDA   quadratic discriminant analysis

USDA   U.S. Department of Agriculture

V   vanadium

Zn   zinc

AUTHORS, CONTRIBUTORS, AND REVIEWERS
(Reviewers of this Summary Report are marked with an asterisk)

AUTHORS

Richard H. Anderson[1]
U.S. Environmental Protection Agency
Office of Research and Development
National Center for Environmental Assessment
Cincinnati, OH 45268

CONTRIBUTORS

David B. Farrar
U.S. Environmental Protection Agency
Office of Research and Development
National Center for Environmental Assessment
Cincinnati, OH 45268

Sharon R. Thoms
U.S. Environmental Protection Agency
Region 4 Superfund Division
Atlanta, GA 30303

REVIEWERS

Brad Venner
U.S. Environmental Protection Agency
National Enforcement Investigations Center
Denver, CO 80225

Evan Englund
U.S. Environmental Protection Agency
Environmental Response Team - West
Las Vegas, NV 89119

---

[1] Permanent Address: Air Force Civil Engineer Center (AFCEC), Technical Support Branch, Lackland AFB, TX.

AUTHORS, CONTRIBUTORS, AND REVIEWERS (continued)

Robert G. Garrett
Geological Survey of Canada
Natural Resources Canada
Ottawa, Ontario K1A 0E8

Lena Q. Ma
University of Florida
Gainesville, FL 32611

Fangjie J. Zhao
Rothamsted Research
Hertfordshire AL5 2JQ, UK

Michael Griffith*
U.S. Environmental Protection Agency
Office of Research and Development
National Center for Environmental Assessment
Cincinnati, OH 45268

Tim Frederick*
U.S. Environmental Protection Agency
Region 4 Superfund Division
Atlanta, GA 30303

ACKNOWLEDGMENTS

PREFACE

Meaningful estimates of background contaminant levels are a critical component of site assessments. A request was submitted by the Ecological Risk Assessment Forum (ERAF) to the Office of Research and Development's Ecological Risk Assessment Support Center (ERASC) relating to the issue of background soil trace metal and metalloid (hereafter referred to as "chemical") demarcation at metals contaminated sites. Specifically, the request pertained to the validity of an empirical methodology (geochemical association plots) that utilizes covariation between chemical concentrations and concentrations of major soil elemental constituents (i.e., reference metals) to identify samples that deviate from "natural" variation.[2] Consequently, a comprehensive investigation of this methodology was conducted and assumes assessments are conducted with chemical and reference metal data collected from reference sites (i.e., background data) and site-related locations. This document summarizes the results of this investigation as described in the ERASC draft response and two peer-reviewed articles (Anderson and Kravitz, 2010, and Anderson et al., 2009). Part 1 of the document tests chemical/reference metal associations among uncontaminated soils of contrasting mineralogy and chemical/physical composition to help determine the extent of compatible background data sets. Chemical/reference metal associations are shown to vary significantly among background data sets. Thus, geochemical association plots are a useful screening tool for environmental site assessments, but ubiquitous application of generic background data sets could result in erroneous conclusions. Additional methodologies are needed as objective lines of evidence to conclude that a chemical occurs as site-related contamination. Part 2 of the document proposes a novel application to environmental site assessments. This application uses a multivariate-analysis methodology utilizing discriminant analysis with clustered chemical concentrations to determine, in relative order of magnitude, contaminated chemicals.

---

[2]Trace metals and metalloids (i.e., chemicals) are minor constituents of many geologic materials. In general, cationic metals can be found in a variety of silicate and aluminosilicate minerals such as olivines, amphiboles, micas, and feldspars (Wilson et al., 2008). In contrast, high levels of metalloids can be found in sedimentary deposits associated with pyritic sulfur-containing materials (Chen et al., 2002; Strawn et al., 2002). Metamorphic rocks comprised of serpentine minerals are also associated with high background chemical levels (Burt et al., 2001; Lee et al., 2001). Because of the variety of sources, regional background chemical levels can vary by up to three orders of magnitude (Gustavsson et al., 2001; Wilson et al., 2008).

EXECUTIVE SUMMARY

Empirical associations among trace metals and a major (i.e., reference) soil elemental constituent, such as iron (Fe), are used during environmental site assessments to screen for contaminants of potential concern (COPCs). These "geochemical association plots" use empirical log-log relationships to discern sites with naturally elevated chemical levels from sites with anthropogenic contamination. Point of fact, log-log relationships have been consistently observed between chemicals and reference metal concentrations and are often implicitly assumed to be constant. This investigation tests that assumption by using a regional geochemistry data set to evaluate background chemical/Fe log-log associations across soils with highly diverse composition. The results indicate that although geochemical associations may be proportional, they differ statistically across predominant U.S. Department of Agriculture (USDA) soil orders. Also, intraorder variability in geochemical ratios generally ranged multiple orders of magnitude, which suggests that the order level of the USDA soil taxonomic system is insufficient to reasonably classify background chemical concentrations. Geochemical association plots are a useful screening tool for environmental site assessments, but ubiquitous application of generic background data sets could result in erroneous conclusions (Anderson and Kravitz, 2010).

Reference soils are used to define baseline chemical values during remedial investigations and are selected based on professional judgment, usually predicated on factors such as soil type, proximity to site, topoedaphic landscape position, and habitat. Often, however, representative reference soils are difficult to delineate. One could argue that a preferred methodology for determining COPCs should be independent of the assessor and capable of meaningful interpretation despite potential bias from incompatible reference soils or when reference soils cannot be collected at all (e.g., logistic or monetary constraints). Additional methodologies are needed as objective lines of evidence for concluding a chemical occurs as site-related contamination—especially for sites with spatially heterogeneous soil composition where soil matrix composition can be highly variable.

A multivariate-analysis methodology utilizing discriminant analysis with clustered chemical concentrations is proposed as a novel application to environmental site assessments that determine, in relative order of magnitude, contaminated chemicals. Finite mixture models are

presented as a means to assess latent chemical clusters with some basis in statistical inference. The methodology is illustrated with a typical localized data set containing total chemical concentrations, extracted from bulk soil collected from reference (i.e., background data) and site-related locations, obtained from a former military installation in the southeast United States. The illustration is particularly applicable because site-related soils inherently possessed higher background chemical levels relative to reference soils, which biases conventional analyses. However, two distinct chemical signatures were observed within site-related samples illustrating the versatility of the proposed methodology. Using these results along with known information regarding the history of contamination at the site, a qualitative and quantitative assessment of contaminated chemicals was made. Results are intended for illustration purposes only and are discussed within the context of environmental site assessment.

1. EVALUATION OF GEOCHEMICAL ASSOCIATIONS AS A SCREENING TOOL FOR IDENTIFYING ANTHROPOGENIC CHEMICAL CONTAMINATION

1.1 INTRODUCTION

The focus of this investigation is the bivariate analysis between soil chemical concentrations and concentrations of major soil elemental constituents (i.e., reference metals), collectively referred to as a geochemical evaluation (NAVFAC, 2003). Geochemical evaluations are currently used to help determine contaminants of potential concern (COPCs) at some metals-contaminated sites (NAVFAC, 2003; U.S. EPA, 2007) and are predicated on the covariation between chemicals and reference metals, which have been repeatedly documented (Hamon et al., 2004; Myers and Thorbjornsen, 2004; Thorbjornsen and Myers, 2007a,b). In this approach, chemicals that produce trends with anomalous observations are assumed to occur as site-related contamination. All uncontaminated soil samples are implicitly assumed to follow similar, if not consistent, trends (Hamon et al., 2004; Myers and Thorbjornsen, 2004; Thorbjornsen and Myers, 2007a,b). Although the validity of this assumption has been empirically supported (Hamon et al., 2004), it has not been formally tested.

Soil matrix composition may modify geochemical associations. Although formal causality cannot be inferred from scatter plots alone, covariation between chemicals and reference metals reflect soil mineral composition (Lopez et al., 2005) and soil chemical sequestration due to chemical adsorption to soil solid phases (Sparks, 2003). However, total soil chemical concentrations are ambiguous in that they provide no information regarding the solid phase from which the analyte was extracted.

Variable geochemical associations could occur across soils with contrasting siliceous and hydrous oxide clay minerals. For example, chemical reactivity (i.e., binding capacity) of hydrous oxides is influenced by the degree of crystallinity, which is influenced by various pedogenic (i.e., soil forming) factors (Gerth, 2005). Chemicals also have variable affinities for different hydrous oxide minerals (Huelin et al., 2006; Sparks, 2003). Similarly, cationic trace metal affinity for the exchange complex of negatively charged siliceous clay minerals increases with surface area, which can be regulated by clay mineral structure and expandability rather than total clay content (Meier and Kahr, 1999). Thus, anomalous observations among geochemical

association plots may not necessarily indicate anthropogenic contamination. Further investigation is required to determine applicable background data sets for geochemical evaluations of sites based on log-log trace metal/reference metal associations.

Central to the issue of identifying contaminated samples is identifying appropriate background chemical concentrations fundamental to geochemical evaluations. Co-contamination of reference metals and spatial heterogeneity could impede accurate conclusions and result in statistical uncertainty. Without representative background data or with a predominance of suspect site-related data, geochemical evaluations could result in erroneous conclusions if, on average, site-related data were collected at a scale impacted by contamination (Molinaroli et al., 1999; Kazemi et al., 2008; Koptsik et al., 2003; Vanderlinden et al., 2006). *Sufficient representative background chemical data are fundamentally critical to geochemical evaluations.* Consequently, the objective of this investigation was to test chemical/reference metal associations among uncontaminated soils of contrasting mineralogy and chemical/physical composition to help determine the extent of compatible background data sets. A regional geochemistry data set, obtained from the USDA Natural Resource Conservation Service (NRCS) Cooperative Soil Survey Program, was used to test the implicit assumption that all uncontaminated soil samples follow similar, if not consistent, trends (Hamon et al., 2004; Myers and Thorbjornsen, 2004; Thorbjornsen and Myers, 2007a,b). In this context, the central tendencies of the bivariate relationships are evaluated among dissimilar soils in terms of their estimated Y-intercept and slope values. Log-linear relationships between selected chemicals and Fe, a surrogate soil reference metal, were analyzed for selected chemicals.

1.2 STATISTICAL ANALYSES

Log-log trace metal/reference metal associations were evaluated across individual pedons of the same soil series (i.e., the finest level of US taxonomic classification) and analyzed for differences across nine predominant soil orders (i.e., the broadest level of US taxonomic classification). Further evaluations on sub-order classifications were operationally difficult because most soils were not classified to sub-order levels. Proportionality of geochemical log–log associations among soil orders was evaluated by analysis of covariance (ANCOVA) with a

model described in Anderson and Kravitz, 2010. See that publication for further detail on statistical analyses.

1.3 RESULTS AND DISCUSSION

Soils derived from geologic materials with appreciable trace metal concentrations retain a similar chemical composition throughout early genesis. However, as soils weather and undergo pedogenesis (i.e., chemical/physical transformations), there is a redistribution of analytes, organic matter accrual, and secondary (siliceous and hydrous oxide) clay mineral precipitation resulting in soil chemical/physical properties with variable similarities to that of the original parent material (Minasny et al., 2008). Organic matter and secondary clay minerals both affect trace metal retention through ion exchange reactions and semi- to highly stable covalent binding (Sparks 2003). Hence, a highly diverse population of reference soils comprised of contrasting chemical/physical properties was essential to comprehensively evaluate geochemical trace metal/reference metal log–log associations.

Geochemical trace metal/reference metal log–log associations are usually illustrated with Fe, Al, or Mn as the representative soil reference metal. Log-linear relationships between Fe and Al concentrations and Fe and Mn concentrations were plotted by Anderson and Kravitz, 2010 (see Figure 1 in that paper). Because significant associations were observed, trace metal/reference associations were evaluated with Fe as the surrogate reference metal.

Log–log plots are used to compress metal distributions in order to obtain linear relationships and assume log-normality for statistical inference. Log–log plots between selected trace metals and total soil Fe concentrations are presented in Figure 1. Obvious trends are readily apparent with the exception of Cd (and to a lesser extent Pb) where relationships are highly variable. Significant interaction was observed for all metals except Cd (Table 2 in Anderson and Kravitz, 2010). Interaction indicates a departure from parallelism suggesting disproportional relationships among predominant soil orders. However, as discussed in Anderson and Kravitz, 2010, evidence of disproportionality is marginal at best and can be seen by evaluating trends in Figure 1. Thus, further evaluation using a method robust to the influence of sample size was necessary.

Log trace metal/Fe ratios were calculated according to Eq. 3 in Anderson and Kravitz, 2010. Figure 2 shows box plots illustrating distributions of ratios among predominant soil orders. Interestingly, ratios generally ranged multiple orders of magnitude within individual soil orders except for V, which only ranged two orders of magnitude reflecting the lack of variability observed in Figure 1. Numerous extreme data points were observed (i.e., shaded circles) and were determined as data below and above the 5th and 95th percentiles, respectively. Distributions of all ratios generally fell on the same scale—except for Cd, which was considerably lower (Figure 2).

Geochemical ratios were evaluated at the order level of USDA soil taxonomy to specifically test the broadest hierarchical level of soil classification. Similar ratios across predominant soil orders would support ubiquitous application of generic background data sets for site assessments based on geochemical evaluations. However, nonconstant ratios were observed as evidenced by statistically significant differences among soil orders for all chemicals evaluated (Table 1). These results suggest that soil composition affects the ratio of selected chemicals to Fe. Thus, although geochemical associations may be proportional (Figure 1), the application of generic background data sets could result in Type II or Type I statistical errors. Alternatively stated, geochemical evaluations can result in the identification of anthropogenic contamination as background (contaminated observations at or below the trend line) or identification of background as anthropogenic contamination (background observations above the trend line).

The variability observed in trace metal/Fe ratios within soil orders suggests that the order level of soil classification may be too broad of a category to determine applicable background data sets for site assessments based on geochemical association. The USDA Soil Taxonomy system sub-classifies soils on the basis of mineralogy only at the family (i.e., sub-order) level and may provide a better hierarchical scale at which to evaluate trends in geochemical ratios (NRCS, 2003). Hence, further study is required to explicitly determine the taxonomic level of classification where constant trace metal/reference metal ratios occur.

Geochemical surveys have repeatedly shown that background chemical concentrations strongly depend on geologic and pedogenic processes (Burt et al., 2003; Gustavsson et al., 2001; Klassen, 1998; Miretzky et al., 2001; Wilson et al., 2008). Overall, this investigation suggests that parent material composition and pedogenic factors may all influence geochemical ratios given the diversity of soils evaluated (Burt et al., 2003; Wilson et al., 2008). Assuming consistent

effects of geologic and pedogenic processes on geochemical ratios during environmental site assessments is an over simplification and may result in erroneous identification of anthropogenic contamination. Site-specific background chemical data should be utilized if ample reference sites with similar (i.e., suborder) soil composition can be identified and sampled.

1.4 RATIONALE FOR ALTERNATIVE METHODOLOGY

Although geochemical evaluations provide a tool for assessing potential soil contamination, this investigation has quantitatively illustrated that results can be biased if site-related soils and/or reference soils are of dissimilar composition, although further study is required to explicitly determine the appropriate taxonomic level of classification for geochemical evaluations. Reference soils for background data sets are selected at the discretion of the assessor and based on professional judgment usually predicated on factors such as site geology, soil type/parent material, proximity to site, topoedaphic landscape position, and habitat. One could argue that a preferred methodology should be independent of the assessor and capable of meaningful interpretation despite potential bias from incompatible reference soils or when reference soils cannot be collected at all (e.g., logistic or monetary constraints). Additional methodologies are needed as objective lines of evidence for concluding a chemical occurs as site-related contamination—especially for sites with spatially heterogeneous soils where soil matrix composition can be highly variable.

Multivariate analysis techniques can simultaneously evaluate multiple chemicals and allow the development of a site-specific chemical profile that has been referred to as a signature (Ridgway et al., 2003). Multivariate chemical signatures have been used to identify contaminated sites through techniques such as robust multivariate outlier detection (Filzmoser et al., 2005), principal component analysis (Imrie et al., 2008; Korre, 1999a), and factor analysis (Dragovic et al., 2008; Korre, 1999a; Reimann et al., 2002; Vaccaro et al., 2007). Formal inference among these procedures assumes multivariate lognormality out of convenience because it leads to chi-square distributed distance values for significance tests (Filzmoser et al., 2005).

Univariate and bivariate applications are perfectly acceptable exploratory analyses for some situations, but multivariate methods are relatively more comprehensive and can reveal features of environmental data that univariate and bivariate methods cannot. The methodology

summarized here, from Anderson et al., 2009, is emphasized for sampling schemes that collect a large percentage of contaminated samples making it difficult to separate outliers from population mixtures. In this context, a method capable of separating samples into categories that represent sites with relatively homogeneous multivariate signatures seems logical for reliable inference.

2. APPLICATION OF DISCRIMINANT ANALYSIS WITH CLUSTER ANALYSIS TO DETERMINE ANTHROPOGENIC CHEMICAL CONTAMINATION

2.1 INTRODUCTION

Cluster analysis (CA) is an assortment of multivariate data analysis methods that aim to discover groups of similar observations. A considerable variety of CA techniques have been proposed, as reviewed in specialized texts (Everitt, 1993; Gordon, 1999) as well as general texts on multivariate data analysis (Rencher, 2002; Seber, 1984; Venables and Ripley, 1998). Many CA methods summarize information in the form of a matrix giving similarities (or distances) relating pairs of observations, as determined using some appropriate index, and generate classifications such that observations in a given class are more similar to one another (as measured using the chosen index) than to observations in different classes. Such methodology has been driven particularly by the needs of biological taxonomy and ecological community analysis. However, applications to environmental site assessments have increased in recent years (Dragovic et al., 2008; Martinez et al., 2007; Mico et al., 2006; Molinaroli et al., 1999).

In addition to similarity-based procedures, there are recent methods of model-based CA, which are based on distributional models (Fraley and Raftery, 2002; Fruhwirth-Schnatter, 2006). Basing CA on a probability model leads to likelihood-based statistical inferences, jointly for classifying observations and estimating parameters of class-specific distributions. In practice, model-based CA is usually based on the theory of finite mixtures (Fruhwirth-Schnatter, 2006; Ter Braak et al., 2003; Yang and Chang, 2005). A model-based approach assumes that sites can be grouped into classes that can be described with distinct multivariate models (Banfield and Raftery, 1993). Results from model-based clustering may be directly useful in the context of environmental site assessments possibly leading to discovery of heterogeneity within and/or among reference and/or site-related sample soil populations that could cast doubt on standard

statistical comparisons. Multivariate application of model-based CA to environmental site assessment remains uninvestigated.

Discriminant analysis (DA) is commonly applied with samples grouped using CA, in order to evaluate the variables that account for cluster differences. DA methods are based on assumptions similar to those of a finite mixture of multivariate normal distributions. The primary advantage of DA is that the classification of samples is used to derive a linear combination of the original variables (i.e., the discriminant function) that can serve to discriminate among clusters. Canonical DA allows the identification of a few, orthogonal discriminant functions based on the number of classifications, which can be determined by CA (see Lambrakis et al., 2004; Mariner et al., 1997; Petalas and Anagnostopoulos, 2006; Sielaff and Einax, 2007). Despite an explosion of data-mining methodologies, basic DA methods continue to perform a critical role in many disciplines, particularly for small data sets (Hastie et al., 2009).

The remainder of this paper illustrates the application of CA and DA to environmental site assessments. The methods are illustrated with a typical localized data set containing total soil chemical concentrations, collected from reference and site-related locations, obtained from a small arms munitions firing range in the southeast United States. Specifically, objectives are to (1) illustrate the application of CA to define multivariate chemical signatures and (2) illustrate the use of DA to characterize differences in chemical signatures among clusters identified with CA. Results are discussed within the context of environmental site assessment.

2.2 MATERIALS AND METHODS

The proposed methodology is illustrated with field data collected during the remedial investigation of Ft. McClellan, a former military installation located in Alabama, USA. Ft. McClellan was closed under the Base Realignment and Closure program in 1999. The Fort consisted of the Main Post, Pelham Range, and Choccolocco Corridor covering more than 18,000 ha (Figure 3). Firing ranges located in the Choccolocco Corridor (hereafter referred to as "Site") were used for military small arms training during World War II, the Korean War, and the Vietnam War, and the potentially affected area was subject to this investigation. The range was abandoned in 1974 and mostly restricted to small arms training.

The major soil complexes at the Fort are described in Anderson et al., 2009. Most of the reference samples associated with the existing data set came from the Main Post. Other reference samples came from the Pelham Range (Figure 3).

Soil samples were collected and stratified by depth using a stainless steel hollow stem hand auger in accordance with ASTM method D1452 (ASTM, 2009). Only samples collected from 0–30 cm were used for statistical analyses. Site-related soil samples were analyzed for the Target Analyte List (TAL) by U.S. EPA SW-846 analytical methods (U.S. EPA, 2018). All chemicals, except mercury, were analyzed with Inductively Coupled Plasma-Atomic Emission Spectroscopy (ICP-AES). Reference soil samples were also analyzed for TAL chemicals using ICP-AES. Further information on soil collection and analytical measurements techniques are described in Anderson et al., 2009.

Hierarchical cluster analysis (CA) was used to assess patterns of geochemical association. Results of this approach are displayed by a tree-like structure called a dendrogram. A classification with a specified number of clusters was obtained from the dendrogram. Model-based CA was implemented using the R package mclust Version 3 (Fraley and Raftery 2002, 2006) with options set to defaults.

Multivariate DA was used to summarize between-cluster variability as a function of all measured chemicals. Because results from model-based CA suggested three components, DA was performed with the three clustered chemical groups (i.e., signatures) allowing estimation of within-cluster covariances. Assuming equal covariance matrices for each cluster leads to the method of linear discriminant analysis, while quadratic discriminant analysis (QDA) allows for different covariance matrices. However, discriminant plots for QDA are currently a topic of statistical research (Khattree and Naik, 2000; Pardoe et al., 2007). Discriminant plotting based on linear DA gave highly interpretable results, despite results from model-based clustering suggesting differences among covariance matrices. As a result, the structure of the standardized linear discriminant function of interest (defined for each chemical as the linear correlation coefficient between the canonical scores and actual values) was evaluated in order to determine, in relative order of magnitude, the chemicals that best distinguished between inferred background and contaminated signatures. Multivariate DA was conducted using PROC DISCRIM in SAS Version 9.2 for Windows.

2.3 RESULTS AND DISCUSSION

Summarized total chemical concentrations for both reference and Site-related sample soil populations are presented in Table 2. All mean cationic metal levels are elevated in Site-related soils relative to reference soils, while the opposite is true for the anionic species (arsenic, chromium, and vanadium). This suggests dissimilarity in the chemical/physical matrices of the two sample populations, which can be assessed by evaluating relationships between concentrations of dominant clay mineral constituents such as aluminum (Al) and iron (Fe). Figure 4 shows elevated Al concentrations in Site-related soils relative to Fe concentrations. Significantly ($p < 0.0001$) different intercepts, as determined by Analysis of Covariance, verify that differences in soil matrices are in part responsible for elevated cationic trace metal concentrations in Site-related soils. This is because 1) Al is a dominant constituent of the clay minerals involved in cationic metal sequestration (Sparks, 2003) and 2) neither Al nor Fe contamination has ever occurred at the Fort. Consequently, conventional assessment methodologies (e.g., univariate tests, tests of location, and bivariate associations) would likely result in high false-positive error. However, assessment of latent chemical signatures can still be assessed illustrating the versatility of CA.

2.3.1 Cluster Analysis

All data were subjected to hierarchical CA to assess the empirical pattern of multivariate association among chemical concentrations. A dendrogram depicting clusters that minimize within-cluster sum of squared error at all possible cluster configurations is shown in Figure 5. The three cluster configuration accounted for over half the variability in chemical concentrations (R-square=0.652). Subsequent cluster configurations accounted for increasingly less variability (Figure 5).

2.3.2 Discriminant Analysis

Once divergent chemical signatures have been determined and have been associated with a contamination event, the objective is to determine the chemical or chemicals responsible for the divergent signatures. Obviously, simple statistics (mean, min, and max) for each chemical

within each observed signature are relevant diagnostics. However, for purposes of identifying multivariate signatures, DA provides quantitative and graphical results that are particularly suited for evaluating the signatures that best account for cluster differences. While the mclust software includes DA capabilities, for present, illustrative purposes we prefer the current relatively standard approach based on use of specialized canonical DA software.

The discriminant plot of the standardized canonical coefficients is presented in Figure 6. Notably apparent is the vertical separation of all reference and Site-related samples (signatures 1 and 2 vs. 3) characterized in the first canonical variable (Can 1). Thus, Can 1 quantifies the degree to which reference and Site-related samples differ in chemical concentrations, which we argue to be the result of differences in soil matrix composition (Figure 2). Can 1 accounted for the vast majority (92.5%) of the overall variability in chemical concentrations. However, the second canonical variable (Can 2) was still highly significant (p <0.0001) and resulted in roughly equal separation of Site-related samples in the horizontal space. Thus, although minor compared to the matrix effect, substantive differences in chemical concentrations were also evident between Site-related signatures quantified by Can 2 (Figure 6).

Of primary importance to environmental assessors is the ability to determine contaminants among a suite of measured chemicals, with the most important differences between site-related and reference subsets. Typical assessments involve an independent evaluation of each chemical using the previously discussed methodologies (NAVFAC, 2003). Such procedures generate large quantities of output, which can be labor intensive to evaluate. Multivariate procedures, in general, and our proposed CA with subsequent DA approach provide a simultaneous assessment based on information from multiple chemicals. Using the structure of the relevant canonical variable(s) along with known information regarding the history of contamination events at a site, both a qualitative and quantitative assessment of contaminated chemicals can be made. By evaluating the structure of Can 2, the relative order of elevated chemicals (i.e., chemicals above background) can be determined and is illustrated in Figure 7. Cu, Pb, and Zn were determined as the three most elevated chemicals, respectively, and are consistent with contamination associated with small arms firing ranges. Combining site history information with these results, one could conclude that Cu, Pb, and Zn are contaminated chemicals, while the other chemicals are within background ranges.

2.3.3 Potential Applications

A common deliverable from studies designed to investigate the nature and extent of contamination are geographic maps of chemical concentrations. Spatial interpolation methods such as kriging are commonplace in Geographic Information Systems and are used to estimate spatially continuous chemical distributions within a discretely sampled area. Usually, however, chemical distributions are mapped individually without regard to co-occurring chemicals (Johnson and Ander, 2008). On the other hand, multivariate assessment of complex chemical signatures can consolidate output necessary for decision making (Korre, 1999b). For example, classification output from a CA has been used to map multivariate chemical signatures for predetermined grid cells within a study site (Martinez et al., 2007). Alternatively, spatial interpolation methods can be applied to discrete variables (Mancho et al., 2006), such as categories of chemical concentrations for continuously distributed maps. Another objective of environmental site assessments is the development of site-specific benchmarks of chemical concentrations that delineate the upper limit of background values. Typically, background benchmarks are estimated as the upper tolerance limit (90% coverage and 95% confidence) of a reference population of soil samples (U.S. EPA, 2007). Reference soils are selected at the discretion of the assessor and based on professional judgment usually predicated on factors such as proximity to site, topoedaphic landscape position, and habitat. Inherent in their application is the assumption that reference and site-related soils are of similar composition which is contradicted in the current study (Figure 4). Thus, methodologies are needed that estimate background benchmarks where incompatible reference soils are inadvertently collected (e.g., the current study) or when reference soils can not be collected at all (e.g., monetary or logistic constraints).

CONCLUSION

Chemical/reference metal associations among uncontaminated soils of contrasting mineralogy and chemical/physical composition were tested to help determine the extent of compatible background data sets. Chemical/reference metal associations were shown to vary significantly among background data sets. Thus, geochemical association plots are a useful

screening tool for environmental site assessments, but ubiquitous application of generic background data sets could result in erroneous conclusions. Additional methodologies are needed as objective lines of evidence to conclude that a chemical occurs as site-related contamination. A multivariate-analysis methodology utilizing discriminant analysis with clustered chemical concentrations is proposed as a novel application to environmental site assessments that determine, in relative order of magnitude, contaminated chemicals. The methodology is illustrated with a typical localized data set - from a former military installation in the southeast United States - containing total chemical concentrations, extracted from bulk soil collected from reference (i.e., background data) and site-related locations. Site-related soils inherently possessed higher background chemical levels relative to reference soils, yet two distinct chemical signatures were observed within site-related samples, illustrating the versatility of the proposed methodology. Using these results, along with known information regarding the history of contamination at the site, a qualitative and quantitative assessment of contaminated chemicals was made. Copper, lead, and zinc were the contaminated chemicals, while the other chemicals were within background ranges.

REFERENCES

Anderson, R.H., D.B. Farrar, and S.R. Thoms. 2009. Application of discriminant analysis with clustered data to determine anthropogenic metals contamination. Sci. Total Environ. 408:50-56.

Anderson, R.H. and M.J. Kravitz. 2010. Evaluation of geochemical associations as a screening tool for identifying anthropogenic trace metal contamination. Environ. Monit. Assess. 167:631-641.

ASTM. 2009. Standard practice for soil exploration and sampling by auger borings.

Banfield, J.D. and A.E. Raftery. 1993. Model-based Gaussian and non-Gaussian clustering. Biometrics. 49:803−821.

Burt, R., M. Fillmore, M.A. Wilson, E.R. Gross, R.W. Langridge, and D.A. Lammers. 2001. Soil properties of selected pedons on ultramafic rocks in Klamath Mountains, Oregon. Comm. Soil Sci. Plant Anal. 32(13):2145−2175.

Burt, R., M.A. Wilson, M.D. Mays, and C.W. Lee. 2003. Major and trace elements of selected pedons in the USA. J. Environ. Qual. 32(6):2109−2121.

Chen, M., L.Q. Ma, and W.G. Harris. 2002. Arsenic concentrations in Florida surface soils: Influence of soil type and properties. Soil Sci. Soc. Am. J. 66:632−640.

Dragovic, S., N. Mihailovic, and B Gajic. 2008. Heavy metals in soils: Distribution, relationship with soil characteristics and radionuclides and multivariate assessment of contamination sources. Chemosphere. 72(3):491−495.

Everitt, B. 1993. Cluster Analysis. Edward Arnold, London.

Filzmoser, P., R.G. Garrett, and C. Reimann. 2005. Multivariate outlier detection in exploration geochemistry. Comp. Geosci. 31:579−587.

Fraley, C. and A.E. Raftery. 2002. Model-based clustering, discriminant analysis, and density estimation. J. Am. Stat. Assoc. 97:611−631. Available at http://www.stat.washington.edu/raftery/Research/PDF/fraley2002.pdf.

Fraley, C. and A.E. Raftery. 2006. MCLUST Version 3 for R: Normal Mixture Modeling and Model-based Clustering, Technical Report No. 504. Department of Statistics, University of Washington.

Fruhwirth-Schnatter, S. 2006. Finite mixture and Markov switching models. Springer, New York.

Gerth, J. 2005. Effects of crystal modification on the binding of trace metals and arsenate by goethite. J. Soil Sed. 5:30−36.

Gordon, A.D. 1999. Classification. Chapman and Hall/CRC, New York.

Gustavsson, N., B. Bolviken, D.B. Smith, and R.C. Severson. 2001. Geochemical Landscapes of the Conterminous United States: New Map Presentations for 22 Elements. Department of the Interior, USGS Report # 1648, Denver, CO. Available at http://pubs.usgs.gov/pp/p1648/.

Hamon, R.E., M.J. McLaughlin, R.J. Gilkes et al. 2004. Geochemical indices allow estimation of heavy metal background concentrations in soils - art. No. GB1014. Global Biogeochem. Cyc. 18:B1014.

Hastie T, R. Tibshirani and J. Friedman. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd. Ed. Springer Series in Statistics. Springer, NY.

Huelin, S.R., H.P. Longerich, D.H.C. Wilton, and B.J. Fryer. 2006. The determination of trace elements in Fe-Mn oxide coatings on pebbles using LA-ICP-MS. J. Geochem. Explor. 91(1−3):110−124.

Imrie, C.E., A. Korre, G. Munoz-Melendez, I. Thornton, and S. Durucan. 2008. Application of factorial kriging analysis to the FOREGS European topsoil geochemistry database. Sci. Total Environ. 393:96−110.

Johnson, C.C. and E.L. Ander. 2008. Urban geochemical mapping studies: How and why we do them. Environ. Geochem. Health. 30(6):511−530.

Kazemi, H.V., S.H. Anderson, K.W. Goyne, and C.J. Gantzer. 2008. Spatial variability of bromide and atrazine transport parameters for a Udipsamment. Geoderma. 144(3−4):545−556.

Khattree R. and D. Naik. 2000. Multivariate Data Reduction and Discrimination with SAS Software. SAS Press and John Wiley Sons Inc, New York.

Klassen, R.A. 1998. Geological factors affecting the distribution of trace metals in glacial sediments of central Newfoundland. Environ. Geol. 33(2−3):154−169.

Koptsik, S., G. Koptsik, S. Livantsova, L. Eruslankina, T. Zhmelkova, and Z. Vologdina. 2003. Heavy metals in soils near the nickel smelter: Chemistry, spatial variation, and impacts on plant diversity. J. Environ. Monit. 5(3):441−450.

Korre, A. 1999a. Statistical and spatial assessment of soil heavy metal contamination in areas of poorly recorded, complex sources of pollution Part 1: factor analysis for contamination assessment. Stoch. Environ. Res. Risk Assess. 13:260−287.

Korre, A. 1999b. Statistical and spatial assessment of soil heavy metal contamination in areas of poorly recorded, complex sources of pollution Part 2: canonical correlation analysis and GIS for the assessment of contamination sources. Stoch. Environ. Res. Risk Assess. 13:288-316.

Lambrakis, N., A. Antonakos, and G. Panagopoulos. 2004. The use of multicomponent statistical analysis in hydrogeological environmental research. Water Res. 38(7):1862−1872.

Lee, B.D., R.C. Graham, T.E. Laurent, C. Amrhein, and R.M. Creasy. 2001. Spatial distributions of soil chemical conditions in a serpentinitic wetland and surrounding landscape. Soil Sci. Soc. Am. J. 65:1183−1196.

Lopez, J.M.G., B. Bauluz, C. Fernandez-Nieto, and A.Y. Oliete. 2005. Factors controlling the trace-element distribution in fine-grained rocks: The Albian kaolinite-rich deposits of the Oliete Basin (NE Spain). Chem. Geol. 214(1−2):1−19.

Mancho, A.M., D. Small, and S. Wiggins. 2006. A comparison of methods for interpolating chaotic flows from discrete velocity data. Comp. Fluids. 35(4):416−428.

Mariner, P.E., F.J. Holzmer, R.E. Jackson, J.E. White, A.S. Wahl, and F.G. Wolf. 1997. Fingerprinting arsenic contamination in the sediments of the Hylebos waterway, Commencement Bay Superfund Site, Tacoma, Washington. Environ. Eng. Geosci. 3(3):359−368.

Martinez, J., J. Llamas, E. de Miguel, J. Rey, and M.C. Hidalgo. 2007. Determination of the geochemical background in a metal mining site: Example of the mining district of Linares (South Spain). J. Geochem. Explor. 94(1−3):19−29.

Meier, L.P. and G. Kahr. 1999. Determination of the cation exchange capacity (CEC) of clay minerals using the complexes of copper (II) ion with triethylenetetramine and tetraethylenepentamine. Clays Clay Min. 47(3):386−388.

Mico, C., L. Recatala, A. Peris, and J. Sanchez. 2006. Assessing heavy metal sources in agricultural soils of an European Mediterranean area by multivariate analysis. Chemosphere. 65(5):863−872.

Minasny, B., A.B. McBratney, and S. Salvador-Blanes. 2008. Quantitative models for pedogenesis - A review. Geoderma. 144(1−2):140−157.

Miretzky, P., V. Conzonno, and A.F. Cirelli. 2001. Geochemical processes controlling silica concentrations in groundwaters of the Salado River drainage basin, Argentina. J. Geochem. Explor. 73(3):155−166.

Molinaroli, E., M. Pistolato, G, Rampazzo, and S. Guerzoni. 1999. Geochemistry of natural and anthropogenic fall-out (aerosol and precipitation) collected from the NW Mediterranean: Two different multivariate statistical approaches. Appl. Geochem. 14(4):423−432.

Myers, J. and K. Thorbjornsen. 2004. Identifying metals contamination in soil: A geochemical approach. Soil Sed. Contam. 13(1):1−16.

NAVFAC (Naval Facilities Engineering Command). 2003. Guidance for environmental background analysis, Vol. II: Sediment. In: NFESC User's Guide UG-2054-ENV. Washington, DC. Prepared by Battelle Memorial Institute, Earth Tech, Inc., and New Fields, Inc. under Contract No. N47408-01-D8207.

NRCS (National Resource Conservation Service). 2003. Keys to Soil Taxonomy, 9th ed. U.S. Department of Agriculture, NRCS, Washington, DC.

Pardoe, I, X.R. Yin, and R.D. Cook. 2007. Graphical tools for quadratic discriminant analysis. Technometrics. 49(2):172−183.

Petalas, C. and K. Anagnostopoulos. 2006. Application of stepwise discriminant analysis for the identification of salinity sources of groundwater. Water Res. Mgmt. 20(5):681−700.

Reimann, C., P. Filzmoser, and R.G. Garrett. 2002. Factor analysis applied to regional geochemical data: Problems and possibilities. Appl. Geochem. 17(3):185−206.

Rencher, A.C. 2002. Methods of Multivariate Analysis, 2nd ed. Wiley, New York.

Ridgway, J., N. Breward, W.J. Langston, R. Lister, J.G. Rees, and S.M. Rowlatt. 2003. Distinguishing between natural and anthropogenic sources of metals entering the Irish Sea. Appl. Geochem. 18(2):283−309.

Seber, G.A.F. 1984. Multivariate observations, 1st ed. Wiley Series in Probability and Mathematical Statistics, V Barnett, R.A. Bradley, J.S. Hunter et al., Eds. Wiley, New York.

Sielaff, K. and J.W. Einax. 2007. The application of multivariate statistical methods for the evaluation of soil profiles. J. Soil Sed. 7(1):45−52.

Sparks, D.L. 2003. Environmental Soil Chemistry, 2nd ed. Academic Press, London, UK.

Strawn, D., H. Doner, M. Zavarin, and S. McHugo. 2002. Microscale investigation into the geochemistry of arsenic, selenium, and iron in soil developed in pyritic shale materials. Geoderma. 108(3−4):237−257.

Ter Braak, C.J.F., H. Hoijtink, W. Akkermans, and P.F.M. Verdonschot. 2003. Bayesian model-based cluster analysis for predicting macrofaunal communities. Ecol. Model. 160(3):235−248.

Thorbjornsen, K. and J. Myers. 2007a. Identifying metals contamination in groundwater using geochemical correlation evaluation. Environ. Foren. 8(1):25−35.

Thorbjornsen, K. and J. Myers. 2007b. Identification of metals contamination in firing-range soil using geochemical correlation evaluation. Soil Sed. Contam. 16(4):337−349.

U.S. EPA (Environmental Protection Agency). 2007. Framework for metals risk assessment. Office of the Science Advisor, Risk Assessment Forum, Washington, DC. EPA/120/R-07/001.

U.S. EPA. 2018. SW-846 Update VI On-Line. Chapter Three, Inorganic Analytes. Revision 6. https://www.epa.gov/sites/production/files/2019-06/documents/chapter_three_update_vi_12-11-2018.pdf

Vaccaro, S., E. Sobiecka, S. Contini, G. Locoro, G. Free, and B.M. Gawlik. 2007. The application of positive matrix factorization in the analysis, characterisation and detection of contaminated soils. Chemosphere. 69(7):1055−1063.

Vanderlinden, K., R. Ordonez, M.J. Polo, and J.V. Giraldez. 2006. Mapping residual pyrite after a mine spill using non co-located spatiotemporal observations. J. Environ. Qual. 35(1):21−36.

Venables, W.N. and B.D. Ripley. 1998. Modern Applied Statistics with S-PLUS, 2nd ed. Springer, New York.

Wilson, M.A., R. Burt, S.J. Indorante et al. 2008. Geochemistry in the modern soil survey program. Environ. Monit. Assess. 139(1−3):151−171.

Yang, S.Y. and W.L. Chang. 2005. Use of finite mixture distribution theory to determine the criteria of cadmium concentrations in Taiwan farmland soils. Soil Sci. 170(1):55−62.

Table 1.   Non-parametric Test Results Showing Differences in Trace Metal/Fe Ratios[*] Across Predominant USDA Soil Orders for Selected Trace Metals.

Statistic	Cd	Cr	Cu	Pb	V	Zn
Chi-Square	28.4	37.6	22.8	79.2	22.2	39.6
Degrees of Freedom	8	8	8	8	8	8
p-Value	0.0004	<0.0001	0.0036	<0.0001	0.005	<0.0001

[*]See Eq. 3. in Anderson and Kravitz, 2010

Table 2. **Summarized Total Chemical Concentrations (mg kg⁻¹) Among Reference[a] and Site-related[b] Sample Soil Populations.**

Contaminant	Reference Soils				Site-Related Soils			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Antimony	1.02	1.33	0.04	3.57	6.65	4.40	4.30	50.3
Arsenic	6.59	7.75	1.10	48.5	4.97	2.62	0.57	13.8
Barium	60.2	54.0	11.2	288	136	197	28.9	2290
Beryllium	0.39	0.21	0.06	0.87	0.78	0.39	0.37	2.66
Cadmium	0.17	0.28	0.01	2.01	0.78	1.68	0.50	18.8
Chromium	18.4	20.5	1.99	134	15.4	7.62	3.80	51.2
Cobalt	7.10	11.1	0.39	70.9	10.3	20.4	1.14	214
Copper	6.32	4.30	0.25	23.5	60.0	83.9	4.86	389
Lead	19.4	14.5	2.89	82.8	410	824	8.44	4640
Mercury	0.04	0.05	0.01	0.32	0.08	0.11	0.03	1.18
Nickel	5.08	4.14	0.82	21.8	7.97	6.91	1.07	59.7
Selenium	0.24	0.14	0.13	1.28	1.03	0.62	0.53	3.79
Silver	0.18	0.34	0.01	1.87	1.18	0.10	1.05	1.69
Thallium	0.75	1.32	0.02	3.31	1.23	0.47	0.73	5.68
Vanadium	29.6	26.4	4.66	158	23.3	11.6	5.66	53.1
Zinc	20.4	26.7	2.46	209	182	1060	9.74	9540

[a]Nonaffected by Site-related activity; $n = 67$.
[b]Possibly affected by Site-related activity; $n = 139$.
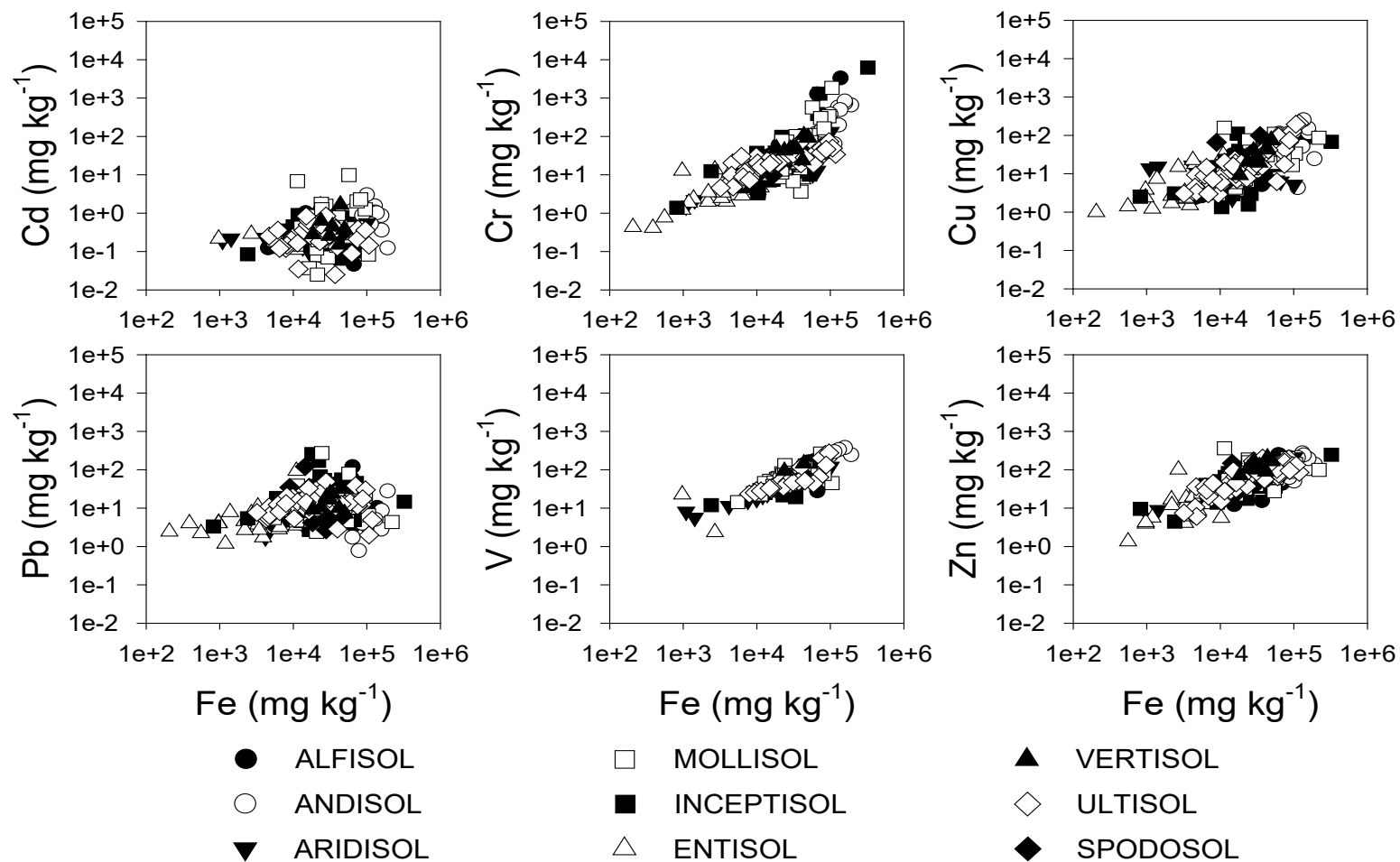
Figure 1. Log-linear Relationships Between Selected Trace Metal and Fe Concentrations Among Predominant USDA Soil Orders.
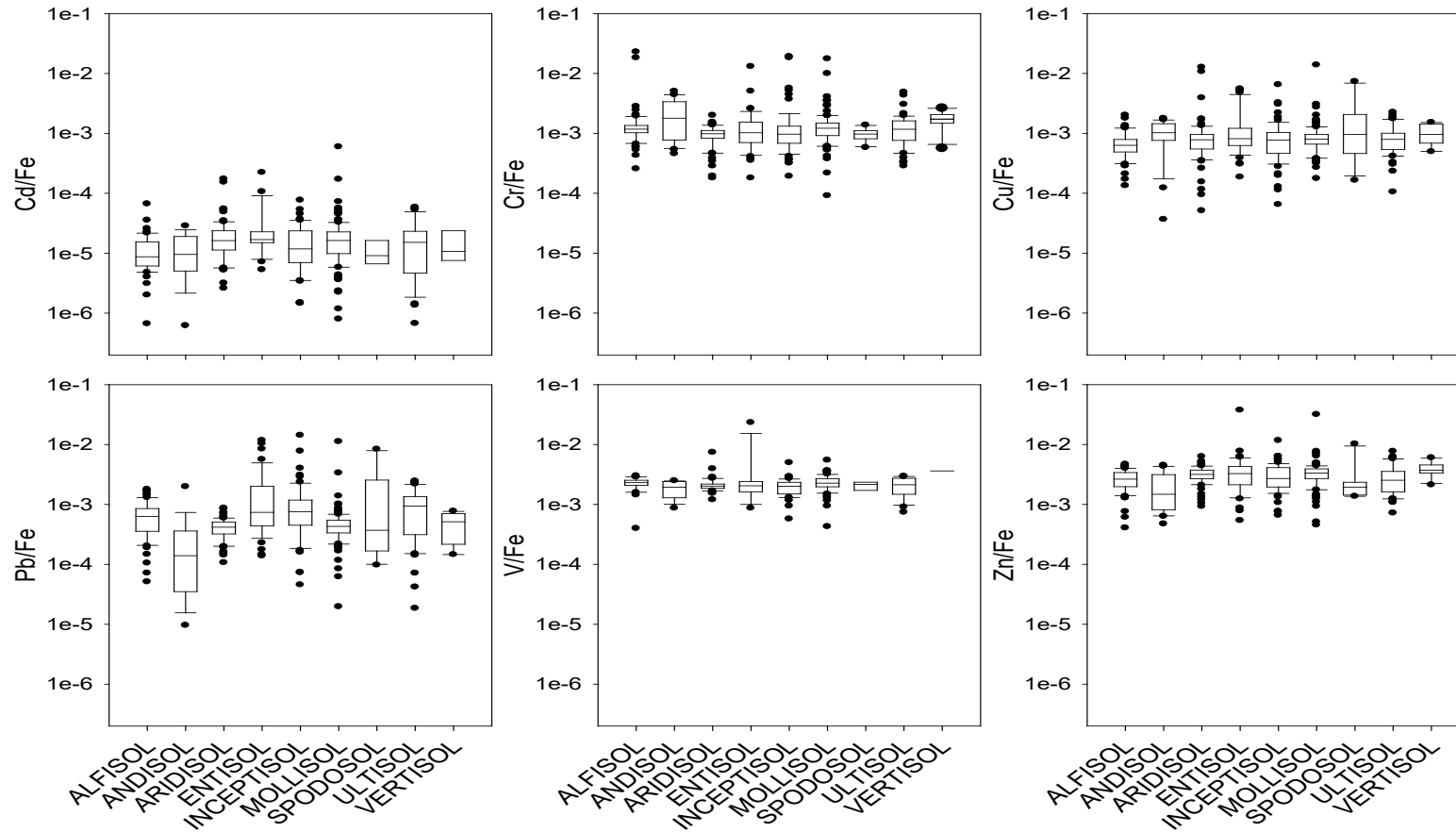
Figure 2. Box Plots Depicting Log Distributions of Geochemical Ratios Among Predominant USDA Soil Orders. Boxes represent the interquartile range (i.e., 25th–75th percentiles) and the center horizontal line represents the 50th percentile or median value. Lower and upper horizontal lines represent threshold values (i.e., 5th and 95th percentiles, respectively) beyond which constitute extreme observations represented by the shaded circles.
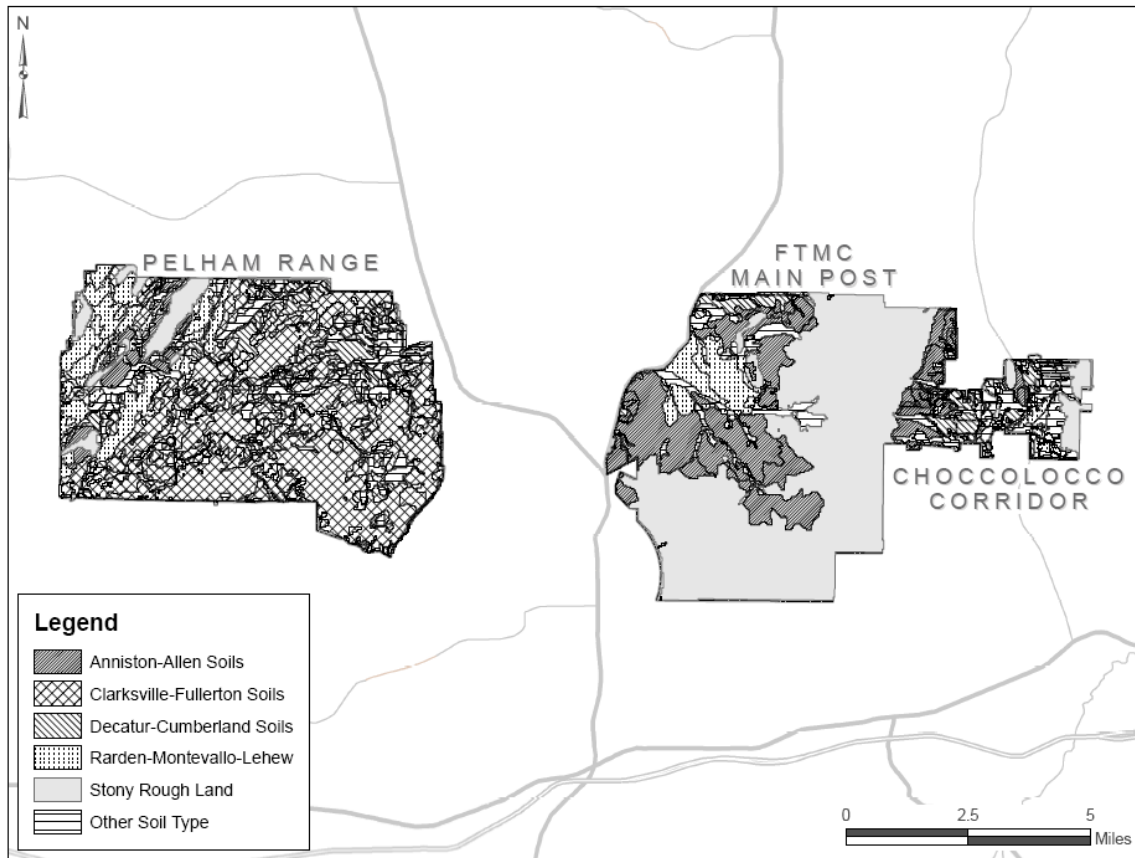
Figure 3. **Site Soils Map.** Former Choccolocco Corridor Ranges, Fort McClellan, AL, USA.
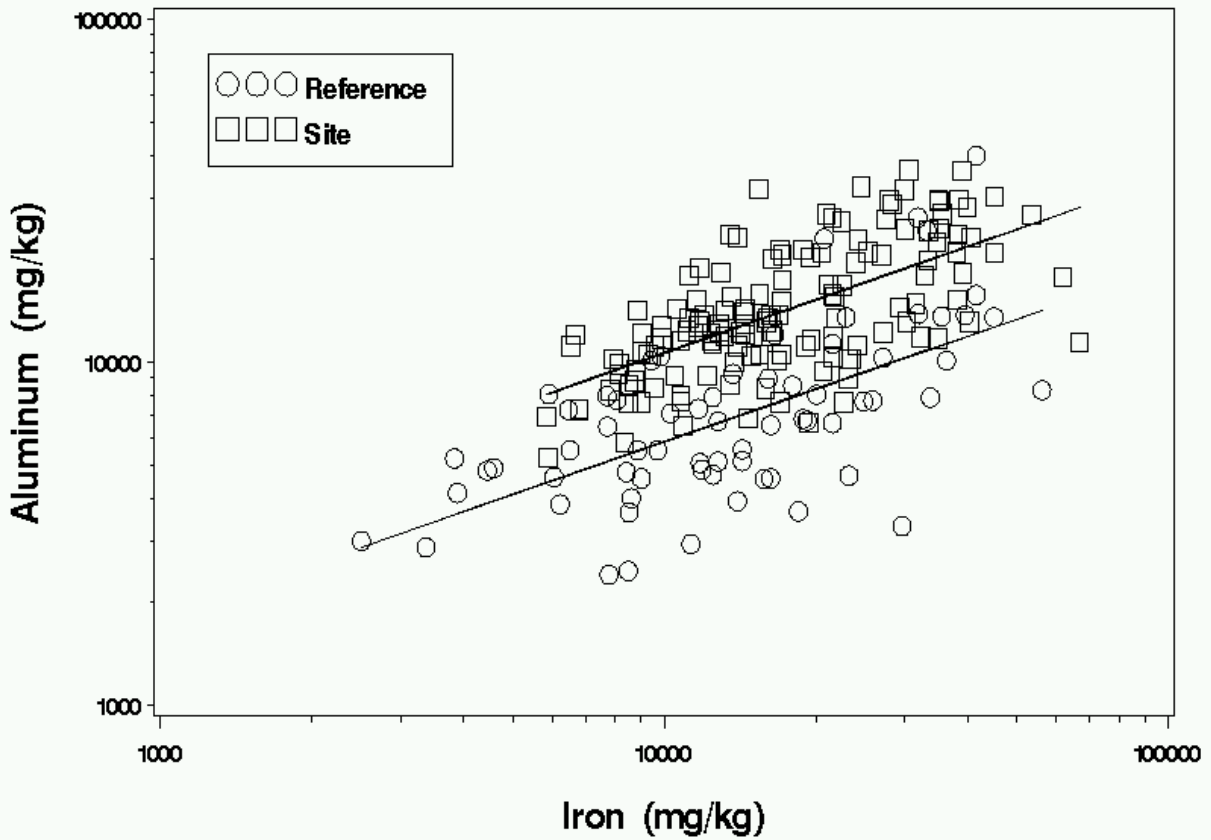
Figure 4. Geochemical Associations Between Total Soil Aluminum and Iron Concentrations Among Reference and Site-related Sample Soil Populations Illustrating Incompatible Matrices.
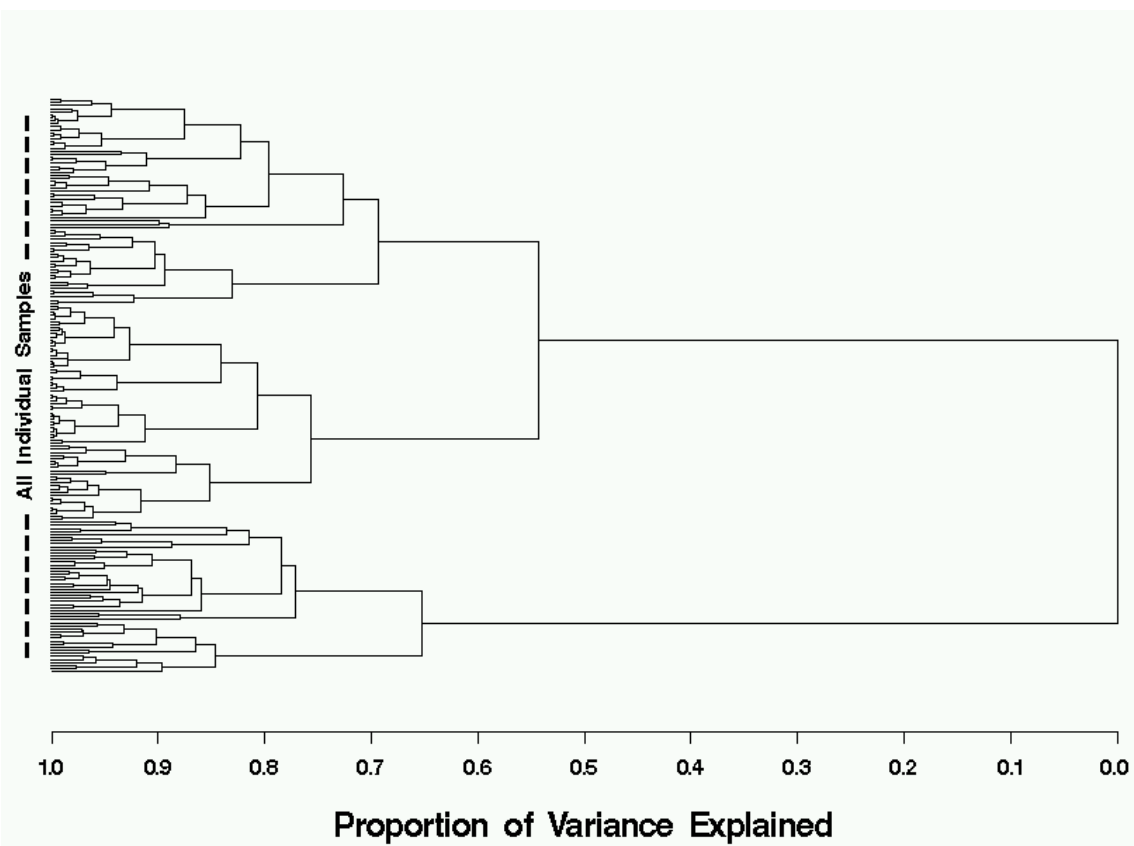
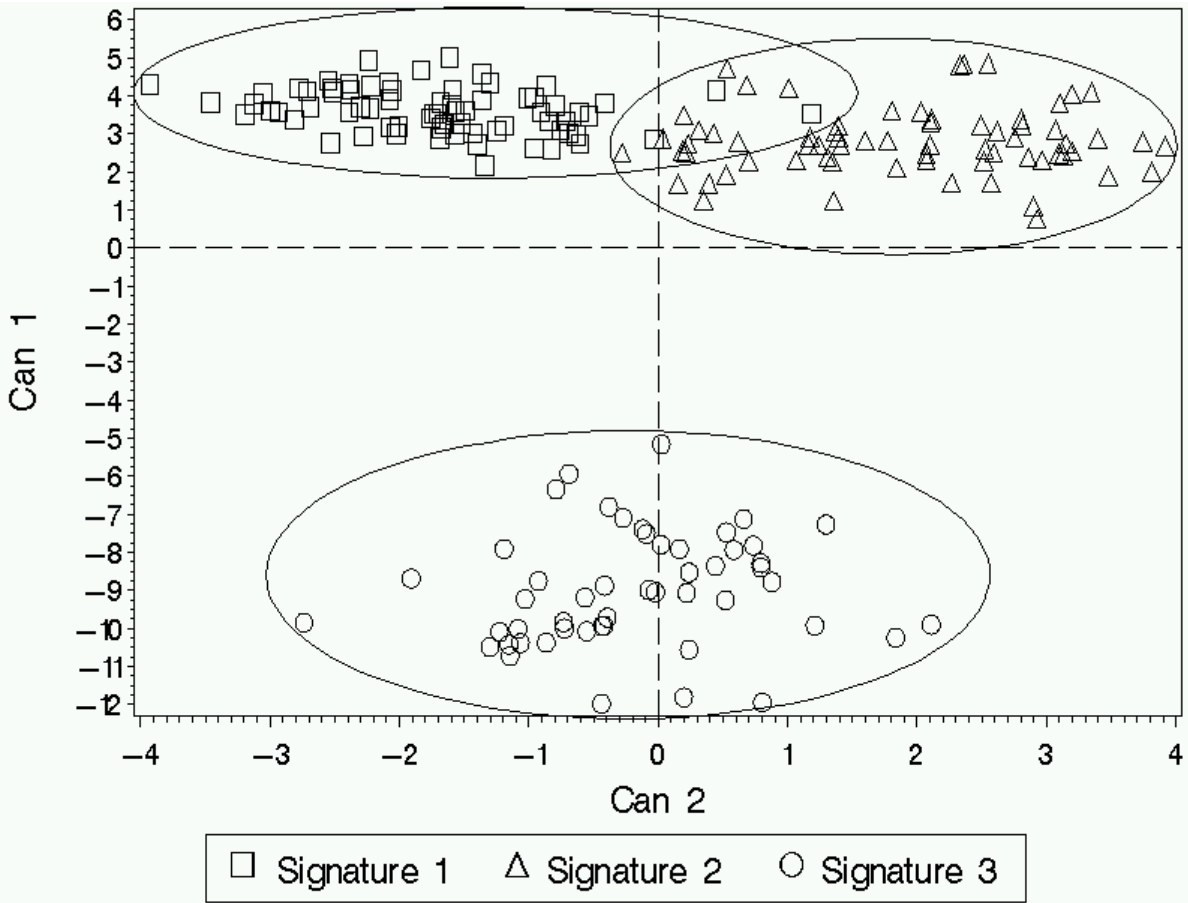**Figure 5.** **Dendrogram Depicting the Hierarchical Clustering of Chemical Concentrations.**

Figure 6. **Results from Discriminant Analysis of the Three Significant Chemical Signatures Illustrating the Multivariate Separation of Chemical Concentrations.** Signature 1 contains 51% of the Site-related soil samples, Signature 2 contains the other 49% of the site-related soil samples, and Signature 3 contains 100% of the reference soil samples.
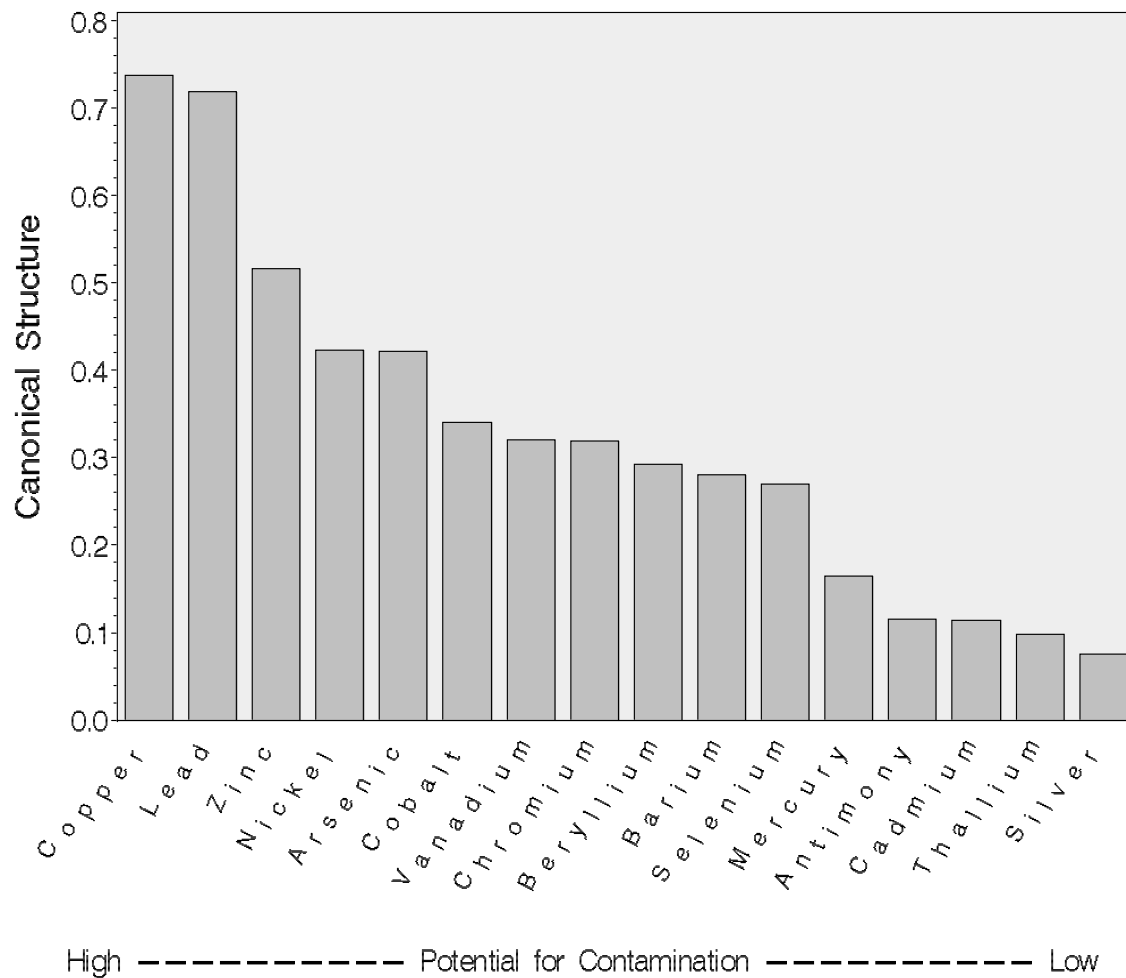
Figure 7. **Results from Discriminant Analysis Illustrating, in Order of Relative Magnitude, the Contaminants of Potential Concern for Site-related Soils (i.e., the second canonical variable).**