SEPARATING BACKGROUND FROM METALS CONTAMINATION: GEOCHEMICAL EVALUATIONS

R.A. Hunter Anderson, Ph.D.

ORISE Post Doctoral Fellow
U.S. Environmental Protection Agency
Office of Research and Development
National Center for Environmental Assessment

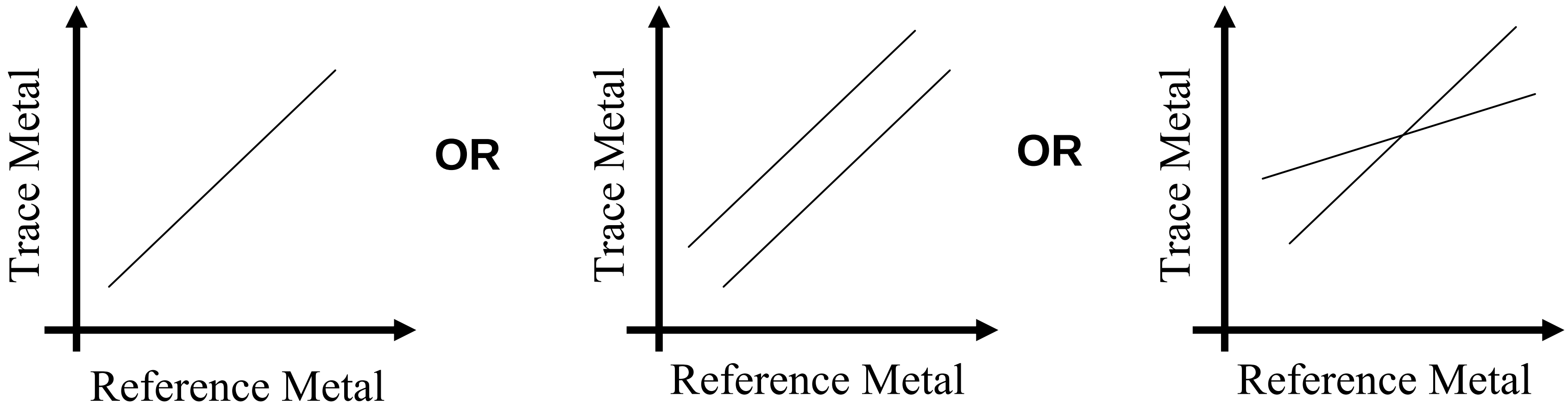
EPA/TSERWG Joint Conference: January, 2009, Vicksburg, MS

REQUEST #18 TO THE ECOLOGICAL RISK ASSESSMENT SUPPORT CENTER (ERASC)

1. Evaluation of Geochemical Associations as a Screening Tool for Identifying Anthropogenic Trace Metal Contamination
   - *Final Draft Complete*
   - *In Review: J. Environ. Mont. Assess.*
2. Use of Statistics in Geochemical Evaluations to Determine Sites Impacted by Anthropogenic Trace Metal Contamination
   - *First Draft Complete*
3. *Application of Discriminant Analysis with Clustered Data to Determine Anthropogenic Metals Contamination*
   - *In Prep.*

EVALUATION OF GEOCHEMICAL ASSOCIATIONS AS A SCREENING TOOL FOR IDENTIFYING ANTHROPOGENIC TRACE METAL CONTAMINATION

Background Geochemical Associations among predominant soil orders were tested for proportionality (i.e., slopes) and consistency (i.e., intercepts = ratios).



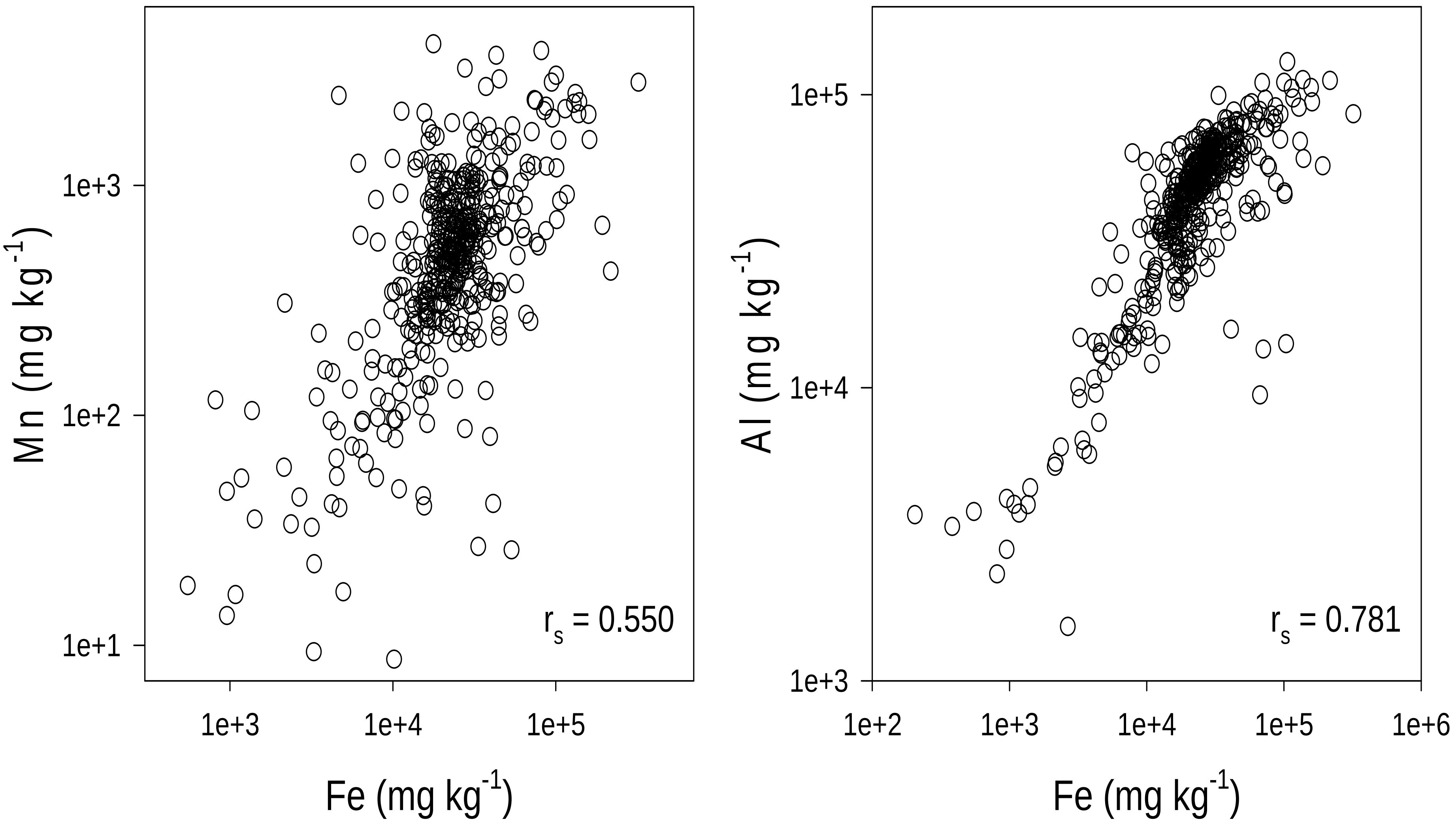
(Eq. 1) $\log[y] = \alpha + \beta_0 \log[x] + (\beta_1 Z_1 + ... + \beta_j Z_{n-1}) + \beta_{j+1} \log[x] * (Z_1 * .... * Z_{n-1}) + \varepsilon$

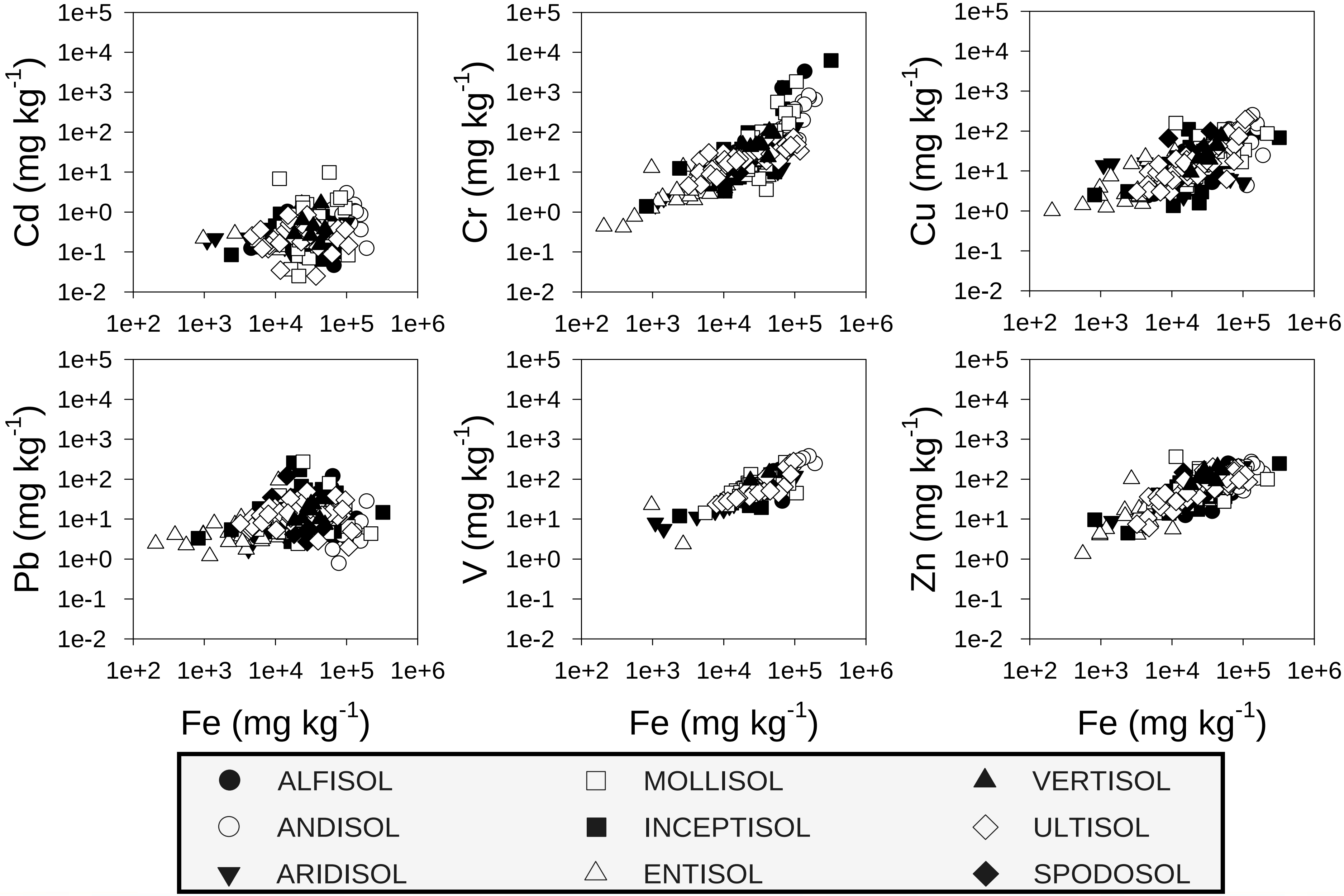
A regional geochemistry data set was obtained from the USDA NRCS Cooperative Soil Survey Program containing total background metal concentrations from 636 soil pedons from around the conterminous U.S.A.

- **http://soils.usda.gov/survey/geochemistry/index.html**
- **Selected trace metals evaluated: Cd, Cr, Cu, Pb, V, and Zn**

Order:	Alfisol	Andisol	Aridisol	Entisol	Inceptisol	Mollisol	Spodosol	Ultisol	Vertisol
N[a]:	64	21	74	46	72	115	11	54	10

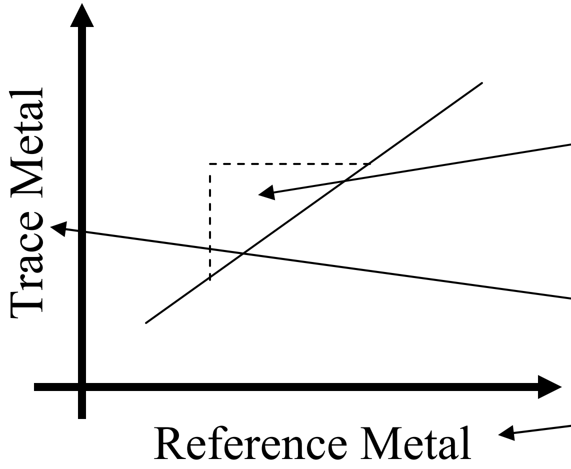
[a] Sample size of individual soil series (i.e., finest level of USDA classification) for each soil order.


Mn (mg kg-1)
1e+3
1e+2
1e+1
1e+3
1e+4
1e+5
Fe (mg kg-1)
rs = 0.550
Al (mg kg-1)
1e+5
1e+4
1e+3
1e+2
1e+3
1e+4
1e+5
1e+6
Fe (mg kg-1)
rs = 0.781


Cd (mg kg-1)
Cr (mg kg-1)
Cu (mg kg-1)
Pb (mg kg-1)
V (mg kg-1)
Zn (mg kg-1)
1e+5
1e+4
1e+3
1e+2
1e+1
1e+0
1e-1
1e-2
1e+2
1e+3
1e+4
1e+5
1e+6
Fe (mg kg-1)
Fe (mg kg-1)
Fe (mg kg-1)
ALFISOL
ANDISOL
ARIDISOL
MOLLISOL
INCEPTISOL
ENTISOL
VERTISOL
ULTISOL
SPODOSOL

Assuming slopes from log-log plots are proportional and roughly 1 (?), differences across soil orders can be tested using log trace metal/major metal ratios with nonparametric tests (i.e., not influenced by sample size).

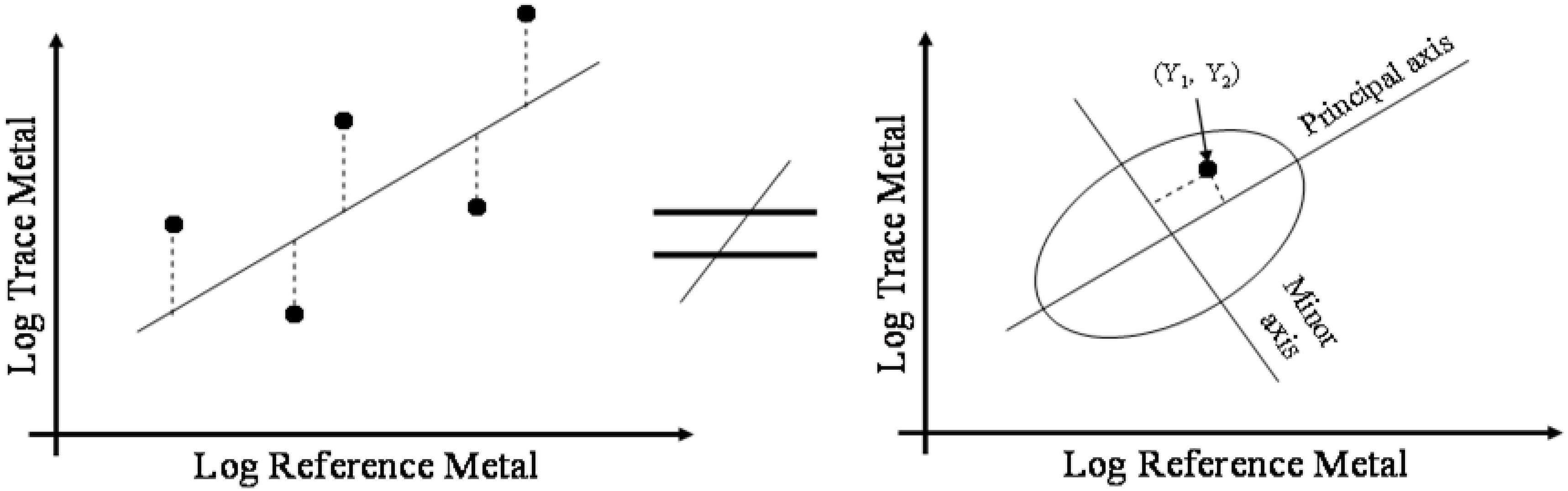
$$\log[y] = Y_{int} + \log[x] + \varepsilon \implies \log\left(\frac{[y]}{[x]}\right) = Y_{int}$$



Are slopes 1?

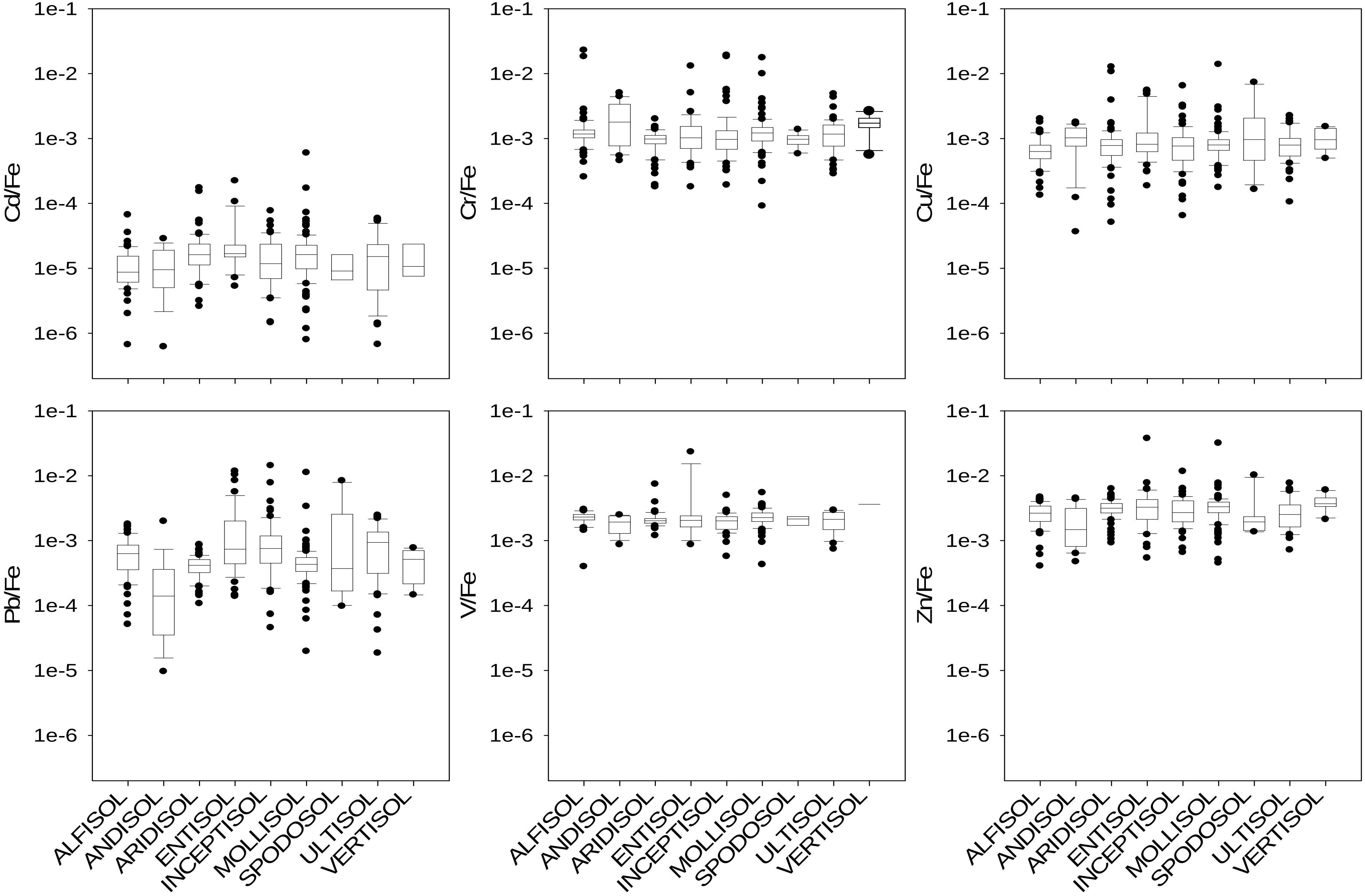
Both are random variables with the same sampling error

Least Squares / *Principal Components*



Principal component analysis showing common slope and 99% C.I.s of the principal axis for log trace metal/Fe relationships for selected trace metals.							
Statistic		**Cd**	**Cr**	**Cu**	**Pb**	**V**	**Zn**
Lower Bound		***0.62***	***1.16***	***0.90***	***0.35***	***0.81***	***0.81***
Slope		***1.22***	***1.27***	***1.02***	***0.59***	***0.90***	***0.90***
Upper Bound		***2.70***	***1.40***	***1.16***	***0.89***	***1.00***	***1.00***

Sokal and Rohlf. 1981. Biometry


Cd/Fe
Cr/Fe
Cu/Fe
Pb/Fe
V/Fe
Zn/Fe
1e-1
1e-2
1e-3
1e-4
1e-5
1e-6
ALFISOL
ANDISOL
ARIDISOL
ENTISOL
INCEPTISOL
MOLLISOL
SPODOSOL
ULTISOL
VERTISOL

Nonparametric test results showing differences in log trace metal/Fe ratios across predominant U.S. soil orders for selected trace metals.

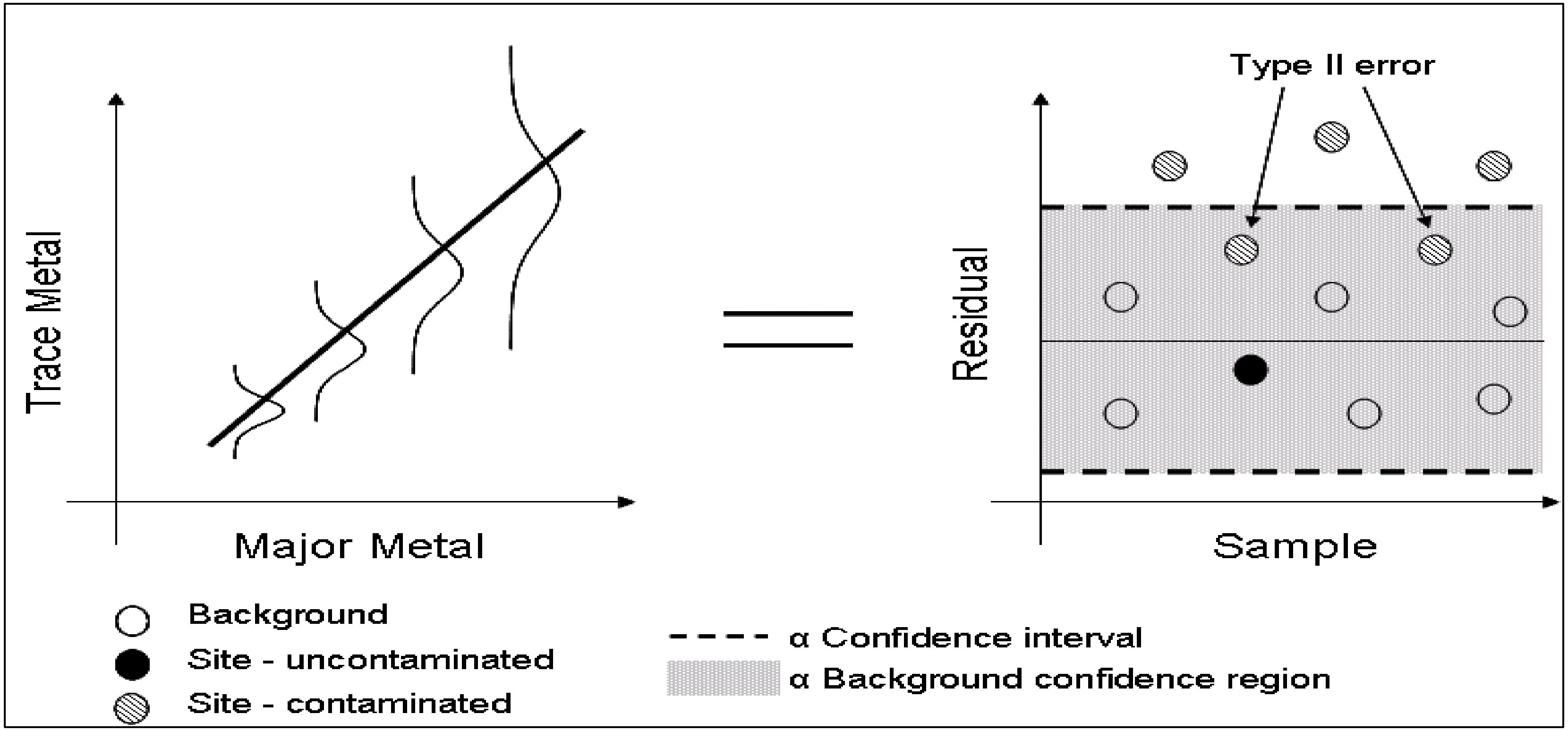
Statistic		Cd	Cr	Cu	Pb	V	Zn
Chi-Square		28.4	37.6	22.8	79.2	22.2	39.6
D.F.		8	8	8	8	8	8
P – Value		0.0004	< 0.0001	0.0036	< 0.0001	0.0050	< 0.0001

Highly significant results suggest non-constant ratios among predominant U.S. soil orders.

Although relationships may or may not be proportional, ubiquitous application of generic background datasets may result in type II or type I statistical error.

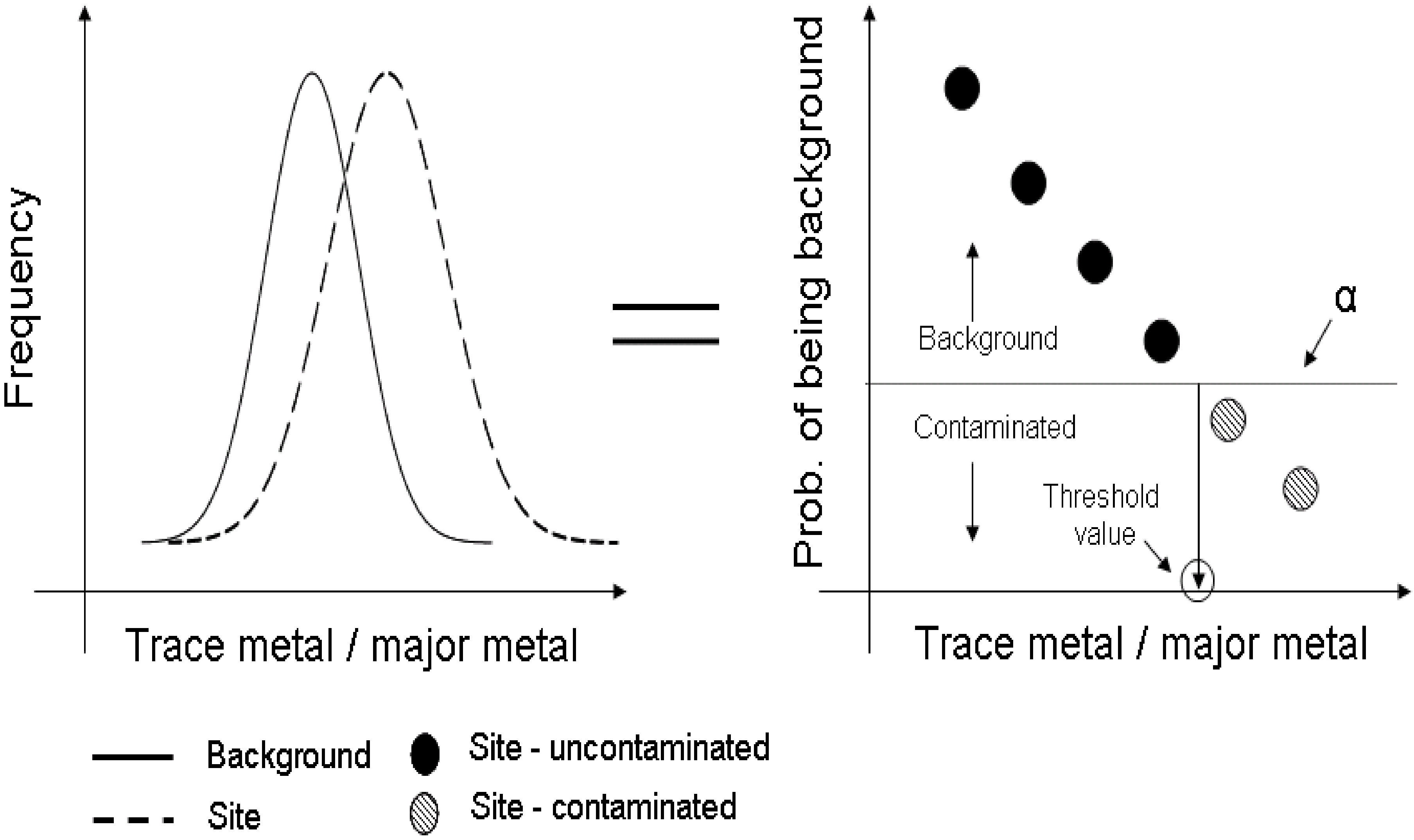
Results from agglomerative clustering technique were used to identify patterns of association among soil orders – may aid environmental assessors in screening candidate background metal datasets for their applicability to site-specific soil composition.

USE OF STATISTICS IN GEOCHEMICAL EVALUATIONS TO DETERMINE SITES IMPACTED BY ANTHROPOGENIC TRACE METAL CONTAMINATION



Theoretical representation of the issue with heteroscedastic populations of background and site geochemical metal ratios for conventional regression-based outlier detection.

Preferred Approach: Cumulative Density Function



APPLICATION OF DISCRIMINANT ANALYSIS WITH CLUSTERED DATA TO IDENTIFY ANTHROPOGENIC METALS CONTAMINATION

Objective: Identify in relative order the metal/metals that optimally distinguishes between clusters (i.e., signatures)

Advantages over Geochemical Evaluations:

1: Nonparametric inference possible (no distributional assumptions)

2: Multivariate rather than bivariate evaluation "chemical signature"

3: Incorporate priors for certainty/uncertainty of "true" background

Step 1: Identify observations with similar signatures (i.e., clusters)

– Model based clustering can determine significant distinct clusters

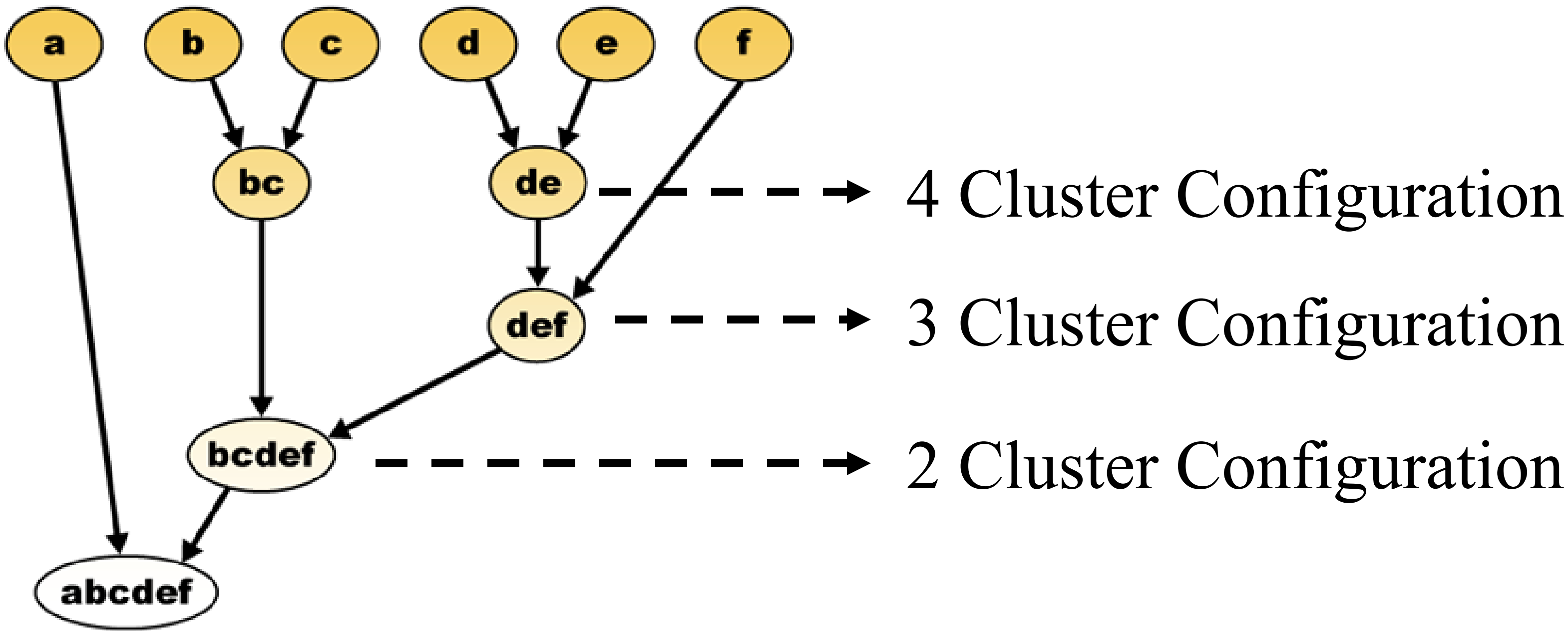
Step 2: Develop discriminant criterion optimally separating clusters

– Creates canonical variables

Step 3: Determine in relative order contaminated metals

– Based on canonical structure (i.e., linear correlation)

1. Cluster Analysis: Defines Multivariate Signatures



- Multidimensional distances determined by metal concentrations

2. Discriminant Analysis: Defines Contaminants

$$\Delta \text{Cluster} = b_1 * x_1 + b_2 * x_2 + \ldots + b_m * x_m$$

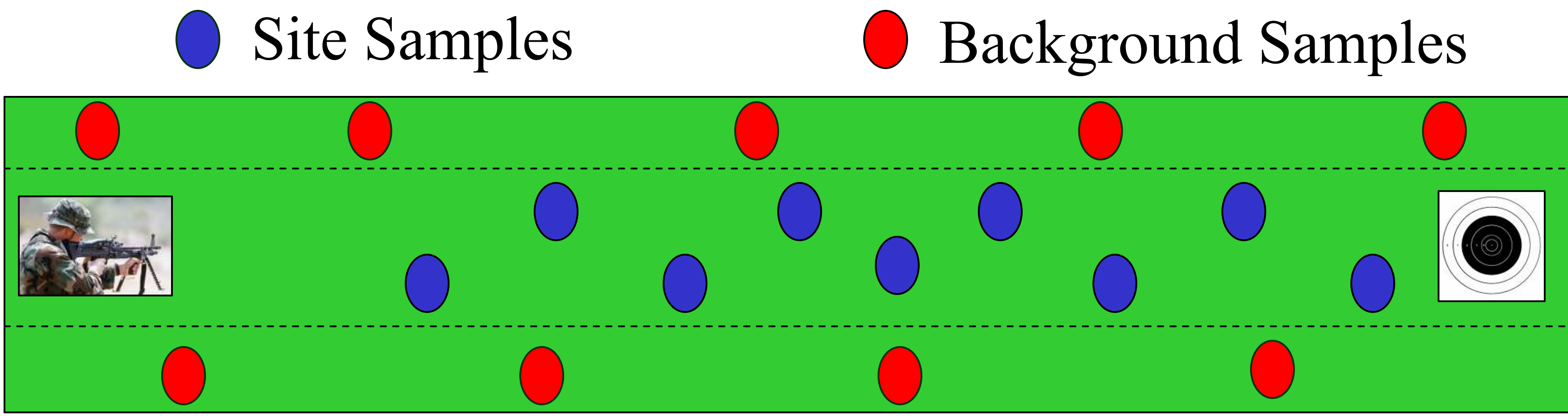
b_i = Canonical Coefficient; x_i = Analyte Suite (i.e., COIs)

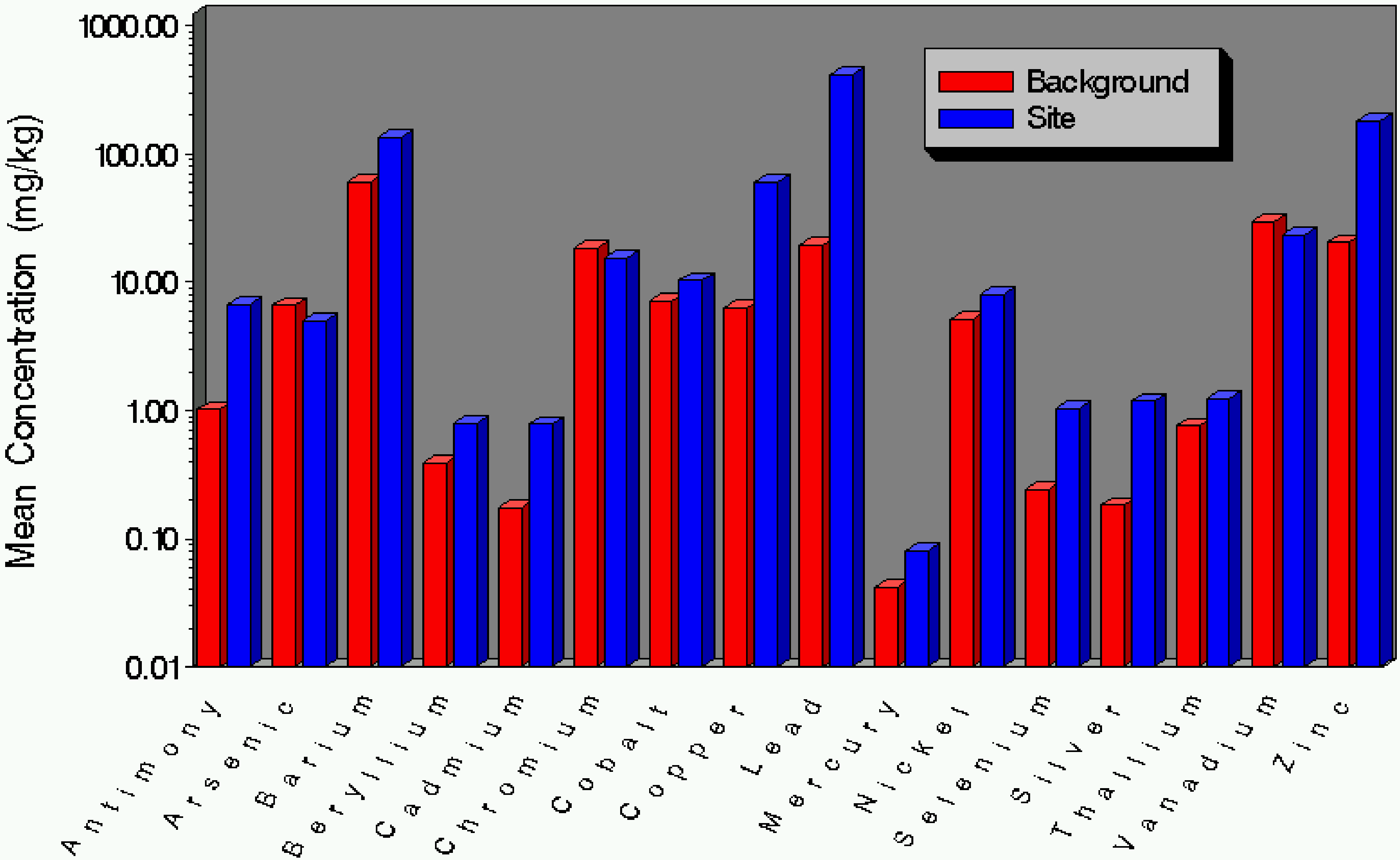
Field Example:

- Data were obtained from an undisclosed firing range
- 16 COIs were evaluated

1. Antimony	5. Cadmium	9. Lead	13. Silver
2. Arsenic	6. Cobalt	10. Mercury	14. Thallium
3. Barium	7. Chromium	11. Nickel	15. Vanadium
4. Beryllium	8. Copper	12. Selenium	16. Zinc

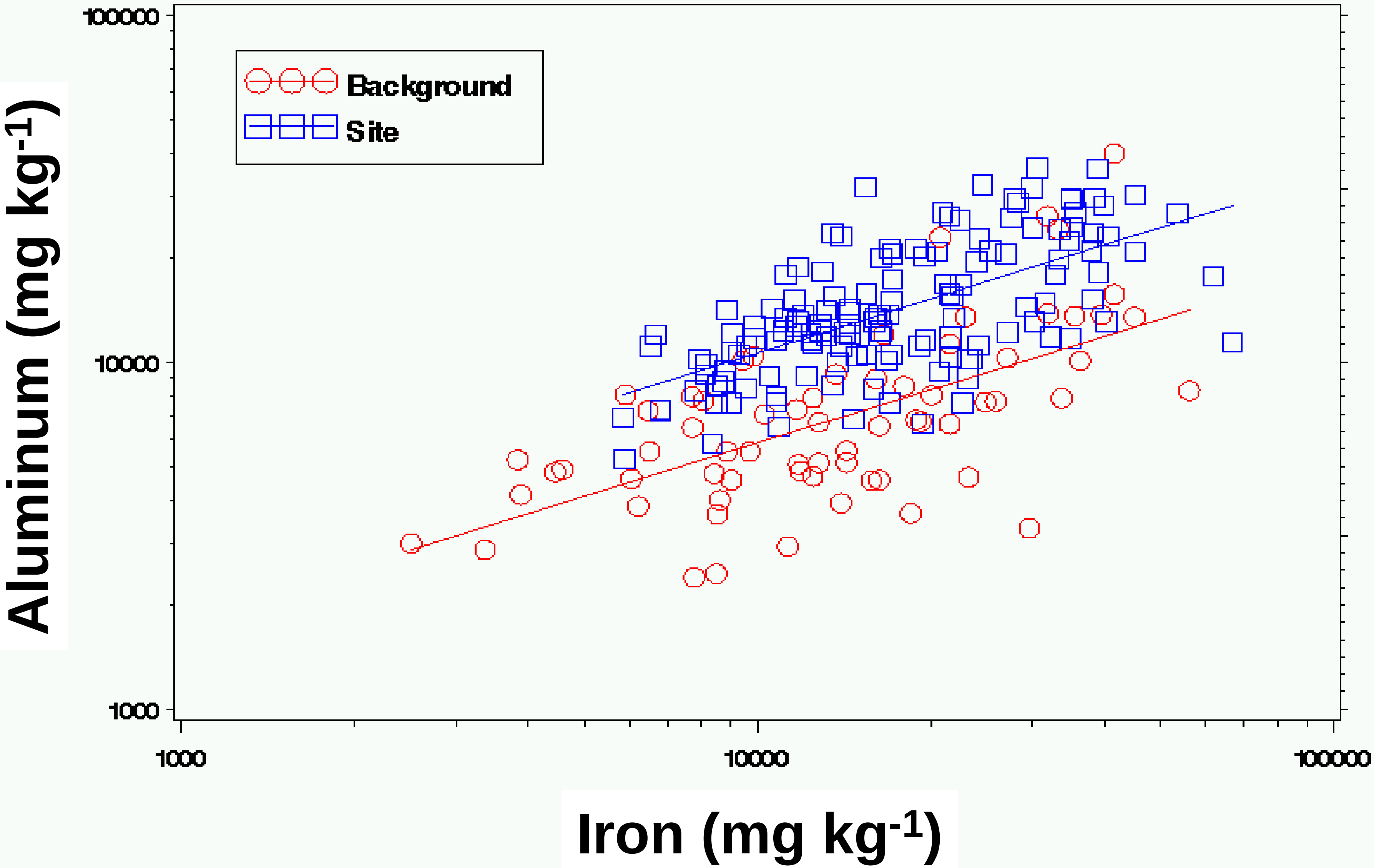
- Observations were classified *a priori* as “background” or “site”



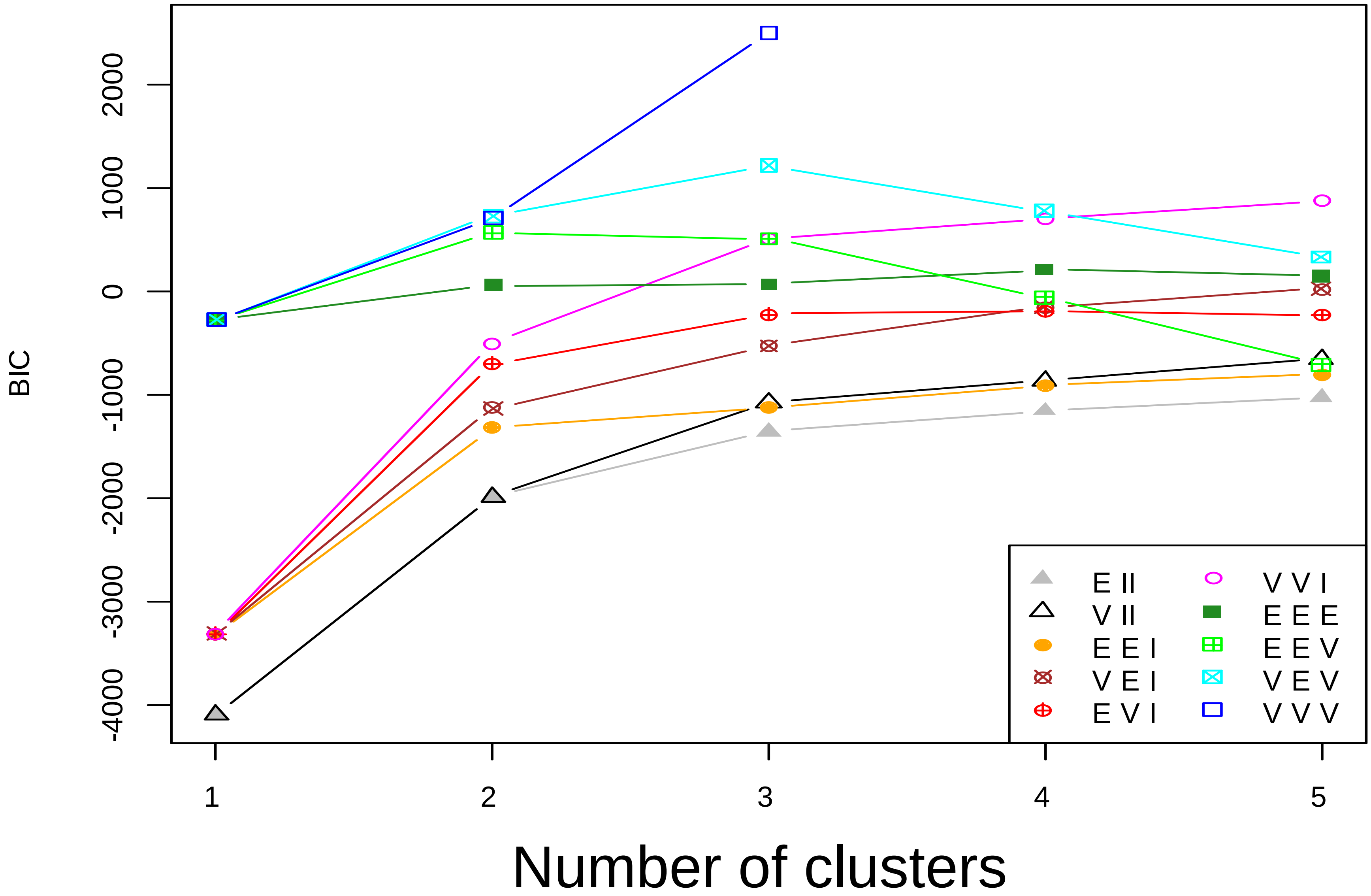

Mean Concentration (mg/kg)
1000.00
100.00
10.00
1.00
0.10
0.01
Background
Site
Antimony
Arsenic
Barium
Beryllium
Cadmium
Chromium
Cobalt
Copper
Lead
Mercury
Nickel
Selenium
Silver
Thallium
Vanadium
Zinc

Caution: Sometimes Things Deserve a Closer Look

Suggests dissimilar site/reference soils (i.e., apples vs. oranges)



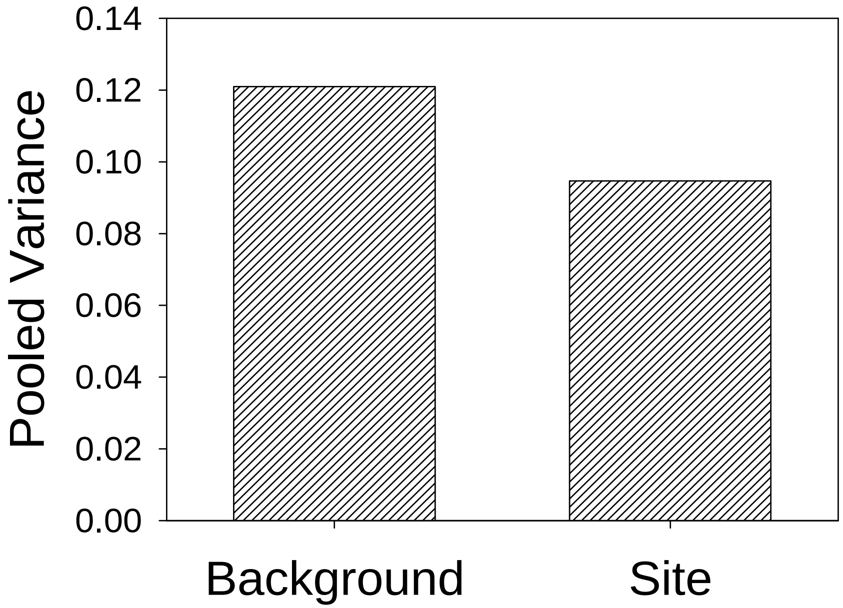
Significant cluster determination: Model-based clustering



Two cluster cross-classification

Cluster:	1	2
Background:	100%	0%
Site:	0%	100%

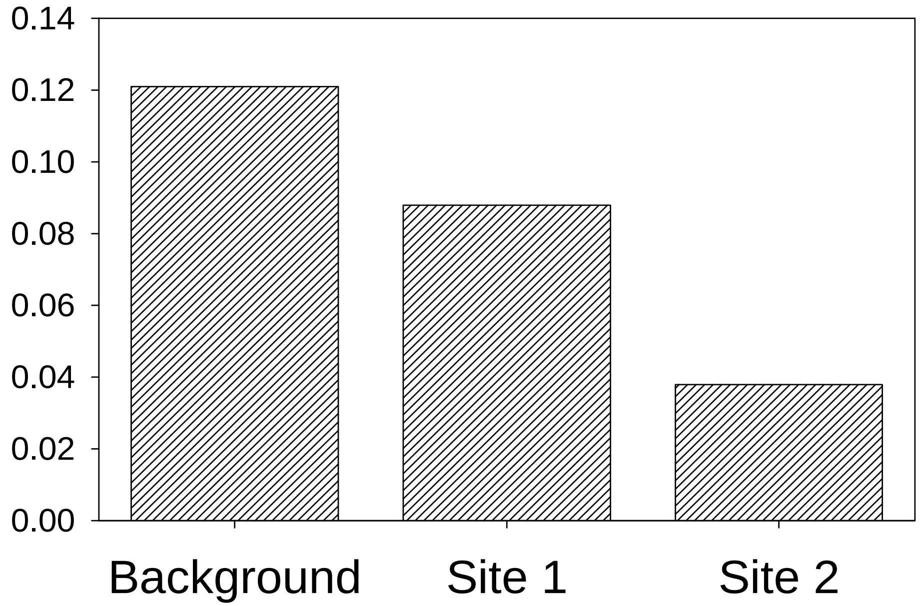
Chi-square = 189; $p < 0.0001$

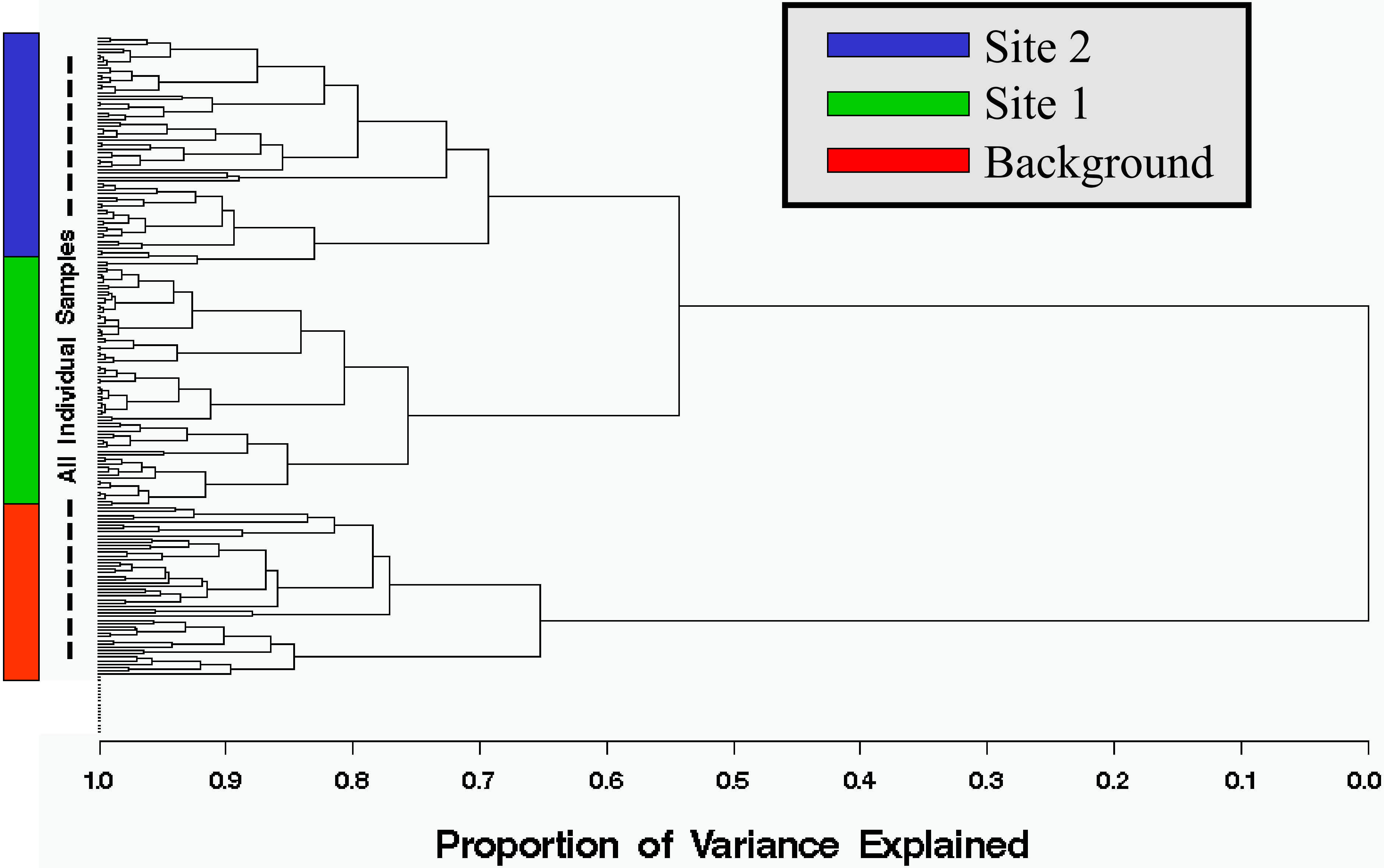


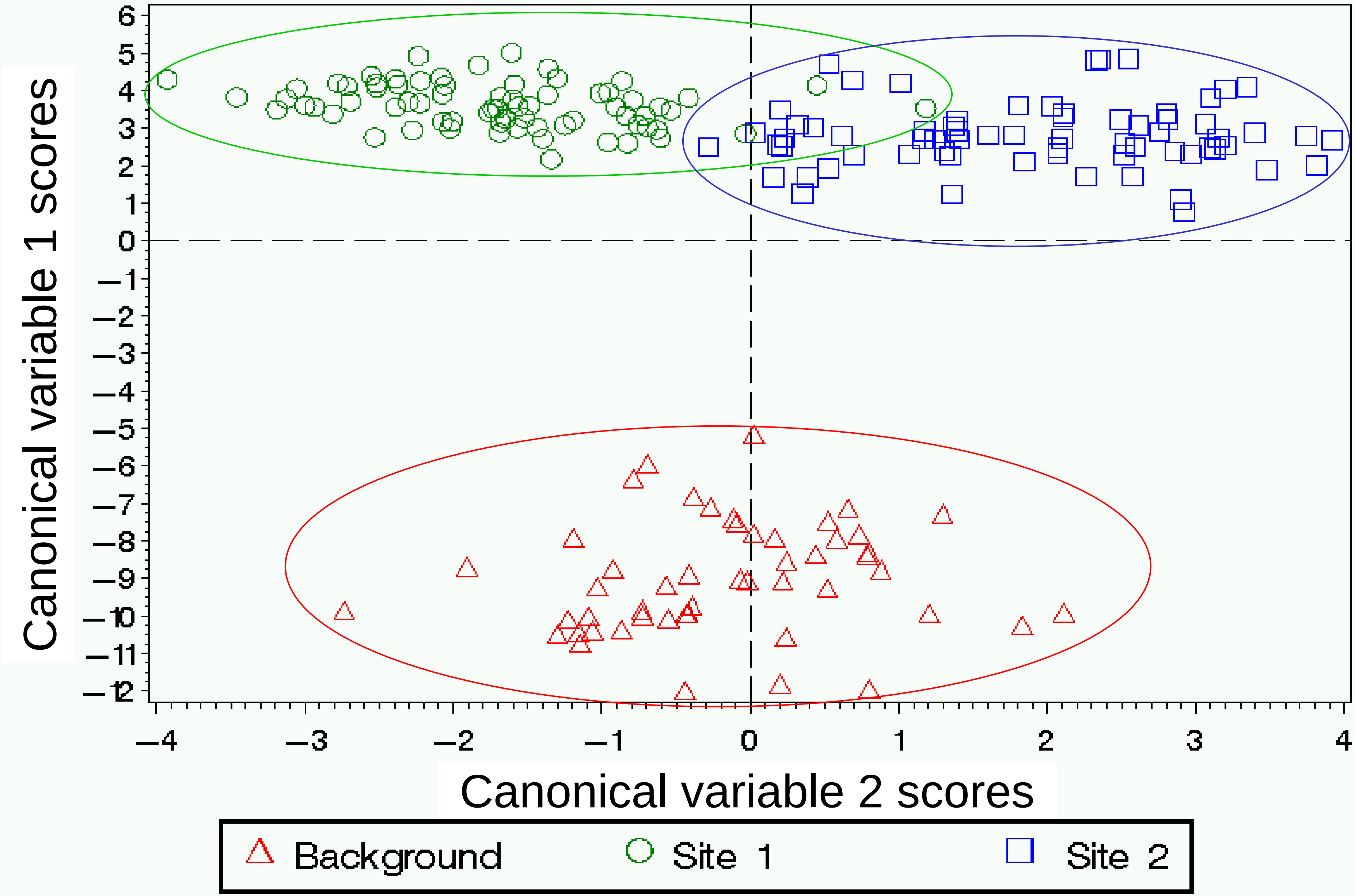
Three cluster cross-classification

Cluster:	1	2	3
Background:	100%	0%	0%
Site:	0%	51%	49%

Chi-square = 189; $p < 0.0001$

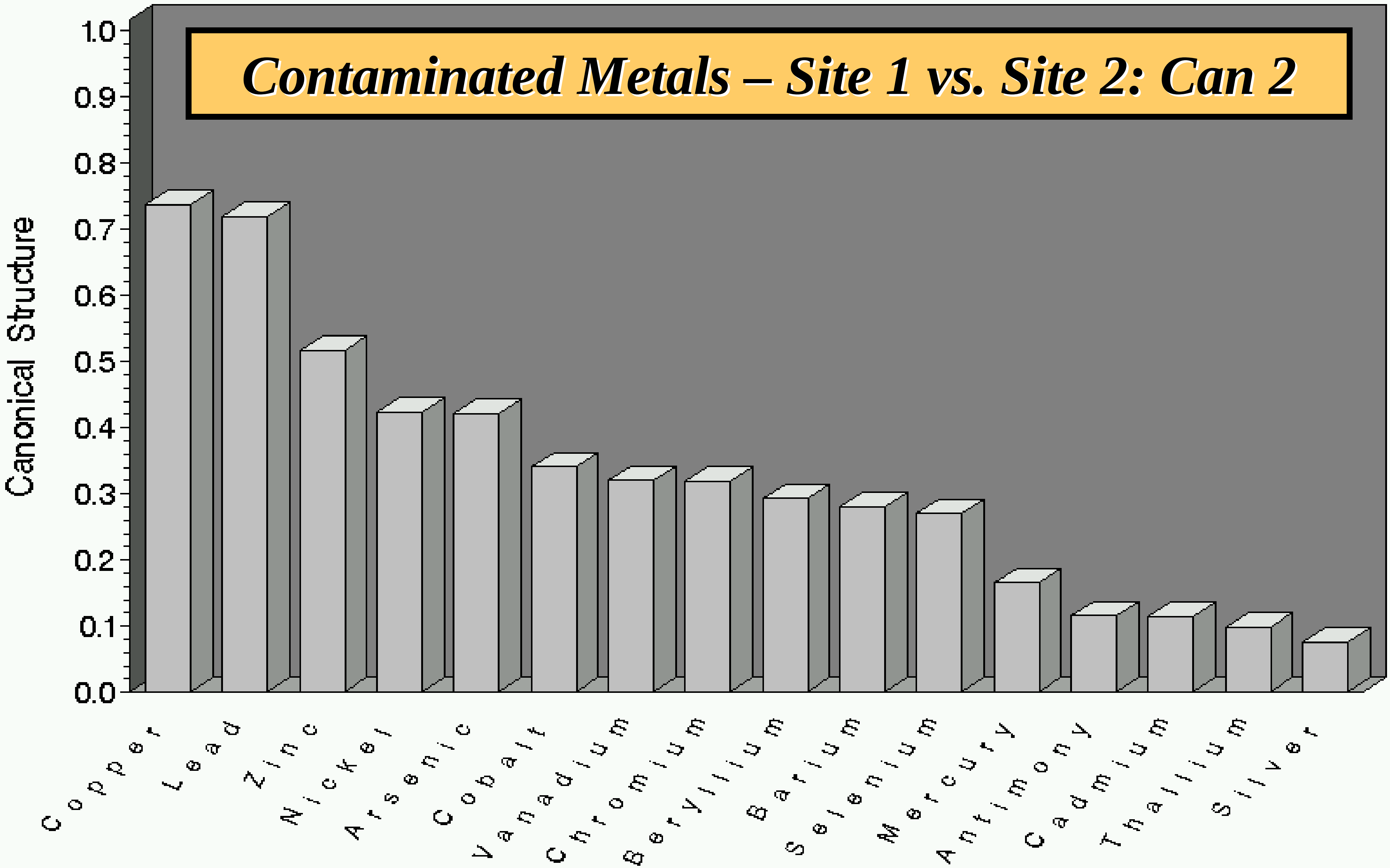



Site 2
Site 1
Background
All Individual Samples
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Proportion of Variance Explained

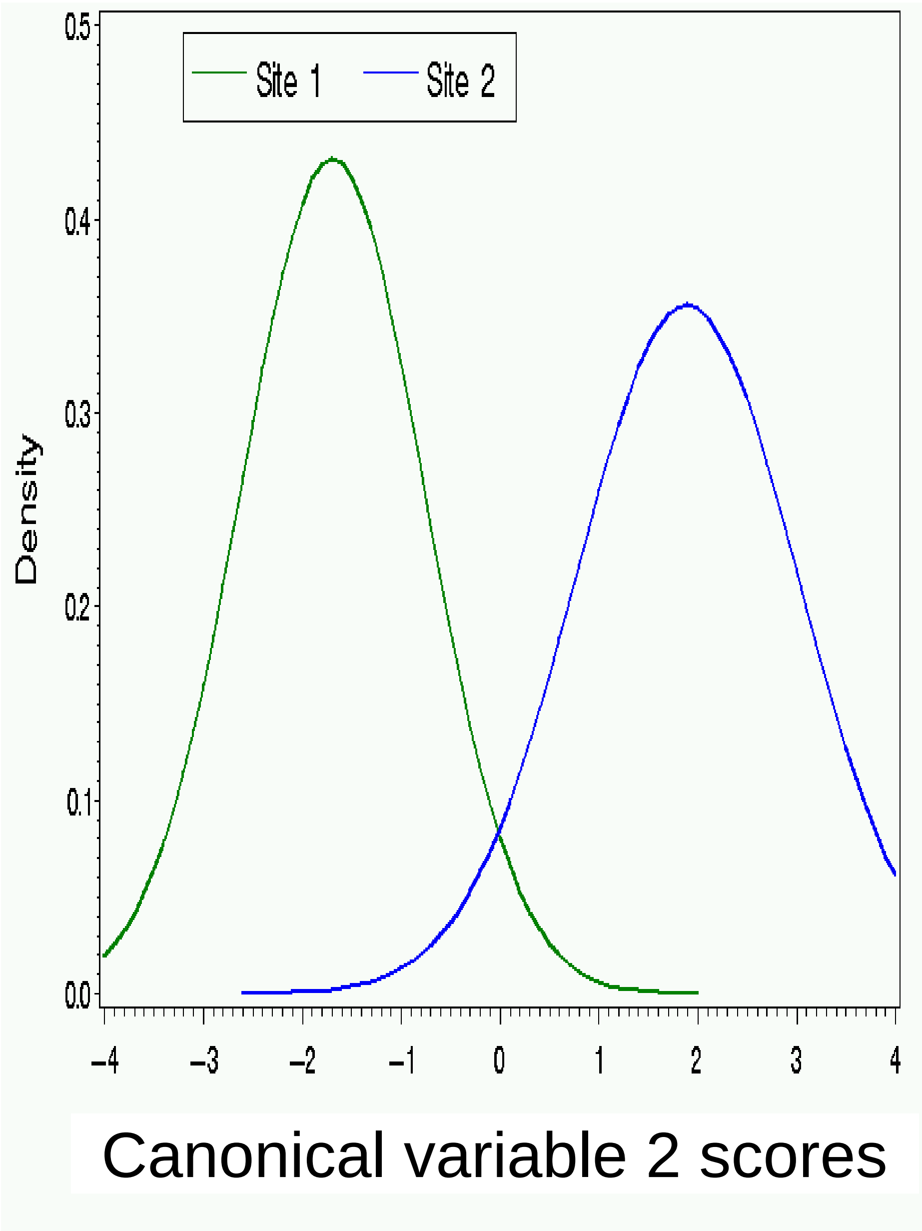


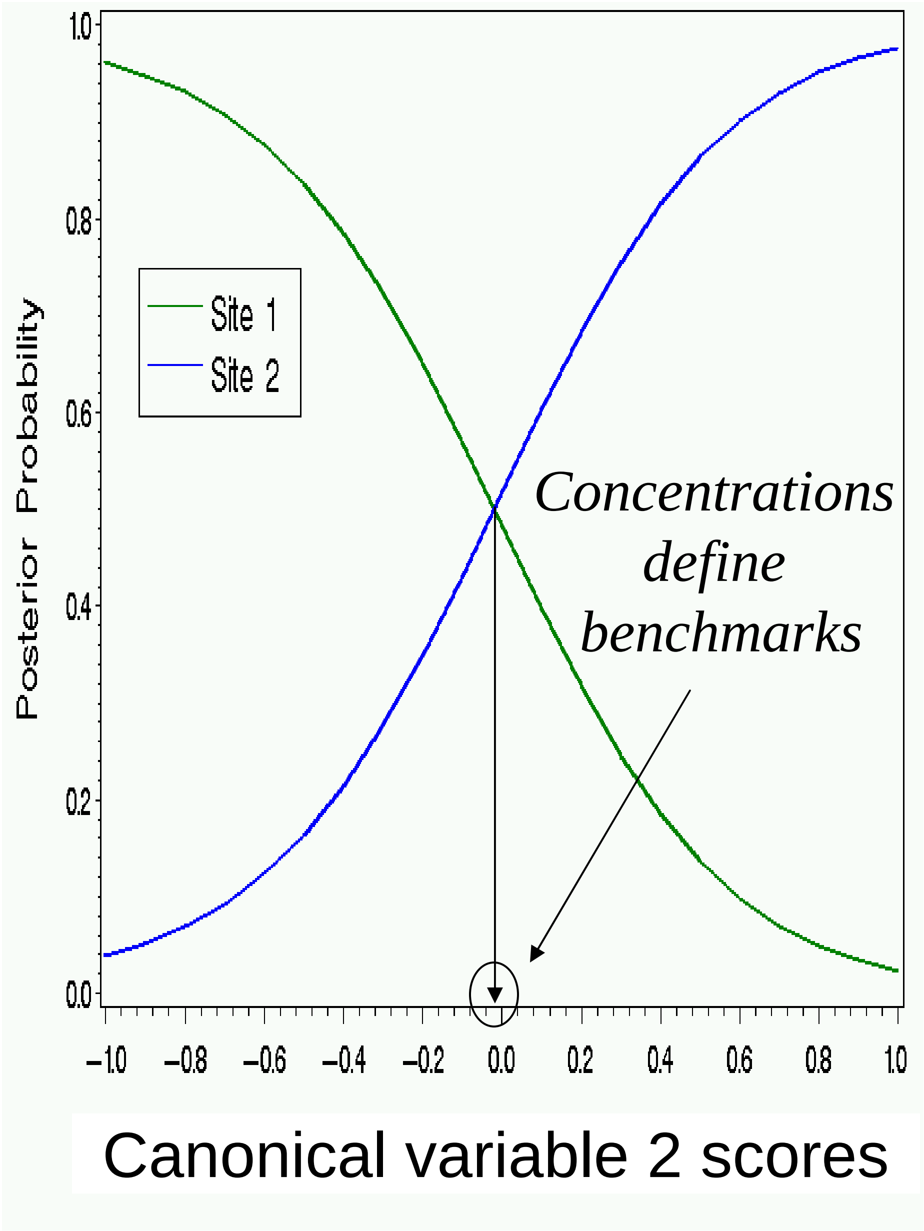
6
5
4
3
2
1
0
−1
−2
−3
−4
−5
−6
−7
−8
−9
−10
−11
−12
Canonical variable 1 scores
−4
−3
−2
−1
0
1
2
3
4
Canonical variable 2 scores
Background
Site 1
Site 2

Contaminated Metals – Site 1 vs. Site 2: Can 2



High ← Relative Magnitude of Contamination → Low


Site 1
Site 2
Density
0.0
0.1
0.2
0.3
0.4
0.5
-4
-3
-2
-1
0
1
2
3
4
Canonical variable 2 scores


Site 1
Site 2
Posterior Probability
0.0
0.2
0.4
0.6
0.8
1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Concentrations define benchmarks
Canonical variable 2 scores

ACKNOWLEDGEMENTS

ORD:

Dave Farrar

Region 4:

Sharon Thoms

ORISE

OAK RIDGE INSTITUTE FOR SCIENCE AND EDUCATION

http://orise.orau.gov/