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

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).

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


- Selected trace metals evaluated: Cd, Cr, Cu, Pb, V, and Zn

<table>
<thead>
<tr>
<th>Order:</th>
<th>Alfisol</th>
<th>Andisol</th>
<th>Aridisol</th>
<th>Entisol</th>
<th>Inceptisol</th>
<th>Mollisol</th>
<th>Spodosol</th>
<th>Ultisol</th>
<th>Vertisol</th>
</tr>
</thead>
<tbody>
<tr>
<td>N(^a):</td>
<td>64</td>
<td>21</td>
<td>74</td>
<td>46</td>
<td>72</td>
<td>115</td>
<td>11</td>
<td>54</td>
<td>10</td>
</tr>
</tbody>
</table>

\(^a\) Sample size of individual soil series (i.e., finest level of USDA classification) for each soil order.
The scatter plots show the relationship between Fe (mg kg\(^{-1}\)) and Mn (mg kg\(^{-1}\)) on the left, and Fe (mg kg\(^{-1}\)) and Al (mg kg\(^{-1}\)) on the right. The correlation coefficients are $r_s = 0.550$ and $r_s = 0.781$, respectively.
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*
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.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Cd</th>
<th>Cr</th>
<th>Cu</th>
<th>Pb</th>
<th>V</th>
<th>Zn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Bound</td>
<td>0.62</td>
<td>1.16</td>
<td>0.90</td>
<td>0.35</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>Slope</td>
<td>1.22</td>
<td>1.27</td>
<td>1.02</td>
<td>0.59</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>Upper Bound</td>
<td>2.70</td>
<td>1.40</td>
<td>1.16</td>
<td>0.89</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Sokal and Rohlf. 1981. Biometry
Nonparametric test results showing differences in log trace metal/Fe ratios across predominant U.S. soil orders for selected trace metals.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Cd</th>
<th>Cr</th>
<th>Cu</th>
<th>Pb</th>
<th>V</th>
<th>Zn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>28.4</td>
<td>37.6</td>
<td>22.8</td>
<td>79.2</td>
<td>22.2</td>
<td>39.6</td>
</tr>
<tr>
<td>D.F.</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>P – Value</td>
<td>0.0004</td>
<td>&lt; 0.0001</td>
<td>0.0036</td>
<td>&lt; 0.0001</td>
<td>0.0050</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

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
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

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

\( b_i = \text{Canonical Coefficient}; x_i = \text{Analyte Suite (i.e., COIs)} \)
Field Example:

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

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

- Observations were classified *a priori* as “background” or “site”
Caution: Sometimes Things Deserve a Closer Look
Iron (mg kg⁻¹)

Aluminum (mg kg⁻¹)

Suggests dissimilar site/reference soils (i.e., apples vs. oranges)
Significant cluster determination: Model-based clustering
### Two cluster cross-classification

<table>
<thead>
<tr>
<th>Cluster:</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background:</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Site:</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Chi-square = 189; p < 0.0001

### Three cluster cross-classification

<table>
<thead>
<tr>
<th>Cluster:</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background:</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Site:</td>
<td>0%</td>
<td>51%</td>
<td>49%</td>
</tr>
</tbody>
</table>

Chi-square = 189; p < 0.0001
Contaminated Metals – Site 1 vs. Site 2: Can 2

High Relative Magnitude of Contamination Low

Canonical Structure

Copper | Lead | Zinc | Nickel | Arsenic | Cobalt | Vanadium | Chromium | Beryllium | Barium | Selenium | Mercury | Antimony | Cadmium | Thallium | Silver

High Relative Magnitude of Contamination Low

RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions
Concentrations define benchmarks
ACKNOWLEDGEMENTS

ORD:
Dave Farrar

Region 4:
Sharon Thoms

http://orise.orau.gov/