PRINCIPAL COMPONENTS ANALYSIS OF DIOXINS FROM THE ROCKY MOUNTAIN ARSENAL & DENVER CO.

Paul D. Jones[1], John P. Giesy[1], John L. Newsted[2], John Slocomb[2], Gerry Henningsen[3], Laura Williams[3], Carol Bicher[4]

[1]Michigan State University, [2]Entrix Inc., [3]USEPA, [4]Gannett Fleming Inc.

Introduction

The statistical analysis of data for 17 PCDD/F congeners in ambient soil samples collected from the Denver Front Range Dioxin Survey are reported. Principal Components Analysis (PCA) was conducted to evaluate the utility of pattern recognition techniques for identification of congener profiles that could characterize typical land uses for risk assessment purposes. The five typical land uses that were studied were: residential, commercial, open space, agricultural and industrial. The results of these analyses are intended to be used to help determine if suspected point sources from the Rocky Mountain Arsenal have released TCDD-like congeners (i.e. "dioxins") above ambient levels and/or in different forms than those found in well characterized reference areas.

Materials and Methods

Samples were collected, analyzed, verified and reported by USEPA Region 8. Sampling locations are provided (Figure 1). Soils were collected from 160 locations divided into 5 pre-designated off-post land-use types (open space, agricultural, residential, commercial and industrial) as well as from 60 locations at the Rocky Mountain Arsenal. The data were validated and provided in a spreadsheet that contained 220 rows, each representing a single surface soil sample location, with 24 columns of results and related sample information. The results consist of 17 columns of data for 17 PCDD/F congeners, 2 columns for G.P.S. data and 5 columns containing sample identifiers and location indicators. Data reports: (http://www.epa.gov/region8/superfund/sites/rma/rmdioxrpt.html).

Assessment of Normality and Data Transformation

Data sets were next tested for normality using a one-sample Komolgorov-Smirnoff test that was visually assisted by using probability plots. Normality is not an absolute requirement for PCA, since PCA is not a strict statistical test (i.e. it does not accept/reject hypotheses based on alpha and beta values). However to minimize the influence of 'outliers' and highly non-normal distributions, data transformations can be used to produce 'nearly normal' data. The issue of data-normality for PCA has been discussed extensively (Wilkinson et al 1996). After log transformation, the distributions for the data were generally normal or near normal. The exceptions to this rule were congeners where a significant number of results below method detection limit (<MDL) (e.g. 2,3,7,8-TCDD).

Comparison of On- and Off-Post Samples
Without specifying a number of factors to be retained, the algorithm returned two factors (PCs) that accounted for 84.5% of the variation, as compared to the 88.3% being explained by three factors. When the algorithm was forced to retain four factors, 90.8 % of the variation in all the data was explained - only an additional 2.7% of the total variation. In addition, the maximum loading for any congener on the fourth Factor was 0.350, indicating that the fourth factor had little power to segregate the samples. For these reasons the PCA that retained three factors is presented here. The component loadings were similar to, but not the same as, those observed for the On-Post or Off-Post PCAs (Table 1). The first Factor was loaded heavily on most of the higher chlorinated PCDF congeners, while the second Factor was more heavily weighted on the higher chlorinated PCDDs. The third Factor was loaded the greatest for 2,3,7,8-TCDD, with the loading for 2,3,7,8-TCDF being the second highest but not as large as it was in the On-Post PCA (represented mostly by Factor 1).

Table 1. Component loadings for PCA on the combined (60 On-post and 160 Off-Post samples) data set.[a]

Congener	Factor 1	Factor 2	Factor 3
PCDF-12378	0.914	0.195	0.205
HxCDF-123789	0.888	0.222	0.13
HpCDF-1234789	0.882	0.351	0.072
HxCDF-123478	0.878	0.406	0.125
HxCDF-123678	0.86	0.385	0.137
TCDF-2378	0.758	0.298	0.37
OCDF	0.753	0.562	0.004
HxCDF-234678	0.701	0.518	0.232
PCDF-23478	0.69	0.557	0.251
HpCDF-1234678	0.604	0.668	0.119
HpCDD-1234678	0.283	0.927	0.174
HxCDD-123678	0.35	0.892	0.191
OCDD	0.246	0.88	0.085
HxCDD-123789	0.352	0.86	0.194
HxCDD-123478	0.37	0.85	0.187
PCDD-12378	0.367	0.786	0.262
TCDD-2378	0.284	0.367	0.847

[a] Relatively large loadings (>0.7) are highlighted

When these results were displayed graphically, the clustering of the various groups of samples could be readily seen (Figure 2). Samples that were collected from the RMA were generally separated from the reference samples along the x-axis (Factor 1). Whereas, the reference sites generally are spread along the y-axis (Factor 2). From the component loadings (Table 1), Factor 1 represents higher chlorinated PCDF congeners (found more predominantly On-site), while Factor 2 represents the higher chlorinated PCDDs.

Conclusions

Many soil samples from the RMA have a PCDD/F signature profile containing higher chlorinated PCDF congeners. In contrast, Off-Post samples are more dominated by the higher chlorinated PCDDs typical of general human activities, such as traffic emissions and light industrial activities. The presence of a unique PCDD/F signature in samples from the RMA does not necessarily indicate a major on-site source of these compounds. Indeed, the relatively diffuse nature of the sample clustering here would argue against the presence of a single large source. The predominance of the PCDD/F congeners in the On-Post samples from the central region of the RMA reflects the historical mixed industrial activities carried out on the site, while many of the 'peripheral' site-wide RMA samples were indistinguishable from the Off-Post reference congener profiles. This PCA analysis of trace-level concentrations of dioxins in surface soil samples was a valuable and sensitive tool that contributed useful information for the environmental risk assessment. The PCA results helped USEPA risk assessors and managers to better distinguish the potential source contributions of the site vs. ambient conditions. This led to improved science-based decisions for remedial action based risk reduction.

References

Wilkinson, L., G. Blank & C. Gruber. 1996. Desktop Data Analysis with SYSTAT. Prentice Hall, NJ, 798 pp.

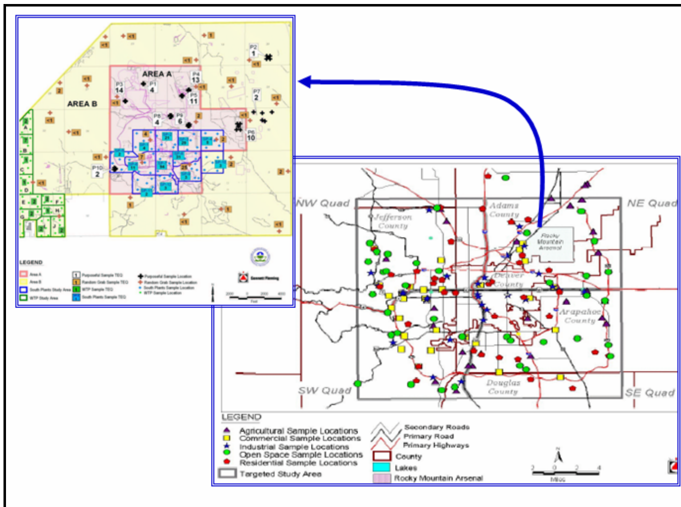

Figure 1. The Denver Front Range Dioxin Survey and soil sample locations.

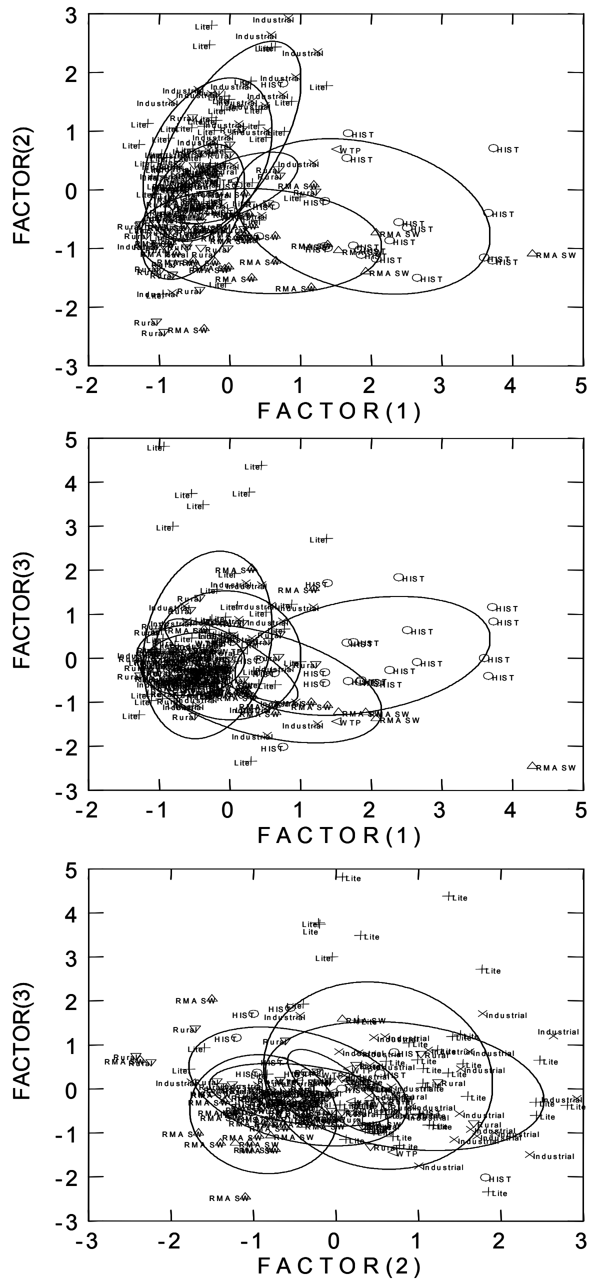

Figure 2 Overall PCA comparing 17 PCDD/F congeners in 60 On-Post and 160 Off-Post soil samples