CLUSTER ANALYSIS FOR THE EVALUATION OF SOIL DIOXIN CONGENER PROFILES FOR A COMMUNITY IN MICHIGAN

Timothy Towey[1], Shu-Chi Chang[2], Peter Adriaens[2], Avery Demond[2], Alfred Franzblau[2], David Garabrant[2], Brenda Gillespie[2], James Lepkowski[2]

[1]LimnoTech, Ann Arbor, Michigan; [2]University of Michigan, Ann Arbor, Michigan

INTRODUCTION AND OBJECTIVES

The University of Michigan Dioxin Exposure Study was undertaken to address concerns among the population of Midland and Saginaw Counties related to the discharge of dioxin-like compounds from the Dow Chemical Company facilities in Midland, resulting in the contamination of soils in the Tittabawassee River floodplain and areas of the City of Midland.

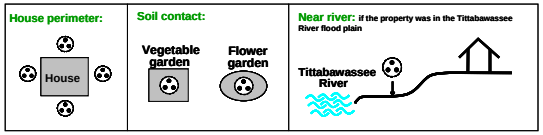
In order to better understand the distribution of PCDD/Fs in the soil of UMDES participants, an analysis of congener patterns in soil samples was performed using multivariate chemometric methods:

- Principal component analysis (PCA) and hierarchical cluster analysis (HCA) were performed on the complete UMDES soil data set (2081 samples).
- This study investigated the impact of linkage methods and cluster quantity on the HCA results.
- The results of the analysis using the preferred linkage method and cluster number are presented.

METHODS

- A total of 2081 soil samples were collected from 766 properties in 5 populations: (Tittabawassee River) Floodplain, Near Floodplain, Other Midland/Saginaw Counties, Midland Plume, and Jackson/Calhoun Counties (the reference population).
- Soil samples were collected from up to seven locations per property from the following zones: house perimeter, soil contact, and near river.

Figure 1. Soil sampling zones



- For this analysis, soil PCB concentrations were not included.
- Data was transformed using natural log (ln(x+1)), constant row sum, and range transformations.
- Seven principal components accounted for 95% of the cumulative variance and were used in the HCA.
- HCA was performed using the distance between samples based on principal component scores. The sensitivity of HCA to different linkage methods and cluster numbers was evaluated.
- The linkage methods that were evaluated were: average, single, complete, centroid, and mean. The sensitivity of cluster quantity on HCA results was evaluated using seven to 14 clusters.

RESULTS AND DISCUSSION

The inclusiveness and specificity of the HCA outcomes were evaluated with respect to an initial, conceptual understanding of the primary dioxin sources.

- A narrowly defined floodplain (NDFP) was delineated as soil samples from the floodplain region with TEQs greater than 100 pg/g.
- A narrowly defined plume (NDPL) was delineated as soil samples from the plume region with TEQs greater than 50 pg/g.
- A broadly defined floodplain (BDFP) was delineated as all soil samples from the floodplain population.
- A broadly defined plume (BDPL) was delineated as soil samples from the City of Midland.
- Analysis that resulted in high percentages in narrowly defined groups and low percentages outside broadly defined groups was considered to have reflected well the initial, conceptual understanding of dioxin sources.
- **Complete linkage was judged to be the best linkage method.**
- **Increasing the quantity of clusters used in HCA offered improved resolution for assessing the distribution of congener patterns.**

Table 1. Evaluation of HCA linkage methods using seven clusters

	Average	Single	Complete	Centroid	Median
% of NDFP samples in FP cluster	99.6	100	99.6	95.3	95.3
% of NDPL samples in PL cluster	91.5	100	90.2	100	100
% outside BDFP in FP cluster	14.1	99.2	16.8	98.6	98.6
% outside BDPL in PL cluster	68.4	99.5	10.5	98.3	98.3

Table 2. Evaluation of HCA cluster quantity using complete linkage

	Number of clusters							
	7	8	9	10	11	12	13	14
% of NDFP samples in FP cluster	99.6	99.6	94.1	94.1	94.1	94.1	75.1	75.1
% of NDPL samples in PL cluster	90.2	90.2	90.2	90.2	90.2	90.2	90.2	90.2
% outside BDFP in FP cluster	16.8	16.8	5.4	5.4	5.4	5.4	1.3	1.3
% outside BDPL in PL cluster	10.5	1.8	1.8	1.8	1.8	1.8	1.8	1.8

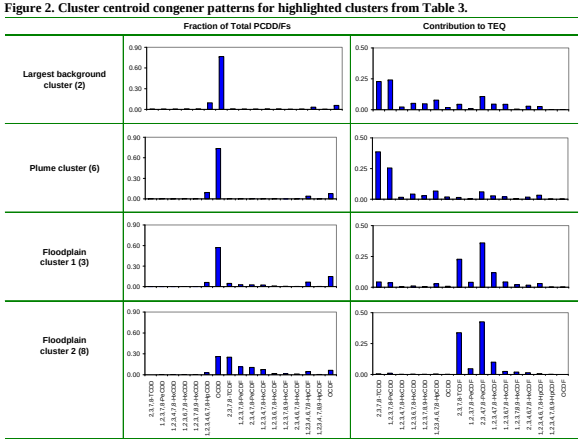
Table 3. TEQ and % contribution from dioxins and furans using 13 clusters and complete linkage

Cluster	Number of samples	Mean $TEQ_{WHO2005-DF}$ (pg/g)	Dioxin contribution to TEQ	Furan contribution to TEQ
1	122	4.95	67.8%	32.2%
2	458	2.35	65.1%	34.9%
3	189	709	15.8%	84.2%
4	379	8.97	67.7%	32.3%
5	430	21.3	24.7%	75.3%
6	132	59.0	77.0%	23.0%
7	152	8.78	45.3%	54.7%
8	113	394	4.3%	95.7%
9	20	5.12	84.8%	15.2%
10	64	0.73	70.7%	29.3%
11	17	10.9	92.0%	8.0%
12	3	1.61	52.2%	47.8%
13	2	29.5	87.5%	12.5%

- Nine clusters performed well with respect to inclusiveness and specificity.
- Thirteen clusters was also thought to be of potential interest because of its increased differentiation of elevated floodplain samples. An" elbow criterion" applied to similarity as a function of number of clusters also supported the use of 13 clusters.
- The results of the cluster analysis using complete linkage and 13 clusters are presented in Table 3 and Figure 2.
- The clusters that have large furan contributions to TEQ are likely associated with the Tittabawassee River contamination.
- Clusters with higher contributions from dioxins are likely associated with combustion sources – both in the Midland Plume and in background.

- **The congener patterns of the plume and background cluster centroids are fairly similar, but are differentiated by the contribution of TCDD.**
- **The two floodplain cluster centroid patterns both have elevated levels of furans, however they differ significantly in their relative fractions of higher chlorinated dioxins.**
- The difference in the two floodplain clusters may be related to time of deposition; FP cluster 2 samples may have been deposited earlier and were therefore not impacted by more recent industrial processes, such as pentachlorophenol production.
- The possibility that the two clusters are differentiated by time is supported by the fact that FP cluster 2 samples occur frequently in the Near Floodplain, where it is possible that soils were moved from the river at an earlier time.

Figure 2. Cluster centroid congener patterns for highlighted clusters from Table 3.



Financial support for this study comes from the Dow Chemical Company through an unrestricted grant to the University of Michigan