CONGENER-SPECIFIC DIFFERENTIATION OF SOIL SAMPLES IN THE CITY OF MIDLAND, MICHIGAN USING MULTIDIMENSIONAL SCALING

Trinh H[1], Garabrant D[2], Franzblau A[2], Hong B[2], Gwinn D[3], Towey T[4], Goovaerts P[5], Demond A[1], and Adriaens P[1]

[1]Department of Civil and Environmental Engineering, University of Michigan College of Engineering, 1351 Beal, Ann Arbor, MI 48109; [2]Department of Environmental Health Sciences, University of Michigan School of Public Health, 109 S. Observatory, Ann Arbor, MI 48109; [3]The Center for Statistical Consultation and Research, University of Michigan, 3550 Rackham Building, Ann Arbor, MI 48109; [4]Limno-Tech Inc, 501 Avis Dr, Ann Arbor, MI 48108; [5]PGeostat, 710 Ridgemont Lane, Ann Arbor, MI 48103

Key words: visualization, spatial analysis, incineration

Abstract
Multidimensional scaling (MDS) was applied to the 51 UMDES soil samples in order to visualize the impact of congener-specific differences on the soil profile. The method was conducted by reducing the dimensions of a multivariate dataset to a lower dimensional space such that distances between points were approximately equal to the dissimilarities (distances) of the points in the multidimensional space. We used classical multidimensional scaling with Euclidean distances in two applications: (1) to map sampling locations based on pair-wise distances of the UMDES sampling locations and of the Dow incinerator without revealing their actual coordinates; and (2) to map a dioxin congener profile by differentiating concentration levels of congeners across each location. This analysis allowed for a spatial differentiation between dioxin contributions to the overall profile as follows: a high level of all congeners were observed close to the Dow incinerator, in particular for 2,3,7,8-TCDD and 1,2,3,7,8-PeCDD. There is a decreasing trend of concentration levels of dioxin congeners further away from the plant incinerator.

Introduction
The UMDES soil sampling campaign was carried out at 51 locations in the city of Midland, Michigan, USA in the vicinity of the Dow Chemical Company incinerator complex for the analysis of dioxins and like compounds. The sampling locations were designed based on the prediction of the dioxins plume from the atmospheric deposition model (Industrial Source Complex, EPA ISCST3) using total TEQ_{98} (dioxins/furans/PCB) measures[1]. Analytical results of 1-inch depth UMDES soil samples provide concentrations of 29 dioxins, furans and PCBs' congeners. Total TEQs excluding PCBs' contributions of these 51 UMDES samples were used to validate the deposition model[2].

Validation of the previously published deposition model indicated that there tended to be an underprediction of the actual measured concentrations in soil, particularly in close proximity of the incinerator. Since the differences in physical-chemical properties of dioxin-like compounds vary widely, it is expected that their deposition behavior may be impacted as well. We are interested in a screening procedure to explain the area where the deposition model failed to predict the total TEQs and in data visualizing methods that enable the release of sampling locations without disclosing the exact sampling locations' coordinates due to binding confidentiality agreements.

Materials and Methods
Multidimensional scaling (MDS) is a data dimension reduction technique which aims to map a multidimensional space (p-dim) into fewer dimensions (k-dim, usually 2 or 3) to visualize the data while still trying to capture as much as possible the underlying variance[3]. Object i (i=1,…,N) is mapped into point x_1, x_2, …, x_N of the reduced k-dim space such that the given dissimilarities $D_{i,j}$ calculated from the p-dim space are well-approximated by the distances $|xi - xj|$ from the k-dim space[4].

The pairwise dissimilarities, $D_{i,j}$ between observations in the data set were calculated by the ***daisy*** function[5] provided in the statistical free software R, based on Euclidean distances, which defined as root sum-of-squares of differences. The dissimilarity matrix was then used as an input for the ***cmdscale*** function in R to return an output as a new configuration of a set of points $x_1,\ldots,x_N$ in which the distances between the points are approximately equal to the dissimilarities[6]. The fitted distances between points are represented by the resulting first, second and third mds components from the ***cmdscale*** function according to the dimensionality (k-dim) selected (k=2 or 3). The scatter plot of these mds components then results in a map of the new data configuration.

$$\mathrm{Stress}_D(\mathrm{x}_1,\ldots,\mathrm{x}_N) = \left(\sum_{i\neq j=1..N} (D_{i,j} - \|\mathrm{x}_i - \mathrm{x}_j\|)^2 \right)^{1/2}$$

The goodness of fits of multidimensional scaling method can be measured by the stress function, the residual sum of squares (Stress_D) or by other measures such as the square root of Stress_D normalized by the sum of fitted distances[3,4].

We applied multidimensional scaling technique to the 51 UMDES data (numbered) with dissimilarities calculated based on the pairwise distance between the locations of 51 samples and the location of the Dow incinerator (number 52) on the one hand, and the pairwise measures of concentrations of each of the 29 congeners at 51 sampling locations on the other hand. The former analysis gives a visualization of the sampling location map while the latter provides a congener profile map. The WHO congeners includes: (i) 7 Polychlorinated dibenzo-p-dioxins: 2,3,7,8-TCDD;1,2,3,7,8-PeCDD; 1,2,3,4,7,8-HxCDD; 1,2,3,6,7,8-HxCDD; 123789-HxCDD; 1234678-HpCDD; OCDD; (ii) 10 Polychlorinated dibenzofuran: 2378-TCDF; 12378-PeCDF; 23478-PeCDF; 123478-HxCDF; 123678-HxCDF; 234678-HxCDF; 123789-HxCDF; 1234678-HpCDF; 1234789-HpCDF; OCDF; (iii) 12 Polychlorinated biphenyls: PCB-77; PCB-81; PCB-126; PCB-169; PCB-105; PCB-114; PCB-118; PCB-123; PCB-156; PCB-157; PCB-167 and PCB-189;

Results and Discussion

Multidimensional scaling was performed on 7 scenarios in which the dissimilarities matrix was calculated based on the first 5 dioxin variables, first 6 (i.e. 2,3,7,8-TCDD;1,2,3,7,8-PeCDD; 1,2,3,4,7,8-HxCDD; 1,2,3,6,7,8-HxCDD; 123789-HxCDD, and 1234678-HpCDD) and 7 dioxin variables; all furans variables; combination of all dioxins and furans variables; all PCBs' variables and all 29 congener concentration variables. Two dimensions (k=2) was selected for the fitted space. The results of the analysis are shown in Figure 1 for the aforementioned set of variables (all points in each plot refer to the sample locations and the incinerator (52)).

It is interesting to note that the model based on 5, 6, and 7 dioxin variables indicated a greater distribution of the data points than the cases which included more variables. The extent to which the points are scattered along both dimensions is an indicator for how well the set of variables explains the differences between the samples. The goodness of fits from the scenario with dissimilarities calculated based on the 6 dioxin variables (lower subset of figure 1) was 0.002, suggesting a perfect fit of the data set into the 2-dim space, thus this scenario was selected for the data visualizing purposes.

The multidimensional scaling map is interpreted by looking how near (closeness) and/or far (disparity) the points are to other points and finding the underlying variance at each data point. The outer most part of the map has two points denoted by open hexagonal symbols. The soil concentrations of all congeners at these two locations are the highest among 51 UMDES samples, placing them far apart from the rest of the samples. The next set of points denoted by cross-hair square symbols, is different from other points due to their congener concentrations scored second highest among all locations. Moving toward to the largest cluster of points on the left, we observed that levels of all congeners, in particular of 2378-TCDD, 12378-PeCDD and 1234678-HpCDD concentration are decreasing. The

plus signs presented a group of samples that is dissimilar from the cross sign symbol group in level of 1234678-HpCDD.

Given the pairwise distances between 51 sampling locations and the Dow incinerator, we projected the MDS map on the sampling location map (figure 2). The proximities between the sampling locations and between sampling locations to the incinerator in the fitted 2-dim space are well-retained but the map rotation and direction are arbitrary. To avoid any attempts of rescaling the map to its actual map, the first and second mds component (on x, y axis) were resized one more time. The MDS of congener profile at 51 locations from previous discussion was superimposed onto the location map by coding with the same symbols. Thus, figure 2 is a joint presentation of multidimensional scaling of both location map (dissimilarities in pairwise distances) and congener profile map (dissimilarities in pairwise congener concentrations).

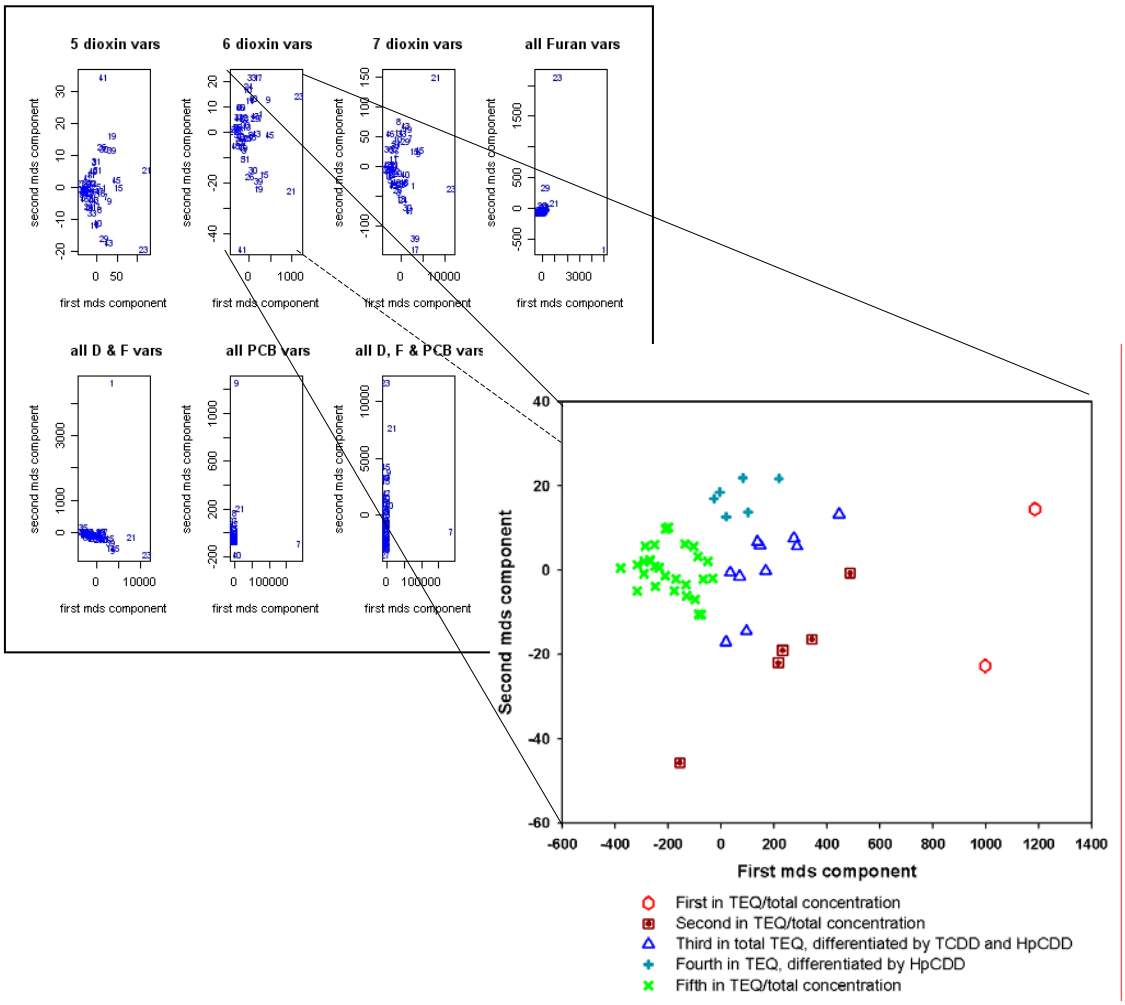

Figure 1 Multidimensional scaling (MDS) of 51 UMDES soil samples based on Euclidean distances between 5, 6, 7 dioxin congeners; all furan congeners, combination of dioxins and furans; all PCBs and all congeners. Subset figure is the enlargement of the multidimensional scaling based on Euclidean distances of 6 dioxin congeners. Color coded based on the congeners' concentration differences.

The combined map revealed a fact that levels of soil concentrations of all congeners are decreasing along the direction further away from the incinerator. Close to the incinerator, the TCDD and PeCDD concentrations were the highest observed. The contributions of TCDD, PeCDD and HpCDD to the TEQ decrease with distance from the incinerator, indicating potentially different sources of soil contamination. Considering the spatially differentiated ratios of the six dioxin congeners, care needs to be taken to spatially interpolate TEQ values. Hence, it is likely that sample heterogeneity has contributed to the underestimation of the total TEQs measured at locations close to the incinerator (figure 3). The implications from an exposure modeling perspective is that the 'plume' population may be limited to study participants living in close proximity, whereas those living further removed from the incinerator (but still predicted to be in the plume) are in fact exposed to background contamination, and other dioxin contamination sources.

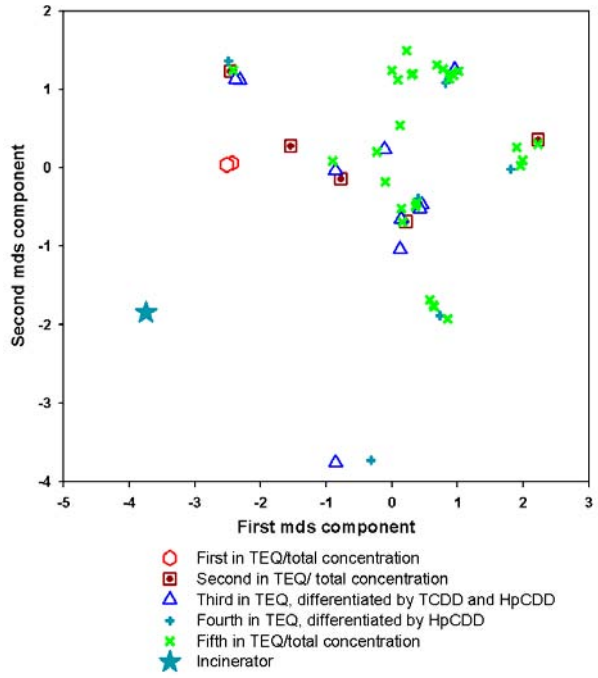

Figure 2 Multidimensional scaling map of sampling locations. Symbols coded according to congener profile.

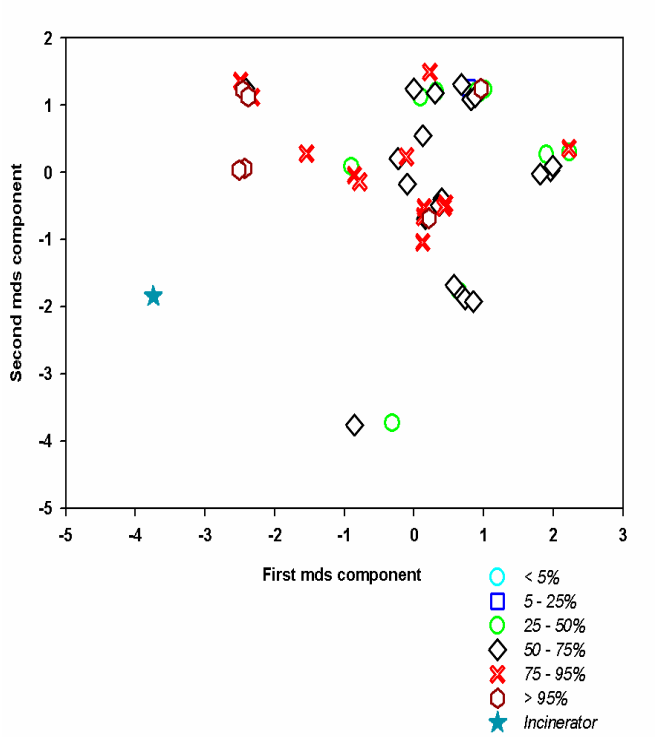

Figure 3 Multidimensional scaling map of sampling locations. Symbols coded according to actual observations falling into prediction range (%) of deposition model.

Acknowledgements

Financial support for this study comes from the Dow Chemical Company through an unrestricted grant to the University of Michigan. The authors acknowledge Ms. Sharyn Vantine for her continued assistance and Drs. Linda Birnbaum, Ron Hites, Paolo Boffetta and Marie Haring Sweeney for their guidance as members of our Scientific Advisory Board.

References

[1] Goovaerts P., Adriaens P. 2006 (abstract Dioxin 2007 Conference)
[2] Goovaerts P., Trinh H. and Adriaens P. 2007 (Dioxin 2007 Conference)

[3] Michailidis G., Multivariate Statistics Lecture Notes, University of Michigan 2007
[4] Buja A., Swayne DF, Littman ML, Dean N., Hoffmann H. 2004, AT&T Bell Laboratories
[5] Kaufman L. and Rousseeuw PJ. In: *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York 1990
[6] Venables WN, Smith DM andand the R Development Core Team. In: An Introduction to R (Manual version 2.4.1) 2006