CONSTRUCTION AND VALIDATION OF A GEOSTATISTICAL MODEL OF PCDD AND PCDF DEPOSITION FROM INCINERATION

Goovaerts P[5], Trinh H[1], Garabrant D[2], Hong B[2], Gwinn D[3], Towey T[4], Lee S-Y[6], 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; [6]Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, Michigan 48104.

Keywords: geostatistics, TEQs, deposition model, stochastic simulation, uncertainty

Abstract
Assessment and predictions of exposure of populations impacted by incinerator plume deposition depends on the accurate spatial delineation of the impacted area. A geostatistical model was constructed using a regression between publicly available soil dioxin data and dry deposition, to provide guidance for the collection of new human and soil data in the study area. Although these 51 new soil measurements were confined to a few census blocks and so are not representative of the entire area that was initially characterized, the validation study showed that the model of local uncertainty is accurate and precise. The predicted values were on average lower than observed values, and the uncertainty model indicates that the underestimation occurs mainly in the vicinity of the plant property line. This information can be used to update the estimation model for spatial distribution at the census block level.

Introduction
Since the deposition of pollutants around incinerators displays complex spatial patterns depending on prevailing weather conditions, the local topography and the characteristics of the source. Deterministic dispersion models often fail to capture the complexity observed in the field, resulting in uncertain predictions that might hamper subsequent decision-making, such as delineation of areas targeted for additional sampling or remediation. Geostatistics[1,2] provides a set of methods for incorporating the spatial coordinates of field data in the mapping of pollutant levels and the assessment of the attached uncertainty. This paper describes a geostatistical methodology to combine field data with the predictions of dispersion model. The approach generates a set of equally-probable maps of the spatial distribution of pollutants which can be post-processed to compute the probability that target thresholds are exceeded locally or on average over polygons of various size (i.e. census blok units). The methodology is used to delineate areas with high level of dioxin TEQ around an incinerator in Midland, Michigan. The accuracy and precision of the geostatistical model is then assessed using recently collected soil data.

Materials and Methods
The information available for modeling the distribution of soil TEQ around the incinerator consists of: 1) 53 geo-referenced soil TEQ concentrations collected during 4 sampling campaigns from 1984 through 1998; and 2) air concentration and total deposition flux values (both dry and wet) predicted at the nodes of a 500×500 receptor grid (spacing = 50 m) using EPA Industrial Source Complex (ICS3) dispersion model[3]. The deposition model was run using hourly meteorological data (e.g. wind speed, ambient temperature, precipitation rate) available for 1987-1991. Major differences were observed between the spatial patterns of dry and wet depositions; while higher air concentration and dry deposition are observed on the North-eastern side of the plant (i.e. downwind), important wet deposition is predicted on the South-western side of the plant (Figure 1). The soil TEQ_{98} (dioxins/furans/PCB) concentrations range between 0.60 ppt and 450 ppt, with a mean value of 73.66 ppt.

1

The uncertainty attached to the TEQ value within each census block was modeled using the following geostatistical methodology: (i) The TEQ concentrations are normal score transformed to correct for the strongly positively skewed histogram; (ii) The transformed data are regressed against the air concentration and deposition (wet and dry) values predicted by the numerical dispersion model. This regression model, which explains 45.3% of the total variance in TEQ data, is used to predict the TEQ concentration and standard error at the nodes of the 500×500 receptor grid; (iii) The spatial variability of regression residuals is modeled using the semivariogram; (iv) Sequential Gaussian simulation[1] is used to simulate the spatial distribution of TEQ values conditionally to the 53 TEQ data, the trend model inferred from the calibration of the deposition data (step 2) and the pattern of correlation modeled in step 3. One hundred realizations were generated using a 500×500 simulation grid with a spacing of 50 m; (v) Point simulated values are aggregated within each census block to yield a simulated block value (upscaling). This aggregation is repeated for each realization, yielding a set of 100 simulated values for each census block.

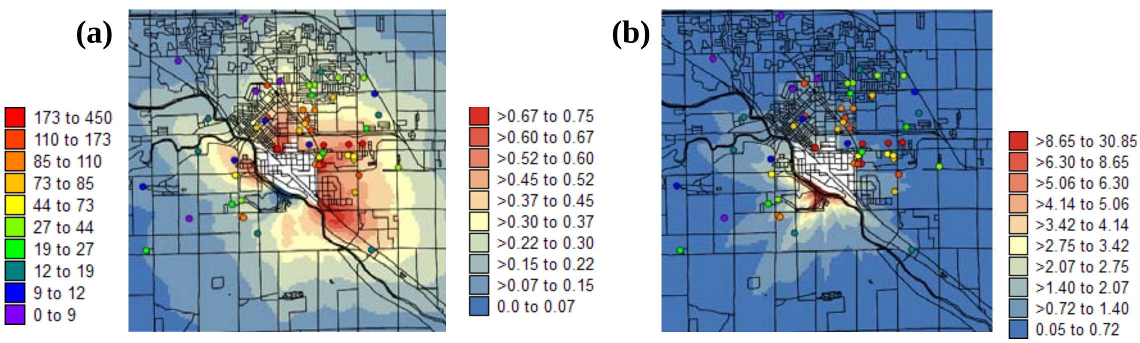


Figure 1. Grid of 5-year dry (a) and wet deposition (b) values predicted by the dispersion model (units= μg/m^2). Dots depict the 53 soil TEQ concentrations, while the outlines of census blocks are displayed in background.

The distribution of 100 block values was used to retrieve the probability for each census block to exceed a negotiated threshold of 90 ppt. Census blocks that were the most likely to exceed this threshold and have the largest population at risk were targeted for a recent soil sampling campaign, which led to the collection of 51 new soil samples. The soil TEQ concentrations range between 4.90 ppt and 923 ppt, with a mean value of 79.15 ppt. Each new soil measurement was compared to the set of 100 TEQ values simulated at the closest grid node. The validation stage proceeded as follows: (i) Boxplots were used to visualize where the measured value falls within the distribution of 100 simulated values; (ii) The correlation between the mean of 100 simulated values and soil measurements was computed; (iii) Prediction errors were mapped to identify any spatial pattern for the over and under-estimation of TEQ values; (iv) The accuracy of the model of uncertainty was quantified by plotting the expected versus observed proportions of measurements that fall within median-centered probability intervals (PI) of increasing size (accuracy plots[4]). The average width of the PIs that include the new observations informs on the precision of the models of local uncertainty. The yardstick is the width of the probability intervals, derived from the global histogram, that include the same proportion of observations.

Results

The regression residuals are spatially correlated with a range of 3,000 feet, which indicates the presence of spatially structured variability that cannot be explained by atmospheric deposition. Figure 2 (top maps) shows two simulated maps of the spatial distribution of TEQ values around the incinerator. As expected, the largest TEQ values are found close to the plant property line. Differences between realizations illustrate the uncertainty attached to the exact TEQ value at those locations. Point simulated values are aggregated to the level of census blocks which represent our decision support, see Figure 2 (bottom maps). The probability map (Figure 2e) shows that census blocks located South and East of the plant are the most likely to exceed the State of Michigan cleanup threshold of 90 ppt, which reflects the impact of the wet and dry deposition patterns displayed in Figure 1.

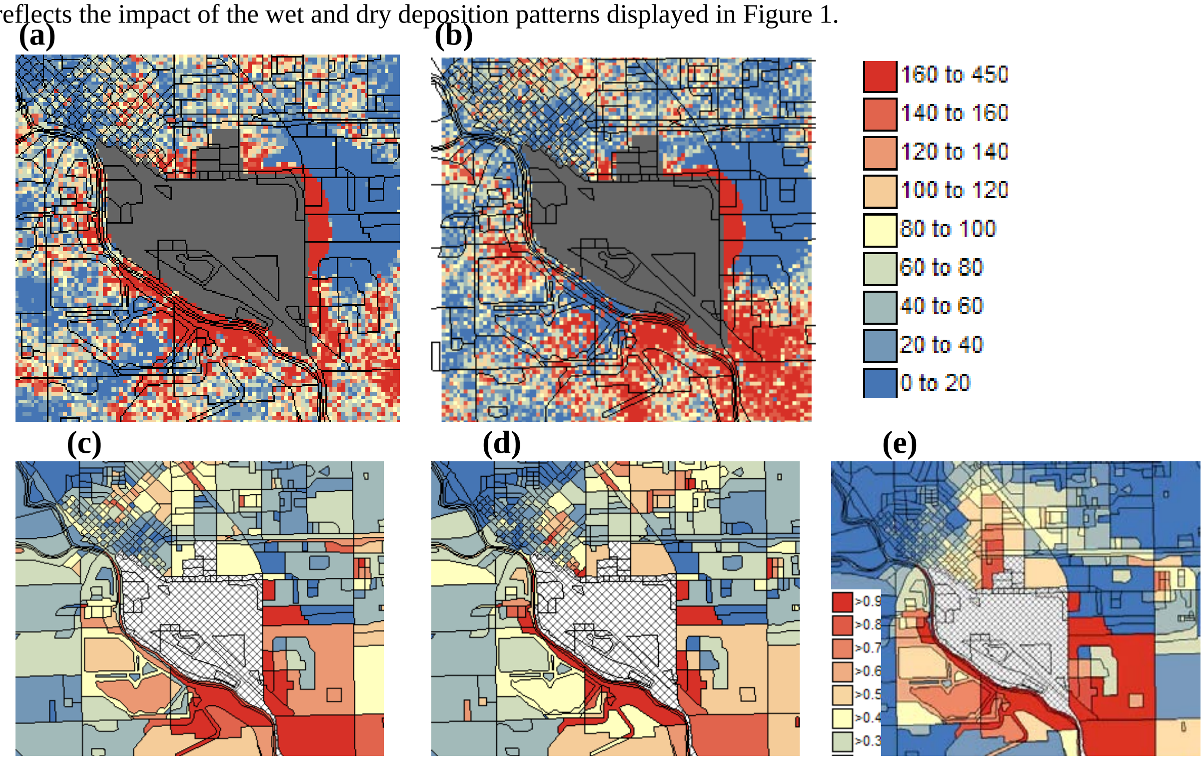


Figure 2. Two realizations (subsets of the original 500×500 grid) of the spatial distribution of TEQ values (a,b), and the results of the aggregation to census block level (c,d). The probability for each census block value to exceed a threshold of 90 ppt is computed from the set of 100 realizations (e).

Each of the 51 new observations is compared to the distribution of 100 TEQ values that was simulated at the closest grid node. For the example in Figure 3a, the observed value (53.7 ppt) falls within the range of simulated values and is very close to the mean of the distribution (47.6 ppt). The scatterplot in Figure 3b indicates a good agreement between observed and mean simulated values (correlation=0.42), although predicted values are on average lower than observed values (48.57 versus 79.15 ppt). The map of prediction errors (not shown for confidentiality reasons) indicates that the underestimation occurs mainly in the vicinity of the plant property line. Figure 4a shows that the geostatistical model of uncertainty is accurate: the proportion of observations that fall within probability intervals (PI) exceeds what is expected from the model. For example, the observed TEQ value is included in the 0.5-PI for 61% of the 51 new samples (expected proportion=50%). Not only should the true TEQ value fall into the PI according to the expected probability p, but this interval should be as narrow as possible to reduce the uncertainty about that value. The average width of these local PIs should also be smaller than the global PI inferred from the sample histogram. The scatterplot in Figure 4b indicates that, for all probabilities p, the local PIs are narrower than the corresponding global PIs, which means that the geostatistical model of uncertainty is both accurate and precise.

Discussion

The approach described in this paper combines the detailed process-based modeling of atmospheric deposition from an incinerator with the probabilistic modeling of residual variability. The benefit of stochastic simulation over spatial interpolation is two-fold: 1) maps of simulated point TEQ values can easily be aggregated to the geography that is the most relevant for decision making (e.g. census block, ZIP codes), and 2) the uncertainty at the larger scale is simply modeled by the empirical distribution of aggregated simulated values. The geostatistical model provided guidance for the collection of new human and soil data in the study area. Although these 51 new soil measurements were confined to a few census blocks and so are not representative of the entire area that was initially characterized, the validation study showed that the model of local uncertainty is accurate and precise. Probability intervals provide a realistic assessment of the range of possible TEQ values that could be observed at unsampled locations, and it is more precise than the aspatial approach whereby the uncertainty model is based on the global histogram of the data. The incorporation of newly collected data into the regression and simulation procedures is straightforward, leading to updated models of the spatial distribution of TEQ values.

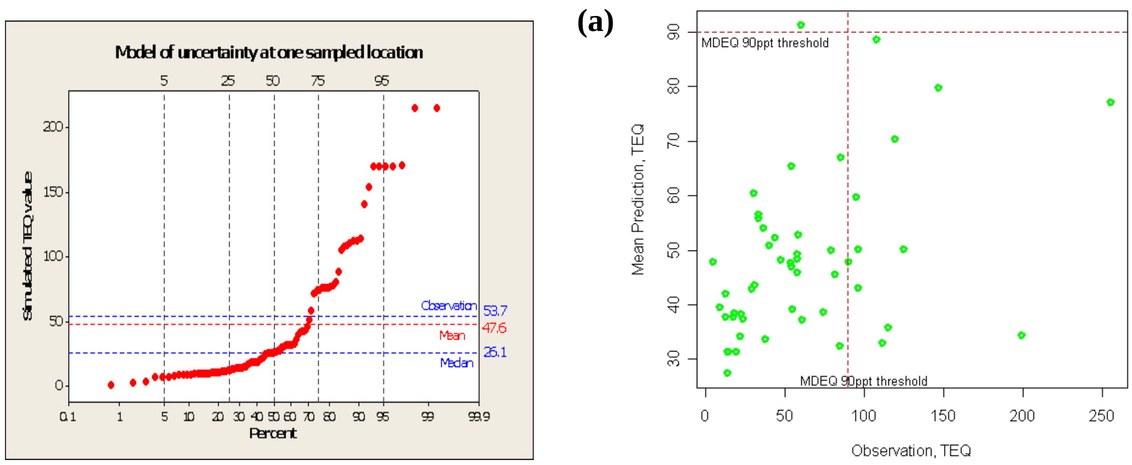
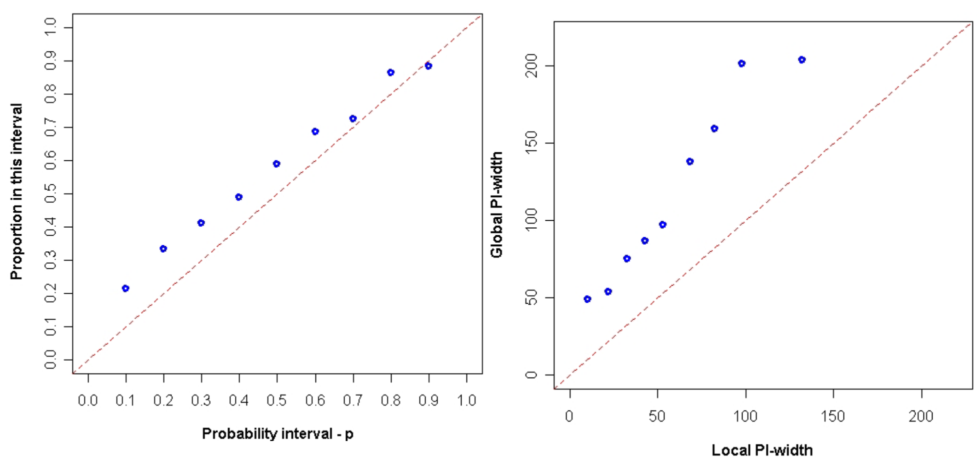


Figure 3. (a) Cumulative distribution of 100 TEQ values simulated at the node the closest to one of the 51 new observations. (b) Scatterplot of observed values versus the mean of the 100 simulated TEQ values (the maximum observation of 923 ppt is not included for graph clarity).

Figure 4. Proportion of observed TEQ values falling into probability intervals (PI) of increasing size (a). The width of these local PIs is plotted against the width of the global PIs derived from the sample histogram (b).



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. *Geostatistics for Natural Resources Evaluation*. New York, Oxford University Press, 1997.
[2] Chiles JP, Delfiner P. *Geostatistics: Modeling Spatial Uncertainty*. New York, JW Wiley and Sons, 1999.
[3] US EPA. *User's Guide for the Industrial Source Complex (ISC3) Dispersion Models*. EPA-454/B-95-003a. Research Triangle, NC, 1995.
[4] Goovaerts, P. *Geoderma* 2001, 103: 3-26.