Journal of Medical Systems, Vol. 28, No. 4, August 2004 (© 2004)

Geostatistics and GIS: Tools for Characterizing Environmental Contamination

Shannon L. Henshaw,[1] Frank C. Curriero,[2,4] Timothy M. Shields,[3] Gregory E. Glass,[3] Paul T. Strickland,[1] and Patrick N. Breysse[1]

Geostatistics is a set of statistical techniques used in the analysis of georeferenced data that can be applied to environmental contamination and remediation studies. In this study, the 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene (DDE) contamination at a Superfund site in western Maryland is evaluated. Concern about the site and its future clean up has triggered interest within the community because residential development surrounds the area. Spatial statistical methods, of which geostatistics is a subset, are becoming increasingly popular, in part due to the availability of geographic information system (GIS) software in a variety of application packages. In this article, the joint use of ArcGIS software and the R statistical computing environment are demonstrated as an approach for comprehensive geostatistical analyses. The spatial regression method, kriging, is used to provide predictions of DDE levels at unsampled locations both within the site and the surrounding areas where residential development is ongoing.

KEY WORDS: 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene (DDE); 1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane (DDT); kriging; R; Superfund.

INTRODUCTION

Evaluating spatial relationships and geographic determinants of health is a growing area of environmental and public health research. Traditional environmental health studies have evaluated temporal changes in diseases and their determinants. More recently, however, there is a growing interest in evaluating both spatial and temporal patterns of health and their environmental determinants. As a result, geographic information systems (GISs) have become increasingly popular

[1]Department of Environmental Health Sciences, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.
[2]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.
[3]Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.
[4]To whom correspondence should be addressed; e-mail: fcurrier@jhsph.edu.

in environmental health applications and to environmental health practitioners.[1–8] As GIS becomes more widespread, various applications of GIS have evolved. Articles in the literature provide various definitions and applications of GIS as a medical, environmental, and public health information system. In its basic form, GIS is a database system with the distinguishing feature that it deals with spatially (or geographically) referenced data. "Where" in addition to "what" that is measured or observed is important and thus recorded and stored in a GIS database. With location information linked to data values, GIS becomes a visual database providing comprehensive mapping capabilities as its most basic and fundamental construct. However, the functionality of a GIS extends beyond producing maps. Some of these more specialized features are highlighted in this study with an application involving environmental contamination at a Superfund site.

The purpose of this paper is to present combined components of GIS and the R statistical computing environment as tools for performing geostatistical- and environmental-based analyses. Data from a western Maryland Superfund site are used to demonstrate different features of GIS and R as they are needed to follow through a geostatistical analysis characterizing potential contamination at the site. Background information on the Superfund site, geostatistics, and the R statistical computing environment are provided first.

BACKGROUND

Description of the Site

In the early 1930s, a large chemical company built a 19-acre facility in western Maryland for the production of fertilizers and pesticides, including blending 1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane (DDT) until 1968. DDT and its derivatives, 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene (DDE) and 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethane (DDD), are classified by the United States Environmental Protection Agency (EPA) to be Group B2, probable human carcinogens.[9] These compounds have also been found to be endocrine disrupters, and acute doses can severely affect the nervous system.[9,10] In 1974, the use and production of DDT was banned by the United States. The chemical company ceased all fertilizer and pesticide operations at the site in 1984, however, many original structures remain today.

Subsequent site investigations involving soil, water, air, and fish sampling during the 1970s, 1980s, and 1990s indicated the presence and migration of these pesticides and other organic and inorganic contaminants to off site areas, some of which are currently under residential development.[11–15] Furthermore, the municipality in which the site is located has a population of approximately 60,000 residents. The EPA placed the site on the National Priority List (NPL) for cleanup in 1997, designating it a Superfund site.[16,17] For demonstration purposes, soil sample data collected on and near the site in the 1990s analyzed using both GIS and statistical techniques are presented in this paper.

Geostatistics and R

The area of statistics known as geostatistics is ideally suited for characterizing the spatial distribution of environmental contamination and for evaluating potential spatial determinants. Geostatistics is a set of statistical techniques used in the analysis of spatially referenced data.[18–20] Originally developed in the mining industry to predict recoverable ore reserves, the practice of geostatistics has found widespread use in environmental applications such as Superfund site characterization.[21–25]

Geostatistics provides the ability to predict levels of contamination at unsampled locations on the basis of known sampled values. Intuitively, predicting contamination levels at unsampled locations can be thought of as a weighted average of neighboring sampled values, with higher weight given to values closer to the prediction location. This can be accomplished formally using the following regression model,

$$Y(s) = \mu(s) + \varepsilon(s), \tag{1}$$

where s denotes spatial coordinates, $Y(s)$ represents contaminant values at location s, $\mu(s)$ the mean component, and $\varepsilon(s)$ corresponds to a zero mean normally distributed random error. Model 1 resembles the common linear regression model with one major distinction; the error component, $\varepsilon(s)$, is assumed to be spatially dependent. Levels of contamination at the site are likely to be more similar at samples closer together than samples further apart. This spatial dependence is accounted for in model 1 by allowing the correlation between error terms to be structured as a decreasing function of the distance between sample locations.

There exists a pool of valid correlation functions that are routinely used in geostatistics, each parameterized to allow the data to help determine its complete structure. The function chosen for the correlation and, more importantly, the estimation of its parameters comprise a crucial step in the geostatistical process. The completely specified correlation function determines the weighting scheme that yields optimal spatial predictions at unsampled locations, generating what are statistically known as best linear unbiased predictions (BLUPs). Once the correlation function has been determined, the BLUPs under model 1 are generated via statistical least squares techniques.[20]

A primary focus in geostatistics is thus specifying this correlation function. Traditionally, this has been accomplished with a closely related function known as the semivariogram.[20] The spherical semivariogram function shown below is an example of a semivariogram commonly used in geostatistics,[20]

$$\gamma(h) = \left\{ \begin{array}{ll} 0 & h = 0 \\ \tau^2 + \sigma^2(1.5 \times (h/\varphi) - 0.5 \times (h/\varphi)^3) & 0 < h \leq \varphi \\ \tau^2 + \sigma^2 & h > \varphi. \end{array} \right\} \tag{2}$$

The parameter, φ, is the range, and corresponds to the distance (h) at which sample locations are no longer spatially correlated. Parameters τ^2 and σ^2, the nugget and partial sill, respectively, define the variance of the process. The nugget represents what is sometimes referred to as microscale variation, or variation between sample locations at a distance smaller than that observed in the sampling. By determining values for these parameters, the semivariogram is specified, and is easily transformed into a correlation structure used for generating BLUPs.[20]

The traditional geostatistical approach first estimates the semivariogram on the basis of the sampled data using the classic method of moments estimator, and then fits a semivariogram function through these estimates with a graphically based procedure such as weighted least squares.[20,26] More recently, because of the subjectivity involved in this method, there has been a push toward a more model-based approach making use of statistical likelihood concepts for semivariogram parameter estimation.[27–30]

Note when information in the form of covariates are available, the mean component, $\mu(s)$, in model 1 may be parameterized as $\mu(s) = \beta_0 + \beta_1 X_1(s) + \cdots + \beta_n X_n(s)$, where the Xs represent covariates and βs, their associated effects. In this case, model 1 is known as a universal kriging model. If no auxiliary information is available, the mean component is assumed to be unknown and constant across the site, $\mu(s) = \mu$, yielding what is known as the ordinary kriging model. For those familiar with regression applications, this may seem unusual. However, prediction is often the primary goal of the kriging model, as opposed to estimation of covariate effects.

Specialized software packages exist to perform geostatistical analyses.[31–34] The open source statistical computing environment R is used in this study. R is similar to the S and S-Plus statistical computing package (MathSoft), with main advantages being that it is free and continually expanded with contributed libraries for a wide range of specialized topics. R has an extensive range of built-in statistical techniques that can be used to perform standard statistical analyses from exploratory to more confirmatory model-based analyses. R is compatible on PC, Unix, Linux, and Macintosh platforms. In addition, R has over 200 official open source add-on libraries for other specialized techniques with over 20 of these libraries containing tools to perform geographic/spatial analyses.[32] Furthermore, R has powerful and unique graphical capabilities as is evidenced in this analysis. R has widespread usage in the academic community and increasing usage in industrial and government sectors. More detailed discussions about R and its contributed libraries are found elsewhere.[35]

METHODS

Environmental Data Source

Soil samples were collected at the western Maryland Superfund site and analyzed for DDT, DDE, and DDD among others at the request of the Maryland Department of the Environment (MDE). The results of these investigations were published in reports which were attained via the Freedom of Information Act.[11–15] Within these reports, assayed values were presented in table form along with survey-type maps of the site that included drawings of buildings and other landmarks. Locations of sampled data were marked on these maps and could be cross-referenced to the assayed value using the corresponding unique sample identifier. However, actual geographic coordinates of each sample or building structure did not exist.

In soil, DDT usually breaks down into two derivate forms, DDE, a product of dehydrogenchlorination, and DDD, a product of dechlorination.[9,36] For this study, 110 DDE surface samples (defined in this study as a depth of less than 1 ft) collected between 1992 and 1997 were used to characterize the spatial distribution of contaminants throughout the site. Multiple samples collected at the same location, time, and surface depth were averaged.

GIS

The DDE data were input into a Microsoft Excel database and imported into ArcGIS 8.3 Desktop (Environmental Systems Research Institute, Inc.). Geographic coordinates for each sample, needed to perform the geostatistical analyses, were determined by creating a georeferenced map of the site and surrounding area, digitizing all sample locations, and creating a geocoded database containing DDE concentrations by location.

Georeferencing is the process by which a map is made electronic, and geographically located on the globe.[37] The process involves obtaining geographic coordinates of some landmark structures on the site. In this study, a hand-held global positioning system (GPS) unit was used to record the geographic coordinates of the corners of each building and fence line during a site walk-through. Maps were scanned into ArcGIS and anchored, or georeferenced, using these landmark coordinates.

Geographic coordinates for the sample locations could not be obtained via a GPS since the ground did not contain any such markings to designate sample locations. Instead, sample locations were identified on the original site maps and made into points on the georeferenced ArcGIS site map with the landmark buildings and boundary fence lines in place. The computer mouse and ArcGIS editor tools were used to produce these points. This process is known as on-screen digitizing and resulted in an approximate set of geographic coordinates for the DDE sample locations.[37]

In order to link the soil contaminant values with the soil sample locations, DDE concentrations were geocoded to the map using ArcGIS. Geocoding allows the user to link any piece of information, or attribute, to a geographical location which already exists in the GIS.[37] In this case, levels of DDE found in soil are linked to the locations from which they were collected on the site layout map via a map identifier that was created when the sample location points were added to the map.

Geostatistics and R

The data, now spatially referenced, were analyzed using geostatistical techniques. The end result of the analysis is a map of the Superfund site characterizing predicted levels of DDE contamination based on the 1992–1997 soil samples.

The ordinary kriging regression model 1 was used to predict levels of DDE at unsampled locations since no auxiliary information was available for the DDE sample data. The spherical semivariogram function 2 was used for the correlation structure. Parameter estimates of the spherical semivariogram were obtained using the composite likelihood approach.[28] The composite likelihood approach for semivariogram estimation has many of the advantages of a full statistical likelihood based approach with the robust and more positive properties of the traditional approach.[20,28]

To assess predictive performance of the kriging model, five DDE samples were randomly chosen from each quartile of the 110 data values, and set aside for validation. These 20 data points did not contribute in the initial analysis. Instead, the remaining 90 values were analyzed and used to predict at these 20 validation locations. The root mean squared error between the true observed DDE values and the predicted DDE values was used to assess the model's ability to predict levels of DDE at unsampled locations. The geostatistical process was then repeated using the full set of 110 DDE samples to produce a map of predicted DDE contamination across the entire site.

All statistical computing and graphics (excluding ArcGIS maps) were performed within the open-source R statistical computing environment. The analysis conducted in this study used the contributed library geoR, a comprehensive set of functions to perform geostatistical-based analyses.[38,39] The geoR library contains written functions to perform most traditional-based geostatistical methods as well as many likelihood-based and Bayesian methods. A more detailed discussion and tutorial about geoR can be found online.[35,39] The flexibility of R also allowed programming the composite likelihood method for semivariogram estimation.

RESULTS

Georeferencing the original Superfund site map as well as digitizing and geocoding the soil sample locations yielded a fully functional GIS map (Fig. 1). The map of the Superfund site displays buildings, boundaries, and the 1992–1997 DDE soil sample locations used for analysis. Buildings on the site included an old pesticide plant (which burned down), fertilizer plant, mixing building, warehouse, laboratory, maintenance shop, electrical shop, and boiler house. The site is bordered by railroad tracks and undeveloped land on the southwest side, a road and industrial property on the southeast border, and residential developments along the northwest and northeast fence lines.

Soil samples collected throughout the site for various reasons characterize the extent of contamination.[11–15] Intensive sampling was performed along the northwest fence line to determine whether the fence should be extended to contain any contamination that had migrated or spilled offsite. Single-family homes were being built in this region and continue to be built to this day. As a result of this sampling, the EPA mandated that the fence be extended by 20 ft to protect the health of those residents in 1996.[12]

From 1992 to 1997, 110 soil samples were taken at the site and analyzed for DDE. The distribution of the DDE sample values were highly skewed to the right and ranged over several orders of magnitude (Fig. 2a). On the basis of subsequent model validations, the DDE samples were log transformed for all analyses with final predictions transformed back to the original scale. Such an approach is common in geostatistics for environmental data with similar characteristics.[20,40–42] An initial summary of the spatial distribution for the transformed DDE samples is shown in Fig. 2b. The apparent clustering of similar values in this plot, with respect to the logged DDE quartiles, is evidence of spatial dependence. Also of interest is the close proximity between several extremely high and moderately low sample values. For example, on the original scale, levels of 0.5 parts per million (ppm) DDE were found in samples within 100 ft of samples with 28 ppm DDE.

Spatial dependence is more formally explored by estimating a semivariogram using the classic method of moments estimator, which are shown as 13 summary points in Fig. 3. The rising and leveling-off pattern in the estimated semivariogram is an

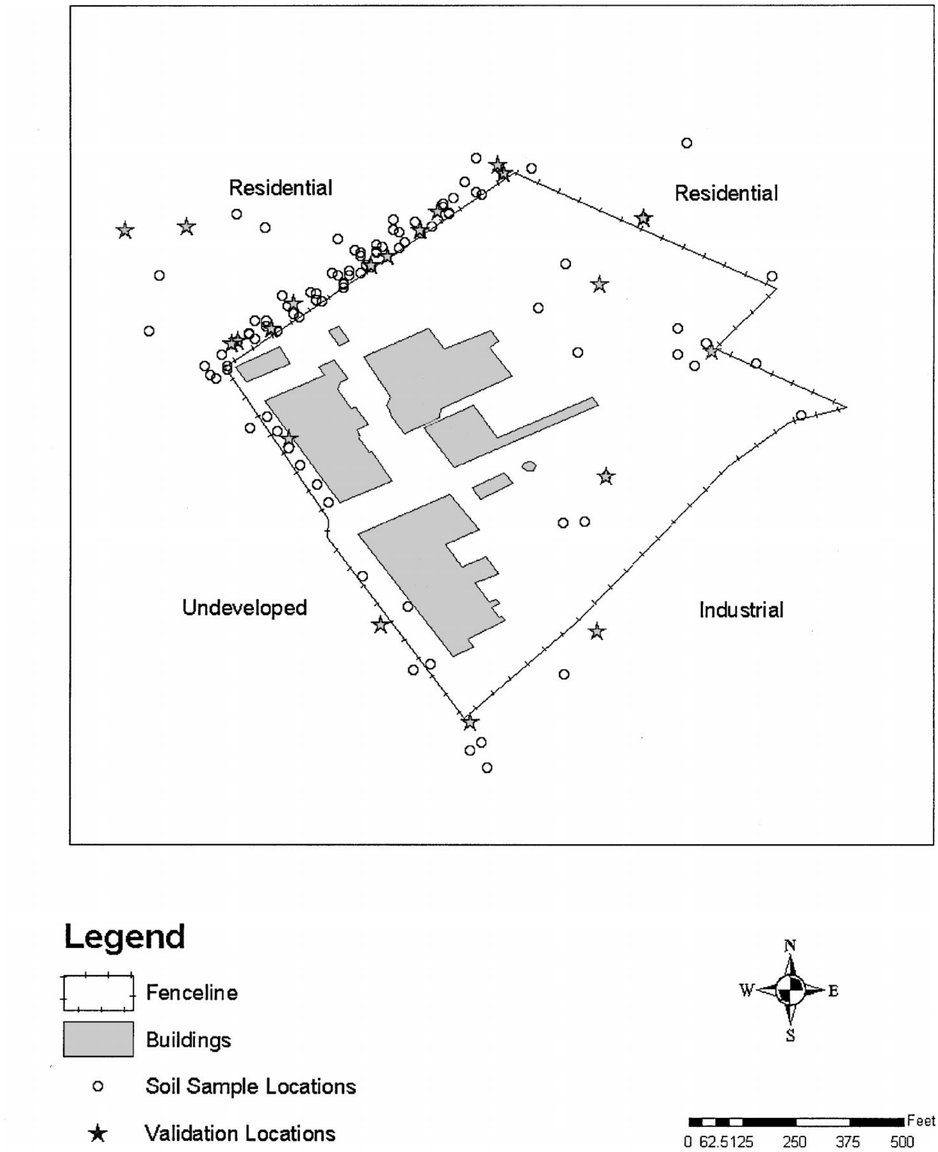

Fig. 1. A GIS-based map of the western Maryland Superfund site. Included are the 110 DDE soil sample locations from 1992 to 1997 indicating the 20 locations set aside for model validation.

indication that the DDE sample values are spatially dependent. Lack of spatial dependence would be evidenced by a horizontal band of estimates showing no pattern. The composite likelihood estimated spherical semivariogram function is also shown in Fig. 3 with a solid line. The range parameter was estimated to be approximately 339 ft beyond which DDE soils samples are assumed to be no longer spatially correlated.

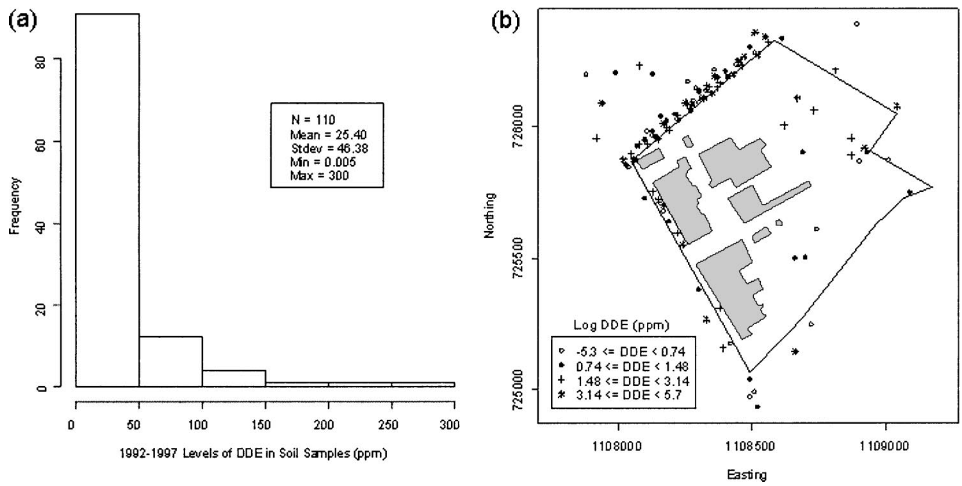

Fig. 2. Spatial and nonspatial exploratory analysis of the 110 DDE samples collected at the Superfund site, 1992–1997. (a) Histogram of the assayed DDE samples. (b) Spatial data posting based on quartiles for the log-transformed DDE samples.

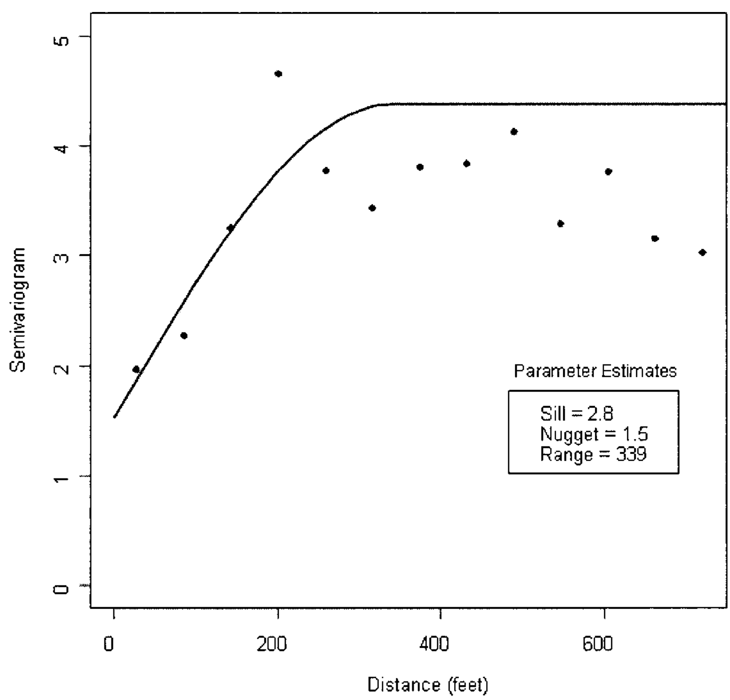

Fig. 3. Method of moments semivariogram estimates (points), composite likelihood fitted spherical semivariogram function (solid line), and corresponding parameter estimates for the log-transformed DDE samples.

The semivariogram estimates shown summarize spatial dependence for only a discrete set of distances. It is widely known that the methods of moments estimator is subjective and unreliable when based on a small number of spatial samples.[28,29] The 110 DDE samples is not a large sample size in this regard.[30] Therefore, the composite likelihood method was used for estimating the spherical semivariogram function because it is based on manipulations involving all 110 data samples, not just the summarized 13 estimates. For this reason, it is expected that the resulting estimated semivariogram function will not necessarily visually appear to fit these estimates well. Rather, the estimated semivariogram is used as an exploratory and diagnostic tool.

As an intermediate step in the analysis, the geostatistical process was validated using the 20 validation samples as described in the Methods section. The spatial locations of this validation set are shown in Fig. 1. Spatial dependence for the remaining 90 DDE values was estimated as outlined above and then used to predict at these validation locations. The predicted as well as measured DDE values are listed in Table I. The overall root mean squared error in prediction (RMSEP) at these 20 validation locations was approximately 23.5 ppm. For validation locations in residential areas, the RMSEP was 16.9 ppm versus a RMSEP of 22.3 ppm for within-site predictions.

Using all 110 DDE samples and the semivariogram function (Fig. 3), kriged predictions (BLUPs) were generated at a grid of locations placed over the entire site and surrounding areas. Grid spacings were approximately 30 ft apart and there were a total of 2500 prediction locations. The predictions were generated on the log scale and transformed back to the original scale.[20] The predicted values were imported into ArcGIS and smoothed using inverse distance weighted interpolation to produce the map of DDE contamination (Fig. 4).

The map indicates high levels of predicted DDE in the northwest corner of the site and in the southwest area just beyond the fence line. Relatively low levels of DDE are predicted in the surrounding northwest and northeast regions of the site (Fig. 4). This is encouraging as these are areas of residential development. However, although these areas seem to have the lowest predicted values on the map, the predicted DDE levels in these regions are up to 39.3 ppm DDE. The EPA recommends cleaning up industrial sites to a contamination level of less than 10 ppm DDE.[43] In addition, the EPA Region III risk based concentration (RBC) is 1.9 ppm for residential soil and 8.4 ppm for industrial soil based on standard exposure scenarios and a fixed level of risk.[44]

DISCUSSION

In this study, GIS and R, together were used to spatially analyze DDE contamination across a Superfund site in western Maryland. Via georeferencing, digitizing, and geocoding techniques, GIS provided the means to attain geographic coordinates of sample locations that otherwise were unknown, and required to spatially analyze the data. The statistical computing environment, R with the accompanying library, geoR, provide the necessary tools to carry out the geostatistical analysis that resulted in a map of predicted DDE levels across the site.

Table I. Measured Levels of DDE at the 20 Validation Locations, Their Predicted Values, Prediction Errors,[a] and Area Designation According to Fig. 1

Area	Measured value (ppm)	Predicted value (ppm)	Prediction error[a] (ppm)	Area	Measured value (ppm)	Predicted value (ppm)	Prediction error[a] (ppm)
Site	1.7	8.2	6.5	Industrial	15.0	5.1	−9.9
Site	2.7	19.3	16.6	Industrial	92.0	153.1	61.1
Site	2.5	15.2	12.7	Residential	1.4	6.2	4.8
Site	3.1	44.1	41.0	Residential	0.6	21.7	21.1
Site	13.0	32.9	19.9	Residential	1.2	36.6	35.4
Site	8.1	54.9	46.8	Residential	1.8	17.5	15.7
Site	67.0	69.1	2.1	Residential	2.3	10.8	8.6
Site	66.6	77.2	10.5	Residential	3.8	11.0	7.2
Site	25.0	29.5	4.5	Residential	19.0	28.4	9.4
Site	47.4	37.9	−9.5	Residential	17.0	27.4	10.4
Overall RMSEP (ppm)[b]				23.5			
RMSEP within Site (ppm)[b]				22.3			
RMSEP ouside of site in Industrial Property (ppm)[b]				43.8			
RMSEP ouside of site in Residential property (ppm)[b]				16.9			

[a]Prediction error = predicted value − measured value.
[b]Root mean square error in predictions (RMSEP) for the 20 locations overall and stratified by area.

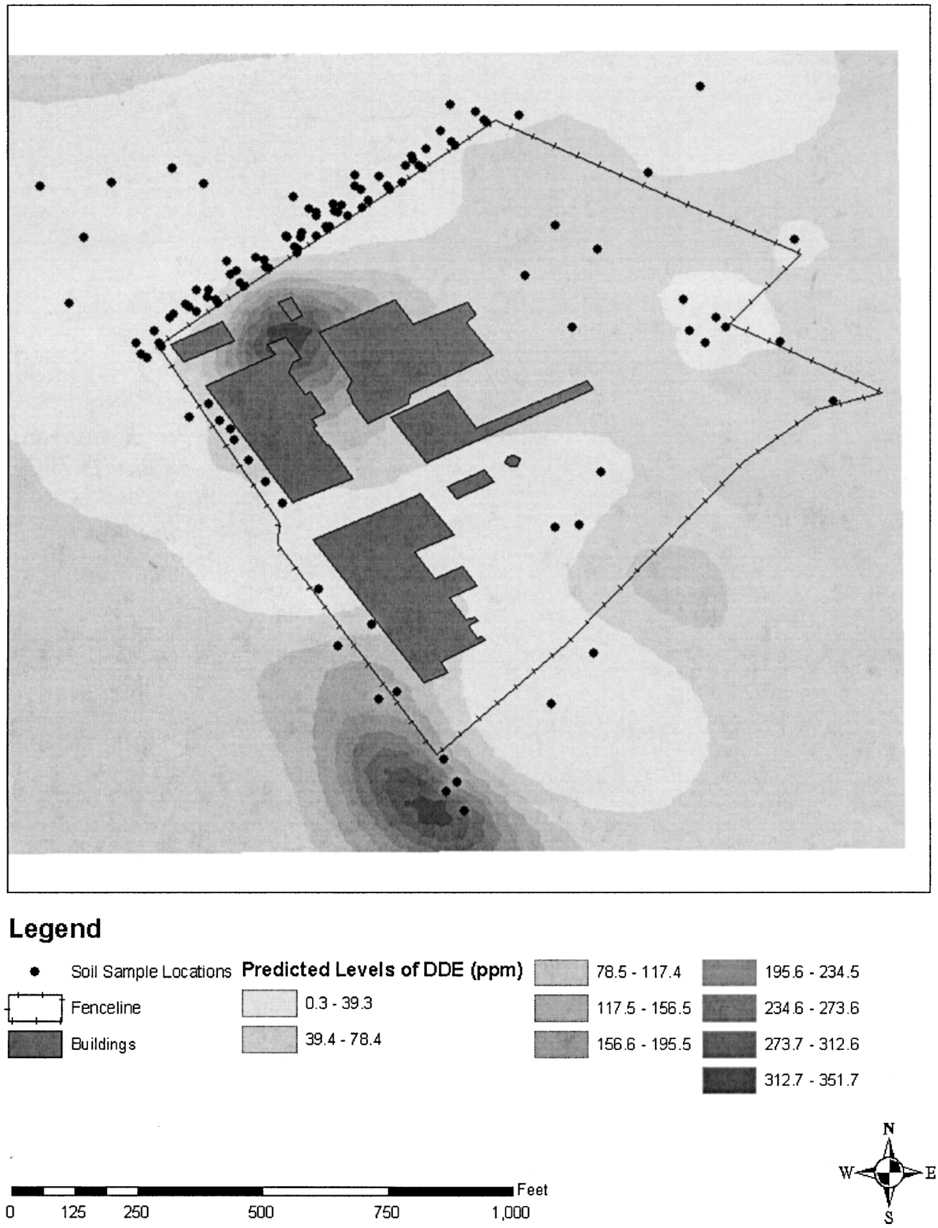

Fig. 4. Kriged map of predicted DDE levels for the western Maryland Superfund site based on soil data collected between 1992 and 1997.

The analysis in this study is for demonstration purposes, and does not reflect a complete analysis of contamination at the site. A more complete analysis would include, for example, DDT and DDD, as well as other known contaminants. In addition, a map of prediction uncertainty for the DDE analysis would be of value to complement predictions and identify areas in need of further sampling. Additional samples are currently being collected by the EPA on and nearby the site to further determine the extent of contamination and identify the best remediation strategy.

A problem encountered in the analysis was the large variability in the sample data, especially that noted for samples close together. The most likely cause of the large variability in contaminant levels in the soil is spillage. Since the site blended DDT in the past, it is likely that there were several points of spillage. In fact, several historical reports mention observed signs of spills.(11,13–15) The spillage of DDT could cause high levels of its derivative, DDE, to remain in these locations for long periods of time since DDE is not a mobile chemical in soil.(9) Another smaller source of variability could be due the fact that the samples were taken over a period of 5 years. It is possible, although unlikely, that environmental degradation over time could have caused the levels of DDE in the soil to change. Furthermore, each year the samples were collected, they were analyzed by different laboratories. This source of variability is most likely small compared to the values observed in the soil.

Both GIS and R have capabilities beyond what were used in this study. For example, GIS can serve as a relational database by providing a common format to link specific data to additional geographical or environmental data collected for other purposes in the same geographic area. Each of these databases can be mapped, and layered atop of one another in a process called overlaying.(37) For example, maps with property lines, proposed development, runoff patterns, and soil types could be overlayed with the DDE data used in this study. This is an avenue for further study.

Likewise, R has the capabilities of performing additional geostatistical techniques as demonstrated in the online geoR tutorial and related publications.(38,39,45,46) With the contributed library, geoRglm, all methods for model-based statistics as described in Diggle *et al*. are available.(27,47) In addition, there are over 20 contributed libraries currently available in R devoted to geostatistics and other methods of spatial statistics. Together, R and GIS make optimal tools for data analysis in the medical and public health fields.

CONCLUSION

This study presents an evaluation of environmental contamination data that were analyzed using GIS and R to predict levels of contamination on and around a Superfund site in western Maryland. The methods and results of this study are important in the public health field because there are over 1200 Superfund sites across the country that are contaminated with substances that adversely affect human health.(17) Furthermore, Superfund sites are often in urban areas surrounded by residences. Therefore, the health of people living near sites like these is dependant on the ability to accurately characterize and evaluate environmental contamination on and around sites like these. Spatial analyses, such as performed in this study using GIS and R, help to improve such characterization, thereby providing important methods in protecting the public's health.

ACKNOWLEDGMENTS

This study was supported by the National Institute for Environmental Health Sciences (NIEHS) training grant (ES 07141) and the Johns Hopkins NIEHS Center in Urban Environmental Health (P30 ES 03819). In addition, the authors thank Humberto Monsalvo of the EPA for his helpful cooperation in obtaining data documents and guidance along site visits.

REFERENCES

1. Lang, L., *GIS for Health Organizations*, 1st edn., Environmental Systems Research Institute, Redlands, CA, 2000.
2. Fitzgerald, E. F., Brix, K. A., Deres, D. A., Hwang, S. A., Bush, B., Lambert, G., and Tarbell, A., Polychlorinated biphenyl (PCB) and dichlorodiphenyl dichloroethylene (DDE) exposure among Native American men from contaminated Great Lakes fish and wildlife. *Toxicol. Ind. Health* 12:361–368, 1996.
3. Shone, S. M., Ferrao, P. N., Lesser, C. R., Norris, D. E., and Glass, G. E., Analysis of mosquito vector species abundances in Maryland using geographic information systems. *Ann. N. Y. Acad. Sci.* 951:364–368, 2001.
4. McKee, K. T., Jr., Shield, T. M., Jenkins, P. R., Zenilman, J. M., and Glass, G. E., Application of a geographic information system to the tracking and control of an outbreak of shigellosis. *Clin. Infect. Dis.* 31:728–733, 2000.
5. Latkin, C., Glass, G. E., and Duncan, T., Using geographic information systems to assess spatial patterns of drug use, selection bias and attrition among a sample of injection drug users. *Drug Alcohol Depend.* 50:167–175, 1998.
6. Glass, G. E., Amerasinghe, F. P., Morgan, J. M., III, and Scott, T. W., Predicting Ixodes scapularis abundance on white-tailed deer using geographic information systems. *Am. J. Trop. Med. Hyg.* 51:538–544, 1994.
7. Glass, G. E., Schwartz, B. S., Morgan, J. M., III, Johnson, D. T., Noy, P. M., and Israel, E., Environmental risk factors for Lyme disease identified with geographic information systems. *Am. J. Public Health* 85:944–948, 1995.
8. Becker, K. M., Glass, G. E., Brathwaite, W., and Zenilman, J. M., Geographic epidemiology of gonorrhea in Baltimore, Maryland, using a geographic information system. *Am. J. Epidemiol.* 147:709–716, 1998.
9. Agency for Toxic Substances and Disease Registry, *ToxFAQs for DDT, DDE, and DDD*. Retrieved September 1, 1995, from http://www.atsdr.cdc.gov/tfacts35.html
10. Longnecker, M. P., Rogan, W. J., and Lucier, G., The human health effects of DDT (dichlorodiphenyltrichloroethane) and PCBS (polychlorinated biphenyls) and an overview of organochlorines in public health. *Annu. Rev. Public Health* 18:211–244, 1997.
11. Maryland Department of the Environment, *Sampling Proposal for Expanded Site Inspection of the Central Chemical-Hagerstown Site, MD-302*, March 1, 1993.
12. Roy F. Weston, Inc., *Site Assessment Technical Assistance Team*, Trip report SAT0300595, 1997.
13. Maryland Department of the Environment, *Expanded Site Inspection of the Central Chemical-Hagerstown Site, MD 302.1*, June 1, 1994.
14. Maryland Department of the Environment, *Expanded Site Inspection Sampling Plan of the Central Chemical-Hagerstown Site, MD 302*, July 1, 1995.
15. Maryland Department of the Environment, *Expanded Site Inspection of the Central Chemical-Hagerstown Site, MD 302.2*, January 1, 1996.
16. Booz Allen Hamilton, Inc., *RCRA, Superfund, and EPCRA Hotline Training Module*, EPA540-R-98-029, 1998.
17. United States Environmental Protection Agency, *Final National Priorities List Sites*. Retrieved October 21, 2003, from http://www.epa.gov/superfund/sites/query/queryhtm/nplfin1.htm
18. Isaaks, E. H., and Srivastava, R. M., *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 1989.
19. Cressie, N., Geostatistics. *Am. Stat.* 43:197–202, 1989.
20. Cressie, N., *Statistics for Spatial Data*, Wiley, New York, 1991.

21. Hwang, S. A., Fitzgerald, E. F., Cayo, M., Yang, B. Z., Tarbell, A., and Jacobs, A., Assessing environmental exposure to PCBs among Mohawks at Akwesasne through the use of geostatistical methods. *Environ. Res.* 80:S189–S199, 1999.
22. Preat, B., Application of geostatistical methods for estimation of the dispersion variance of occupational exposures. *Am. Ind. Hyg. Assoc. J.* 48:877–884, 1987.
23. Juang, K. W., Lee, D. Y., and Ellsworth, T. R., Using rank-order geostatistics for spatial interpolation of highly skewed data in a heavy-metal contaminated site. *J. Environ. Qual.* 30:894–903, 2001.
24. Saito, H., and Goovaerts, P., Selective remediation of contaminated sites using a two-level multiphase strategy and geostatistics. *Environ. Sci. Technol.* 37:1912–1918, 2003.
25. Thayer, W. C., Griffith, D. A., Goodrum, P. E., Diamond, G. L., and Hassett, J. M., Application of geostatistics to risk assessment. *Risk Anal.* 23:945–960, 2003.
26. Cressie, N., Fitting variogram models by weighted least squares. *J. Int. Assoc. Math. Geol.* 17:563–586, 1985.
27. Diggle, P. J., Tawn, J. A., and Moyeed, R. A., Model-based geostatistics. *Appl. Stat.* 47:299–326, 1998.
28. Curriero, F. C., and Lele, S., A composite Likelihood approach to semivariogram estimation. *J. Agric. Biol. Environ. Stat.* 4:9–28, 1999.
29. Myers, D. E., Begovich, C. L., Butz, T. R., and Kane, V. E., Variogram Models for regional groundwater geochemical data. *Math. Geol.* 14:629–644, 1982.
30. Webster, R., and Oliver, M. A., Sample adequately to estimate variograms of soil properties. *J. Soil Sci.* 43:177–192, 1992.
31. Johnston, K., Ver Hoef, J. M., Krivoruchko, K., and Lucas, N., *Using ArcGIS Geostatistical Analyst*, Environmental Systems Research Institute, Redlands, CA, 2001.
32. Ihaka, R., and Gentleman, R. R., A language for data analysis and graphics. *J. Comput. Graph. Stat.* 5:299–314, 1996.
33. Deutsch, C. V., and Journel, A. G., *GSLIB Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 1992.
34. Mathsoft, *S+SPATIALSTATS User's Manual*, Version 1.0 edn., Mathsoft, Inc., Seatle, 1996.
35. R Development Core Team, *The R Project for Statistical Computing*. Retrieved October 24, 2003, from http://www.r-project.org/
36. Holoubek, I., Kocan, A., Holoubkova, I., Hilscherova, K., Kohoutek, J., Falandysz, J., and Roots, O., *Persistent, Bioaccumulative and Toxic Chemicals in Central and Eastern European Countries*, State-of-the-art Report, TOCOEN report no. 150a, RECETOX-TOCOEN & Associates, Brno, Czech Republic, 2000.
37. Ormsby, T., Napoleon, E., Burke, R., Groessl, C., and Feaster, L., *Getting to Know ArcGIS Desktop*, Environmental Systems Research Institute, Redlands, CA, 2001.
38. Ribeiro, P. J., Jr., and Diggle, P. J., geoR: A package for geostatistical analysis. *R-NEWS*, 1, 2, 2001. Available from http://cran.r-project.org/doc/Rnews
39. Ribeiro, P. J., Jr., and Diggle, P. J., *geoR: A Package for Geostatistical Data Analysis Using the R Software*. Available September 30, 2003, from http://www.est.ufpr.br/geoR/
40. Haas, T. C., Lognormal and moving window methods of estimating acid deposition. *J. Am. Stat. Assoc.* 85:950–963, 1990.
41. Armstrong, M., and Boufassa, A., Comparing the robustness of ordinary kriging and lognormal kriging: Outlier resistance. *Math. Geol.* 20:447–457, 1988.
42. Dowd, P. A., Lognormal kriging—The general case. *J. Int. Assoc. Math. Geol.* 14:475–499, 1982.
43. United States Environmental Protection Agency, Region 9, *Soil Screening Users Guide Technical Document*, EPA/540/R-95/128, Appendix A, July 1, 1996.
44. United States Environmental Protection Agency, Region 3. *Risk-Based Concentration Table*, October, 15, 2003.
45. Christensen, O. F., Diggle, P. J., and Ribeiro, P. J., Jr., Analysing positive-valued spatial data: The transformed Gaussian model. In Monestiez, P., Allard, D., and Froidevaux, R. (eds.), *GeoENVIII—Geostatistics for Environmental Applications*, vol. 11, Kluwer, New York, pp. 287–298, 2001.
46. Diggle, P. J., and Ribeiro, P. J., Jr., Bayesian inference in Gaussian model-based geostatistics. *Geogr. Environ. Model.* 2001.
47. Christensen, O. F., and Ribeiro, P. J., Jr., *geoRglm: Software for generalised linear spatial models using R* (0.7-5), September 3, 2003.