Developments in Environmental Science, Volume 5
D. Sarkar, R. Datta and R. Hannigan (Editors)

ISSN: 1474-8177/DOI:10.1016/S1474-8177(07)05032-2

Chapter 32

Surficial characterization of dioxin in Midland, Michigan, using non-euclidean geostatistics

Patrick Kinnicutt

Abstract

Much research has been done characterizing surficial contaminants using geostatistics or other spatial estimation technique. This chapter examines the use of a non-Euclidean distance metric combined with geostatistical techniques to model the surficial distribution of dioxins in Imerman Park near Midland, Michigan. This chapter also examines the applicability of geostatistics to small data sets.

An overview of the dioxin sampling in Midland, MI, will be examined, followed by a brief overview of geostatistical theory, variogram modeling, and the use of non-Euclidean distance metrics to capture the geologic processes. Preliminary results of a case study evaluating the surficial dioxin distribution in Imerman Park downstream from the Dow Midland plants will then be presented, comparing the use of a flood plain non-Euclidean distance norm versus a Euclidean distance norm.

32.1. Background

Dow Chemical Corporation has emitted dioxins into the Tittabawassee River from their Midland, Michigan, plants since 1897 (Dow, 2005). Figure 32.1 shows the location of Midland, Michigan, in relation to the state of Michigan. Midland is located approximately 160 km north of Lansing, Michigan's state capitol. As shown in Fig. 32.2, the Tittabawassee River flows through downtown Midland, with the Dow Chemical plants located adjacent to the river, downstream from where the Tittabawassee and Chippewa Rivers join. The Michigan Department of Environmental Quality (MDEQ) performed some initial tests over the past five years along the Tittabawassee River, finding higher than normal dioxin levels in the Tittabawassee River flood plain, downstream from the Dow Chemical plants. Upstream dioxin levels in the Pine, Chippewa, and

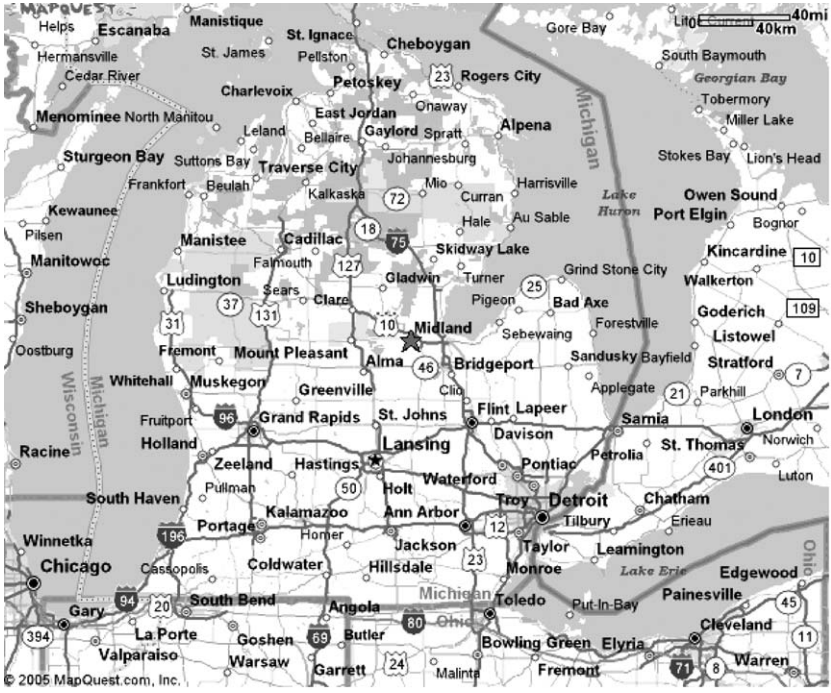

Figure 32.1. Figure showing the location of Midland, Michigan (represented as a star) in relation to the state of Michigan [from Mapquest].

Tittabawassee Rivers had background levels of dioxin (MDEQ, 2003). These elevated dioxin levels were found downstream of the Dow plants, all the way to where the Tittabawassee River joins the Saginaw River and further until the Saginaw River empties into the Saginaw Bay. Further downstream, other plants exist that are dioxin sources. No other plants or point sources of dioxin exist immediately downstream of the Dow plants in Midland, however, so the dioxin found in the Tittabawassee River between Midland and Freeland is all believed to have originated from the Dow plants.

32.1.1. Dioxin overview

The term dioxin is a general term used to refer to approximately 210 chemicals with similar chemical properties (MDEQ, 2003). These chemicals are generally formed as a by-product of burning chlorine-based chemical products with hydrocarbons. Several sources of dioxin include

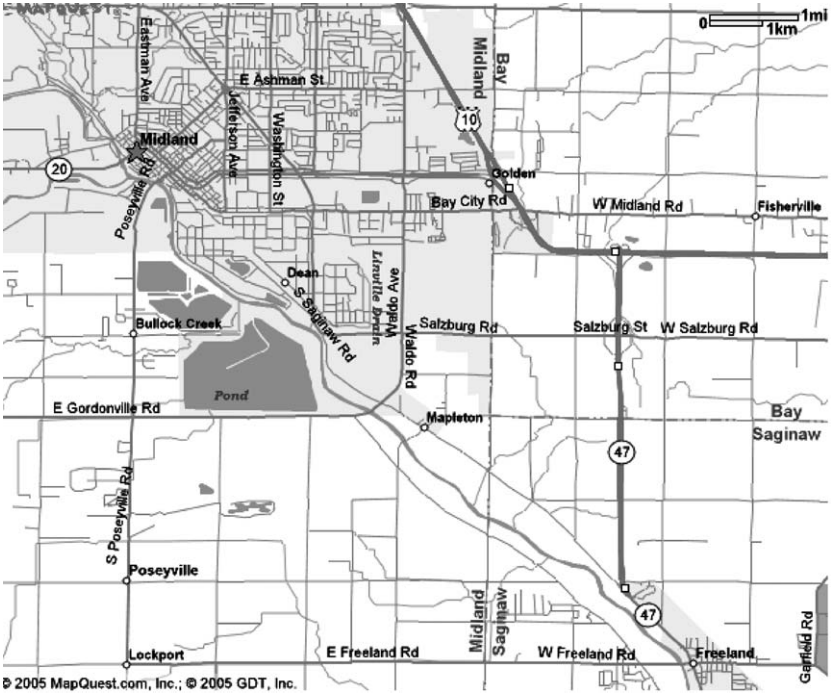

Figure 32.2. Figure showing a street map of downtown Midland, Michigan, with Dow Chemical's location (represented as a star) [from Mapquest].

burning activities from incinerators, pesticides, herbicides, pulp mills, and PVC (polyvinyl chloride). These chemicals may be found pervasively in air, water, soil, and sediments. Agent Orange is one example of an herbicide used by the military during the Vietnam War that led to many health problems in veterans.

Health effects from overexposure to dioxins are an issue of current debate and controversy between Dow and the MDEQ. One universally accepted health effect is the development of chloracne (MDEQ, 2003). Chloracne is a skin condition that may look similar to teenage acne, and is caused by exposure to dioxins. Other possible health effects have been linked to overexposure to dioxins, although these effects have not been universally accepted. These effects include a lowered immune response, hormonal changes, developmental problems in children, reproductive problems in pregnant mothers, and cancer.

Exposure to dioxins is most effectively accomplished by direct ingestion, through eating various foods in ones diet (Clark, 2004). Beef is a

common source of dioxin in the diet, as well as dairy, chicken, pork, fish, and eggs. Vegetables can also be a source of dioxins, although to a lesser extent than meats. Dioxin exposure can also be obtained through direct contact with soil or sediments contaminated with dioxins.

The 210 chemicals that are classified as dioxins all have differing toxicity levels; some chemicals are more toxic than others (Clark, 2004). The most toxic form of dioxin is 2,3,7,8-tetrachlorodibenzo-para-dioxin (2,3,7,8-TCDD). Since all dioxins are usually characterized as a group in environmental remediation studies, a toxic equivalency (TEQ) was developed converting concentrations of other dioxin chemicals to the equivalent concentration of 2,3,7,8-TCDD, with this number representing the concentration of all dioxins at a site.

The EPA has an established human exposure limit of dioxins set at 1000 parts per trillion (ppt) for mandatory cleanup. The MDEQ calculated a different exposure limit of 90 ppt as the threshold for cleanup (MDEQ, 2003). The MDEQ calculated this threshold based on an increased cancer risk of 1 additional cancer per 100,000 individuals above the normal risk. This threshold discrepancy has been a major point of disagreement between Dow and the MDEQ in their negotiations for cleanup in the Tittabawassee River flood plain. Recently, the MDEQ agreed to use the EPA standard 1000 ppt residential contact criteria as the threshold for Dow to mitigate the area, due to complaints from Dow and from the Midland residents (Midland, 2005).

32.1.2. Michigan DEQ dioxin testing summary

The MDEQ tested the sediments under the Tittabawassee River for dioxins and other contaminants (MDEQ, 2002). Dredging the river at various locations within each transect was performed; these samples were then combined to create composite samples by mixing the various dredge samples. A representative sample was then tested for dioxin within each transect's composite sample. Several shallow cores were also taken adjacent to the river. Figure 32.3 shows a part of the Tittabawassee flood plain and a summary of results. Nine transects were sampled over a 22 mile river distance. Each transect is represented in Fig. 32.3 as a rectangle, and each transect has a value next to the rectangle displaying the TEQ for dioxins in ppt. The transect dioxin levels downstream from Dow showed consistent readings well above the 90 ppt threshold set by the MDEQ, but less than the 1000 ppt threshold mandated by the EPA. However, several of the shallow cores did show TEQ concentrations above 1000 ppt adjacent to the river.

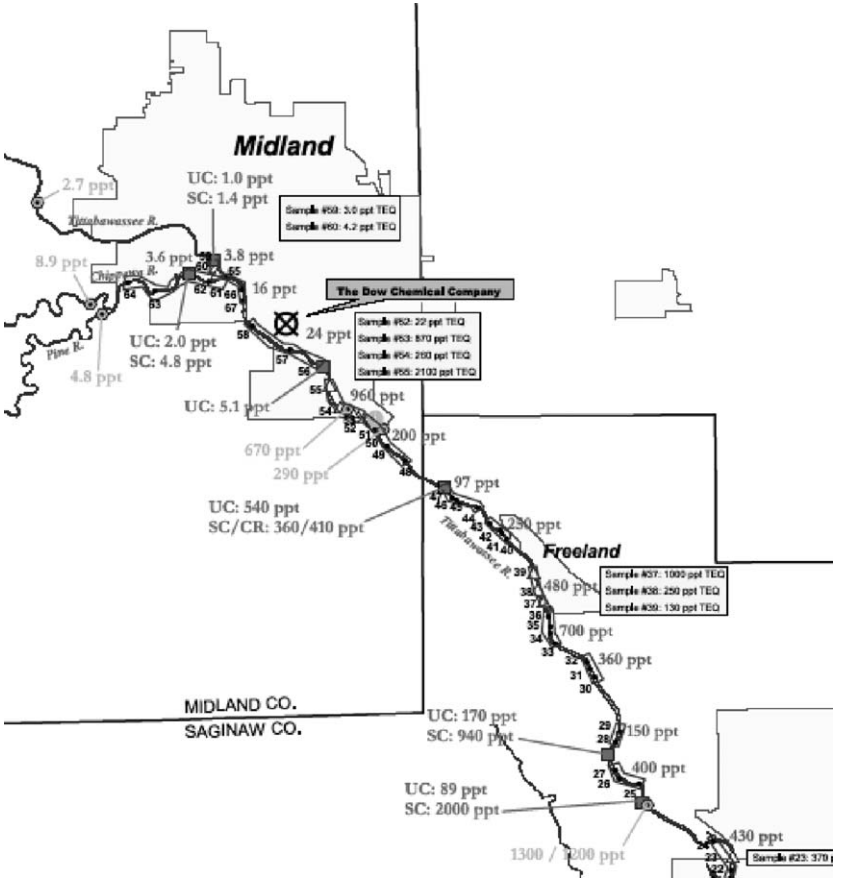

Figure 32.3. Tittabawassee River dioxin sample results between Midland and Saginaw, Michigan (MDEQ, 2002).

During phase 2 of testing, the MDEQ took over 200 shallow soil samples within the Tittabawassee River flood plain (MDEQ, 2003). These samples were taken over a 22 mile length at various locations. The sample depths for the shallow cores were taken at 1″, 3″, 6″, and 15″ below ground level (bgl). These samples also showed many locations where the TEQ was greater than the 90 ppt MDEQ threshold. Several samples were also found to be greater then the 1000 ppt EPA threshold. Dow and the MDEQ have been negotiating the extent of mitigation; recently, the MDEQ agreed to mitigation at the 1000 ppt level (Midland, 2005).

During phase 2, three deep soil borings were taken using a hydraulic probe sampler. Samples were taken at these three borings 24″, 36″, 48″,

and 60″ below ground level. Of these samples, several of the samples at the 36″–48″ depth were found to be above the MDEQ threshold. All the samples tested from the 60″ depth tested at or near the background levels. Figure 32.4 shows the dioxin background levels for various locations in Michigan (MDEQ, 1999). As shown in Fig. 32.4, the background dioxin concentrations in Michigan's Lower Peninsula average well less than 10 ppt, with the area near Saginaw measuring around 13 ppt.

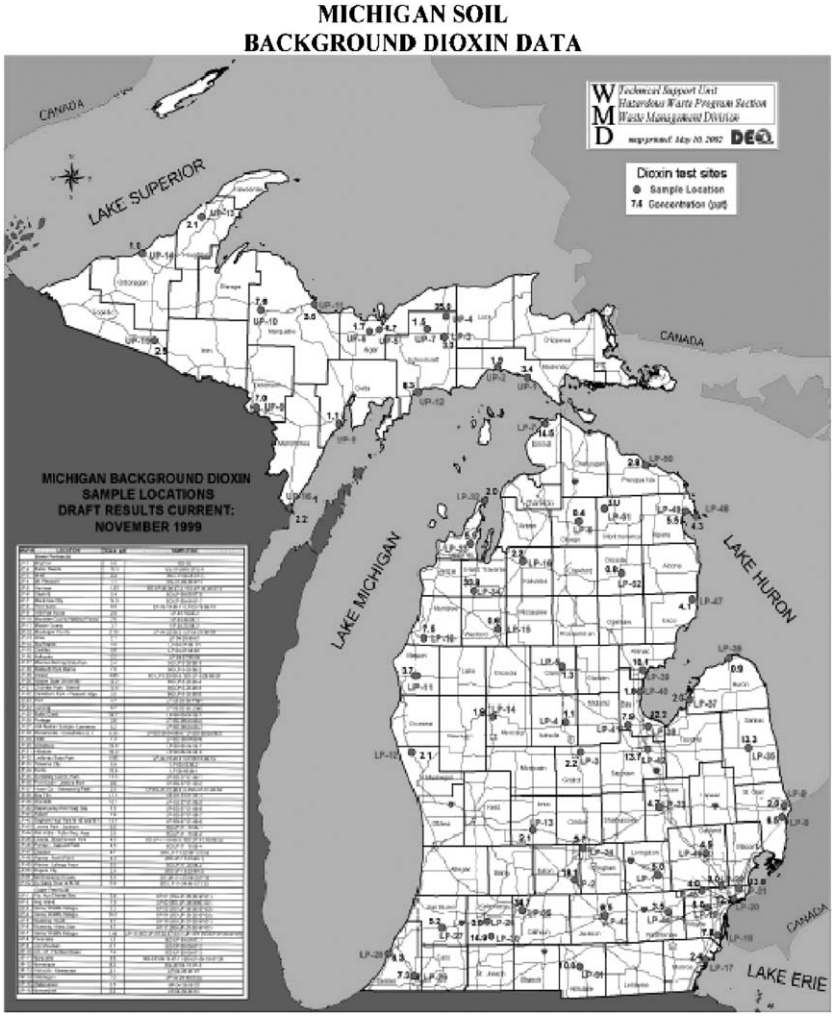

Figure 32.4. Background dioxin levels in Michigan (MDEQ, 1999).

The MDEQ lastly tested 22 drinking wells for dioxins (MDEQ, 2003). In these tests, none of the drinking well samples tested above the 33 ppt level set by the EPA. Current studies are focusing on bioaccumulation in flora and fauna in the flood plain, and blood monitoring of Midland residents that live in the flood plain is being discussed.

32.1.3. Geostatistical characterization of surficial contaminants

Characterization of environmental contaminants is usually done by taking samples at various spatial locations at a particular site and then using spatial interpolation methods to infer the spatial distribution of contaminants. Many different techniques have been used to do this site characterization; several will be described in the following paragraphs. For many contaminants, these spatial interpolation methods extend to three dimensions (x, y, and z) since the contaminant plume can migrate into the groundwater and into aquifers. Much of this discussion will focus on surficial characterization, as studies have shown that dioxins will normally stick to the soil near the surface (MDEQ, 2003). Also, no data samples exist from the MDEQ tests at sufficient depths to do any 3D modeling; the case study will focus on 2D characterization of surficial contaminants.

When looking at natural phenomena in earth science data sets such as with contaminant characterization, the sample data does not exhibit statistical independence; rather, a spatial dependence in the random variable(s) of interest exists (Isaaks and Srivastava, 1989). This spatial dependence is known as spatial autocorrelation. A positive spatial autocorrelation tells us that as two data locations get closer together, their random variable values will be more similar than two data locations further apart. Figure 32.5 shows an example of this correlation. As the separation distance $|\mathbf{u} - \mathbf{v}|$ increases, the covariance $c_{ij}(|\mathbf{u} - \mathbf{v}|)$ decreases

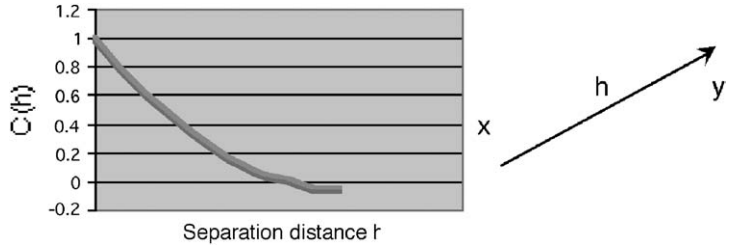

Figure 32.5. As the separation distance between two data points increases, the covariance decreases.

(Deutsch and Journel, 1998), until a certain range is reached. Once two data locations become further apart than this range, the data values no longer exhibit any correlation so become independent. The sample covariance can be calculated from Eq. (1), where **u** and **v** are two data location vectors, $|\mathbf{u}-\mathbf{v}|$ is the separation distance vector, $c_{ij}^{*}(|\mathbf{x}-\mathbf{y}|)$ is the covariance as a function of separation distance, $Z_i(\mathbf{u})$ and $Z_i(\mathbf{v})$ are the data values at the tail and head of the data pair, respectively, and $\bar{Z}_i$ and $\bar{Z}_j$ are the means at the tail and head respectively.

$$c_{ij}^{*}(|\mathbf{u}-\mathbf{v}|) = \left\langle \left(Z_i(\mathbf{u}) - \bar{Z}_i\right) \cdot \left(Z_j(\mathbf{v}) - \bar{Z}_j\right) \right\rangle \tag{1}$$

The vector **h**, as shown in Eq. (2), usually denotes the separation distance between two data locations.

$$\mathbf{h} = |\mathbf{u}-\mathbf{v}| \tag{2}$$

Geostatistics, as defined by Matheron (Journel and Huijbregts, 1978), is the application of the formalism of random functions to the estimation of natural phenomena. Geostatistics studies natural phenomena that change over space and/or time (Deutsch and Journel, 1998). In geostatistics, spatial variability is commonly used instead of spatial correlation. With spatial variability, as two data points become further apart, the spatial variability increases, until the distance becomes large enough where the two data values are independent. The sample (semi-)variogram provides a measure of the spatial variability, and is computed in Eq. (3),

$$\gamma^{*}(\mathbf{h}) = \tfrac{1}{2}E\left[\left\{Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})\right\}^2\right] \tag{3}$$

In Eq. (3), $\gamma^{*}(\mathbf{h})$ is the experimental or sample variogram measure for all data pairs within the separation distance **h**, while $Z(\mathbf{u})$ and $Z(\mathbf{u}+\mathbf{h})$ are the data values at two locations separated by the separation distance **h**. The sample variogram is one half of the second-order moment between these data samples. One advantage of this variogram measure over the covariance measure shown in Eq. (1) is that the means of the corresponding head and tail values are not required. Other spatial measures (Deutsch and Journel, 1998) have been defined to capture the spatial variability in sampled data.

Other than looking at the spatial description of the data samples via the covariance and/or variogram measures, non-spatial statistical measures such as the mean, median, interquartile range, range, variance, skew, and kurtosis are also typically computed in order to obtain a good understanding of the data. In the case of multivariate data, measures such as covariance and correlation coefficient are also computed. Graphs such as

histograms and scattergrams provide excellent visual tools for analyzing univariate and multivariate data.

Once the data analysis has been completed examining the spatial and non-spatial sample statistics, spatial estimation or simulation techniques can be applied to create a map or volume of predicted values (Isaaks and Srivastava, 1989). In the case of surficial contaminant characterization, the relevant predicted values would be either a map of contaminant concentrations or a map of probabilities, showing the probability of the concentration exceeding the desired threshold. All these spatial estimation techniques share a fundamental equation form, as shown in Eq. (4) (Isaaks and Srivastava, 1989),

$$Z^*(\mathbf{u}) = \sum_{i=1}^{N} w_i \cdot Z(\mathbf{u}_i) \tag{4}$$

In Eq. (4), $Z^*(\mathbf{u})$ represents the estimate of the random variable at a particular location, N is the sample size, and w_i are weights assigned to each data value $Z(\mathbf{u}_i)$. Spatial estimation methods differ basically in how the weights assigned to each data value are determined. For instance, with the inverse distance squared method, each weight w_i is a function of $1/d_i^2$, where d_i is the separation distance between the known data sample location and the estimation location.

Kriging methods are a set of spatial interpolation methods that provide a variety of ways to compute the weights, w_i, from Eq. (4) (Isaaks and Srivastava, 1989). These methods are often referred to as BLUE methods; BLUE is an acronym for ''best linear unbiased estimator'' (Isaaks and Srivastava, 1989). Kriging estimates are linear because their general form follows Eq. (4), which is a linear weighted combination of data. They are unbiased since they attempt to set the mean residual or error to 0, and they are best because they try to minimize the variance of the errors. Kriging methods rely on the use of a variogram model to model the spatial variability and compute weights that are ''best'' and ''unbiased''.

Several different variogram models exist in the literature; Deutsch describes five models, although many more exist (Deutsch and Journel, 1998). A variogram model provides a model of variability, describing how the data varies as a function of both the distance and direction. The variogram model replaces the Euclidean distance as used in an algorithm like inverse distance with a structural distance; the variogram model is a two-point statistic, since it only considers two locations at a time. The variogram model usually takes the Euclidean distance between one data value and the estimation location, converting this distance into a variogram value, which is then used to solve for the weights. A typical

variogram model used in surficial contaminant characterization can capture geometric and zonal anisotropies (Deutsch and Journel, 1998). Equation (5) shows an example of an exponential variogram model (Deutsch and Journel, 1998).

$$\gamma(\mathbf{h}) = c \cdot \left[1 - \exp\left(-\tfrac{3\mathbf{h}}{a}\right)\right] \tag{5}$$

In Eq. (5), $\gamma(\mathbf{h})$ is the (semi-)variogram model, c is the sill or maximum variogram value, $\mathbf{h}$ is the separation distance between a data point and the estimation point, and α is the effective range. A variogram ellipse, with a major and minor axis as well as an azimuth direction, models the change in the range a as a function of the direction.

The variogram model does not have to use the Euclidean distance. A non-Euclidean distance norm can be developed to capture objective and/or subjective spatial and/or temporal information about the processes that caused the contaminant to be distributed as it is. In this approach, the spatial coordinates are transformed into a ''flattened'' geographical space, then classical geostatistics is performed in the transformed space; after modeling in the ''flattened'' space, the results are resampled back into the original coordinate space. For example, Fig. 32.6 shows an example of a contaminant characterization model, where the contaminant was deposited via stream processes. In this figure, the use of a Euclidean distance metric may actually lead to the two ovals being too highly correlated when predicting, since the Euclidean distance is less than the non-Euclidean distance. This non-Euclidean distance, referred to as the water distance (Rathbun, 1998), takes account of the contaminant's depositional process. Figure 32.6-B displays the same geologic area as Fig. 32.6-A, although this figure shows the flattened logical grid skeleton generated by using the water distance to model the contaminant distribution. This approach was used by Barbaras et al. (2001) to model uncertainty

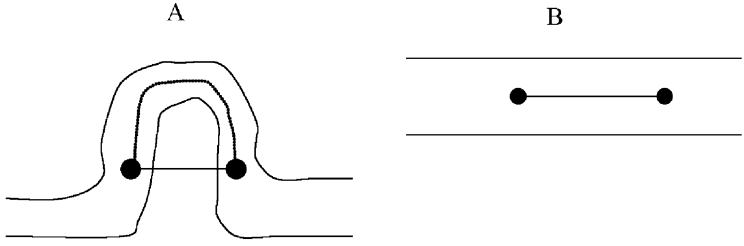

Figure 32.6. (A) Two points appear closer in distance using a Euclidean distance metric than when using a non-Euclidean metric that accounts for the oxbow in a meandering stream; (B) view of the non-Euclidean metric after flattening the model, removing the fluvial meandering process.

for three-dimensional dioxin data in the Passaic River sediments; in Barbaras's study, they used a software package called *Gridgen* to transform the coordinates, using a water distance metric, since their study examined the dioxin distribution in the river sediments.

The selection of a non-Euclidean distance norm must be based on both objective and subjective knowledge regarding the processes that are known to have contributed to the contamination. In the case study presented later, the dominating process leading to the surficial dioxin contamination involved flooding events, where the flood caused contaminated sediments to wash ashore in the flood plain.

The choice of norm, whether Euclidean or non-Euclidean, should not be done a priori at the expense of excluding alternative measures of distance. These alternative norms may provide yet unrealized insights into the nature of the random process. Departures from the Euclidean distance have been suggested based on the geology of the domain of interest, for example non-convex sampling regions. Rathbun (1998) and Rendu and Readdy (1992) discuss such situations. The great-arc distance has also found some use in geostatistics, for example Cressie et al. (1990). The goal is to model spatial variability as well as possible and the model that accomplishes this may not have to correspond perfectly to reality (Curriero, 1996). Much research and development has been done to incorporate non-Euclidean distances in commercial GIS software, in particular ESRI's ArcGIS (Krivoruchko and Gotway, 2003). One method Krivoruchko describes for determining the non-Euclidean metric involves the use of cost-weighted distance to produce a cost surface.

Several flavors of kriging have been developed over the years to account for the use of subjective and objective data, and to address limitations of the variogram model approach (Deutsch and Journel, 1998). The use of a non-Euclidean distance norm as input to the variogram model provides a mechanism to take geologic processes into account in the structural model, as in Fig. 32.5. In Fig. 32.5, the contaminant in most likelihood flow through the river as a dissolved load or suspended load attached to sediments, finally getting deposited downstream in the river. Thus, at any one time it makes sense that the distance norm should follow the water distance (Rathbun, 1998). It also makes sense that when looking at contaminant concentration values on the surface transported by flood processes, the concentrations in the river and in the flood plain should be segregated and modeled separately. In the river, the contaminant is constantly being subjected to fluvial processes, causing temporal variations in the contaminant concentrations. However, in the flood plain the contaminants (in the case of dioxins) will cling to surficial soil and not be transported, until the next flood event occurs. In this scenario,

a non-Euclidean distance norm based on flood plain maps and DEM maps can be used to model the variability in contaminants. If one is examining a contaminant that flows readily in the groundwater, the hydrogeology can be taken into account if information is available to obtain better models of spatial variability.

Many different spatial interpolation algorithms may be used to solve a particular application. Saito and Goovaerts (2000) evaluated several algorithms to characterize dioxins in a well-sampled (over 600 data samples) EPA Superfund site at Piazza Road in Missouri, examining the use of ordinary kriging, log-normal kriging, multi-Gaussian kriging and indicator kriging. During this study, they found that log-normal kriging consistently gave the best results in terms of smallest prediction errors and false positives, with indicator kriging producing good results. Saito and Goovaerts (2000) later examined the use of indicator kriging and p-field simulations at the same site. In this study, indicator kriging was used to provide a probability distribution that the dioxin levels did not exceed the threshold, then p-field simulations were done to see if the average pollutant concentrations in a remediation unit exceed the regulatory threshold. Saito and Goovaerts used Euclidean distance norms in both studies.

32.2. Case study: Midland, Michigan dioxin characterization

This case study shows an application of non-Euclidean distance norms in the estimation of dioxin concentrations in Imerman Park. Many studies concentrate on the use of the water metric (Saito and Goovaerts, 2002; Rendu, 1998) to model contaminant concentrations in streams. This study examines the use of a flood plain distance metric applied on land, with non-ideal data. The data in this study is similar to that found in many sites; there are few data samples and the data samples are clustered in various locations, with large areas left unsampled. Inverse distance and sequential gaussian simulation algorithms were chosen for comparison with this study. GSLIB programs (Deutsch and Journel, 1998) were used for data analysis, simulation, and presentation of results. The inverse distance squared algorithm used a custom program designed by the author, and the coordinate transformation from and to Euclidean space from the non-Euclidean norm was done as pre-processing and post-processing steps to using sgsim and gamv (Deutsch and Journel, 1998), using a coordinate conversion program written by the author.

The MDEQ collected shallow samples at various locations along the Tittabawassee River flood plain, as described earlier (MDEQ, 2003).

One site was examined in a pilot study to see what effect the use of a non-Euclidean distance metric as opposed to a Euclidean distance metric would have on the interpreted results. The site chosen was Imerman Park, located approximately 11½ miles downstream from the Dow plants. This site was chosen because it had the largest number of shallow data sample locations, 14. Imerman Park is a site where Dow Chemical has an established interim response activities (IRA) plan in place with the MDEQ. Figure 32.7 shows an aerial photograph of Imerman Park, adjacent to the Tittabawassee River in Saginaw Township, Michigan.

Figure 32.7. Areal photograph of Imerman Park in Saginaw Township, Michigan.

The blue lines show the extent of the 100-year flood plain, and the stars denote the sample locations for the 14 locations tested in the park. Imerman Park encompasses approximately 96 acres, and it includes public playgrounds, a baseball field, boat launch areas, picnic sites, hiking trails and pet exercise areas. The park entrance is shown in the upper right corner of Fig. 32.7. As shown by the blue lines, most of the park is located within the 100-year flood plain.

Figure 32.8 displays a contour map of the flood plain in Imerman Park that was used in the study. This contour map shows approximate boundaries for flooding events, using a 20-year contour interval. The author converted these contours into a "flattened" coordinate space, using a preprocessing program written by the author to transform the coordinates.

Figure 32.9 displays the cumulative frequency plot of surficial dioxin concentrations measured in ppt from the 14 sample locations. As shown in Fig. 32.9, the minimum measured dioxin concentration was 5.2 ppt, with a maximum data sample of 1400 ppt and a median of 580 ppt. From the interquartile range, it is shown that a little over 25% of the samples tested less than the MDEQ limit of 99 ppt, while a little over 25% of the samples tested greater than the EPA threshold of 1000 ppt.

Figure 32.10 displays the same 14 data samples as shown in Fig. 32.7, color-coded to display the dioxin concentrations. Figure 32.10 shows that the high concentration dioxin values are clustered in the southwest section of the park, near the boat launch and the Tittabawassee River.

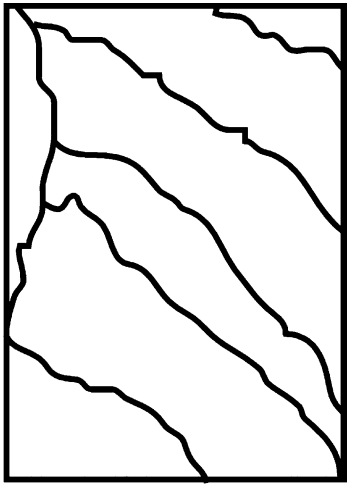

Figure 32.8. Figure showing flood plain contours at a 20 year interval.

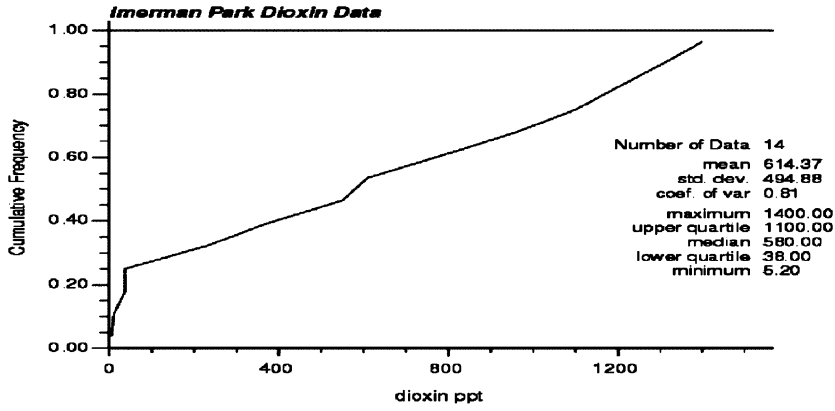

Figure 32.9. Univariate non-spatial statistical measures of the surficial dioxin concentrations in Imerman Park.

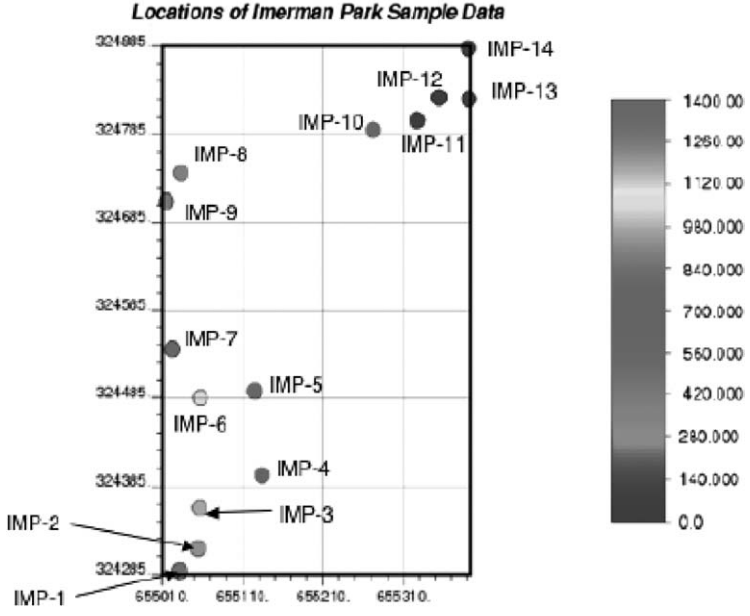

Figure 32.10. Data sample locations shown in a grid outlining Imerman Park.

Low dioxin levels are found in the northeast park section, near the park entrance and the 100-year flood plain contour. Testing was done primarily at the southwest and the northeast corners of Imerman Park, with no sampling performed in the center or southeast sections of the park.

Figure 32.11 shows the dioxin concentrations estimated using an inverse distance squared algorithm and a Euclidean distance norm. The data values are honored at the sampled locations, with the estimated values far from the sampled locations converging on the mean of 614 ppt. This data convergence occurred because that area of the park was equidistant to the two data clusters. The inverse distance algorithm is an estimation technique; this algorithm produces deterministic results and does not produce any uncertainty measurements.

Figure 32.12 shows the spatial dioxin distribution modeled using sequential Gaussian simulation, with a Euclidean distance norm. In all the simulations, 100 realizations were run. As with the inverse distance squared algorithm, the sequential Gaussian simulation algorithm also

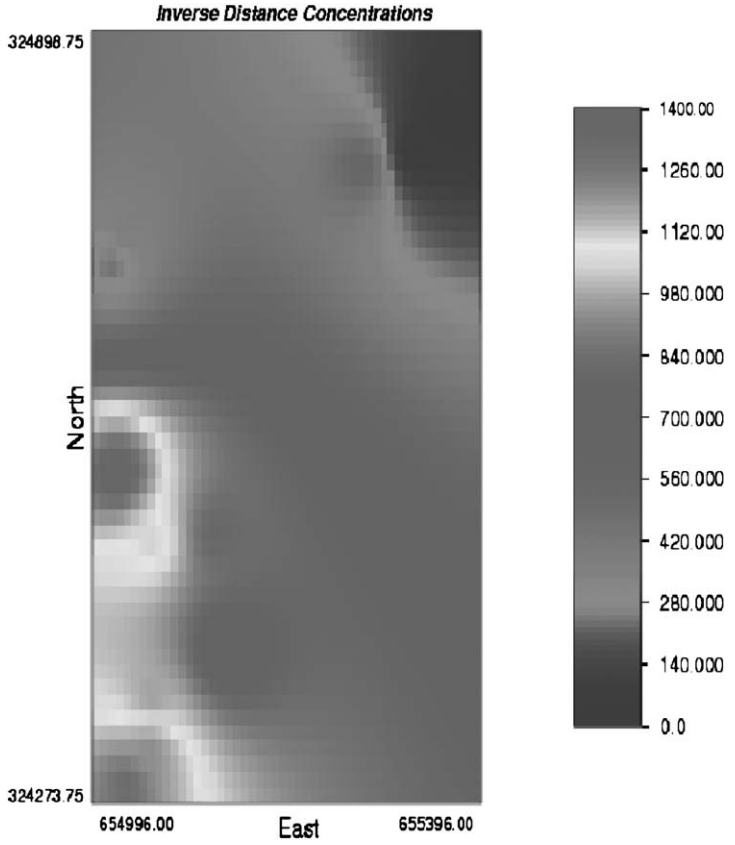

Figure 32.11. Estimated dioxin concentrations, estimated using the inverse distance squared algorithm. This estimation was performed using Euclidean distance norms.

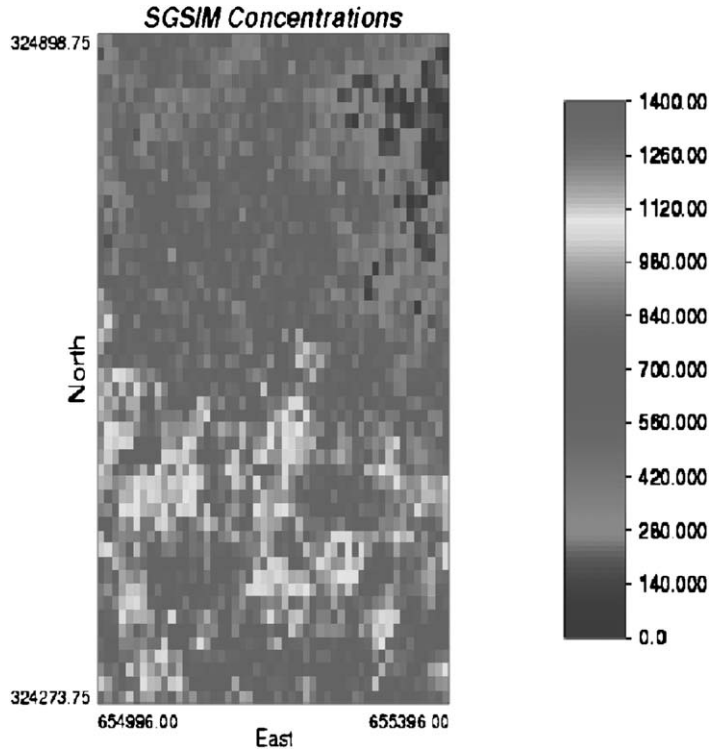

Figure 32.12. Simulation results using the sequential Gaussian simulation algorithm. This simulation was performed using Euclidean distance norms, with 100 realizations.

honors the sampled data values, but as shown generates a ''salt-and-pepper'', non-smoothed concentration map. The sequential Gaussian simulation algorithm is useful for obtaining uncertainty measurements and introducing randomness in the simulated results from multiple realizations, while honoring the data samples and the sample data histogram. In the sequential Gaussian simulations, the data were declustered prior to performing a normal score transform of the data. A normal score transform is performed to transform the data histogram into a Gaussian distribution. Simulation is done in normal space, then the results are back transformed, yielding results that match the data histogram. The models generated were stored in a rectangular geometry; this is a limitation of the GSLIB software.

With both the inverse distance and the sequential Gaussian algorithms, hot spots of high and low dioxin levels are shown. The remediation areas vary, however; with the sequential Gaussian simulation, hot spots can be

seen where no data has been collected, indicating that perhaps further testing may be desired in these areas where the dioxin levels are uncertain.

Figures 32.13 and 32.14 display spatially the probabilities that the dioxin concentration exceeds the MDEQ's 90 ppt threshold for cleanup. In the inverse distance map, no uncertainty is captured in the algorithm. The entire park would have to be remediated with the exception of the far northeast corner, if this map were used for making the remediation decision. The sequential Gaussian simulation algorithm enables multiple realizations to be done, each using a different random walk in a Monte-Carlo simulation; using this map, a probability threshold can be set by the

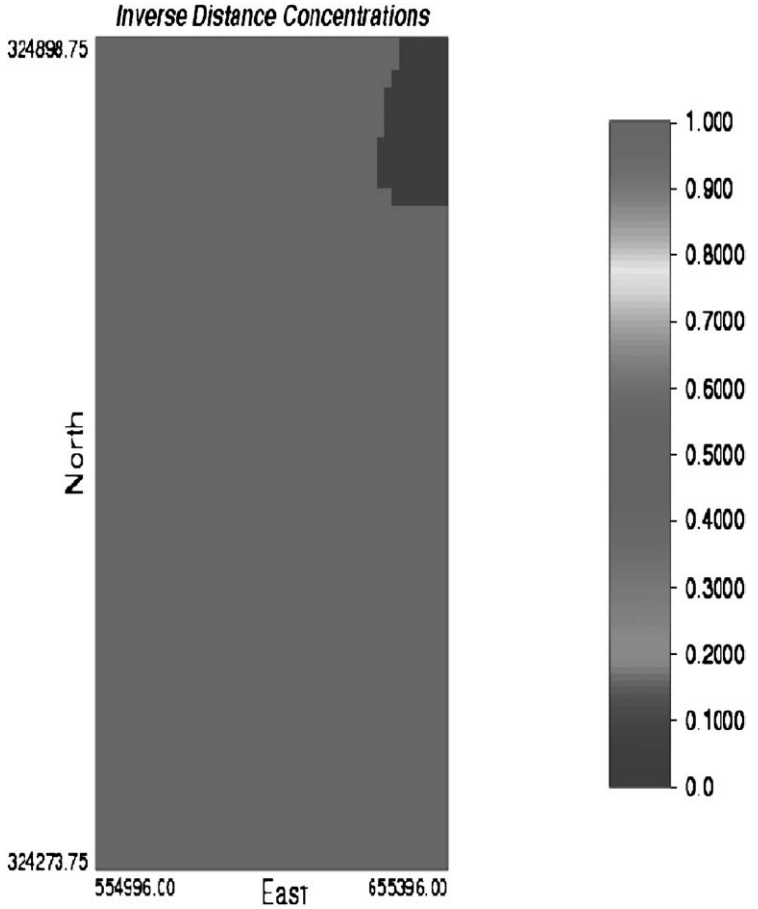

Figure 32.13. Inverse distance probability map, showing the probability that the dioxin concentration is higher than the MDEQ threshold of 90 ppt.

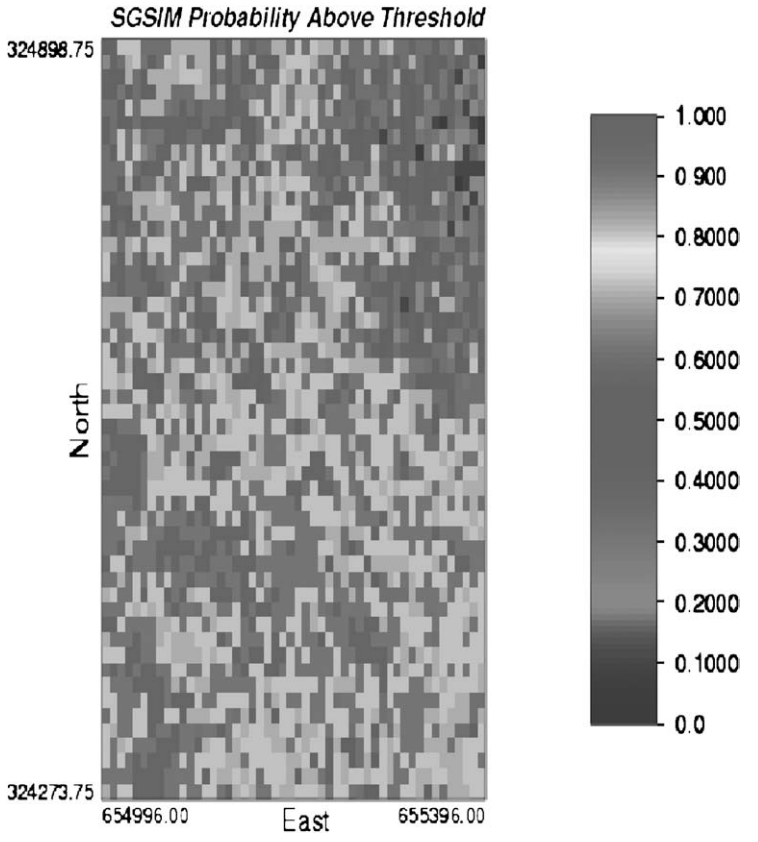

Figure 32.14. Sequential Gaussian simulation probability map, showing the probability that the dioxin concentration is higher than the MDEQ threshold of 90 ppt.

policy makers to determine what extent of the park needs to be remediated. To calculate the probability map for the simulations, the percentage of realizations for each pixel where the simulated value exceeded the target threshold was calculated.

A non-Euclidean distance norm was next used along with sequential Gaussian simulation to perform simulations to characterize the dioxin concentrations in Imerman Park. Figure 32.15 displays the simulated concentrations using a non-Euclidean distance metric. In the non-Euclidean simulation, first a transformed grid was generated as a preprocessing step to sgsim. Gamv was then run to examine the experimental variogram measure using the transformed grid space, then sgsim was run using a variogram model specified using the experimental variogram

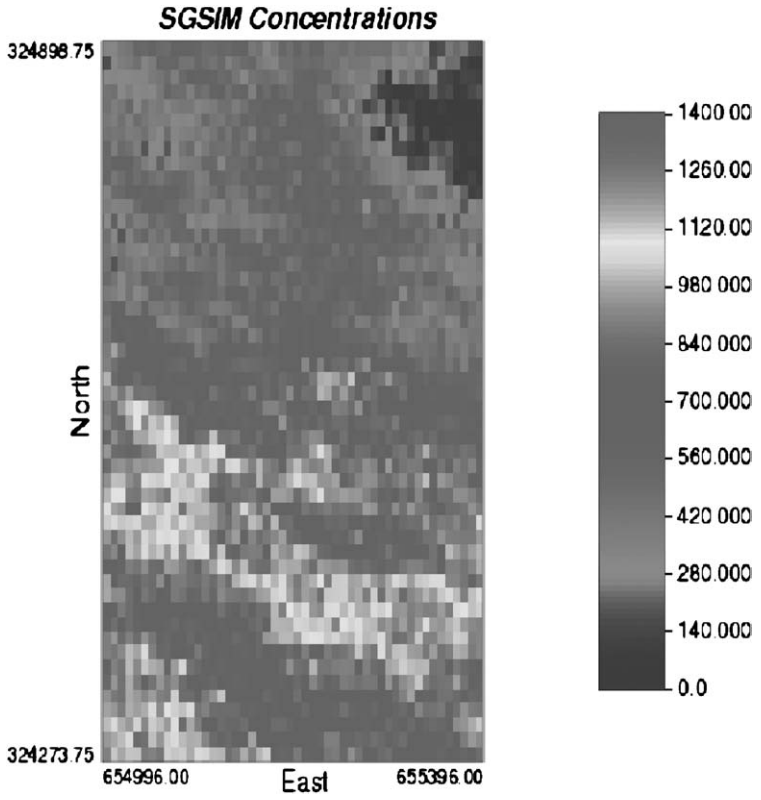

Figure 32.15. Sequential Gaussian simulation results using a non-Euclidean distance norm.

results. The results were then resampled as a post-processing step into the Euclidean space, for display and comparison. Comparing Figs. 32.12 and 32.15, it can be seen that the dioxin concentrations have more structure in the distribution, trending in the direction of the flood plain distance metric. The univariate data distribution still matches that of the sample data, but the spatial distribution has taken into account the flood stages would affect how the dioxins get transported and deposited.

When examining the probabilities of dioxin existing in concentrations above the MDEQ threshold of 90 ppt, the non-Euclidean results again show more structure, trending along the direction of movement of water in the flood plain. Figure 32.16 shows non-Euclidean sequential Gaussian's probability maps. Based on the non-Euclidean results and based on the acceptable probability threshold value as set by the decision makers, Fig. 32.16 shows that it looks more certain that the southwest

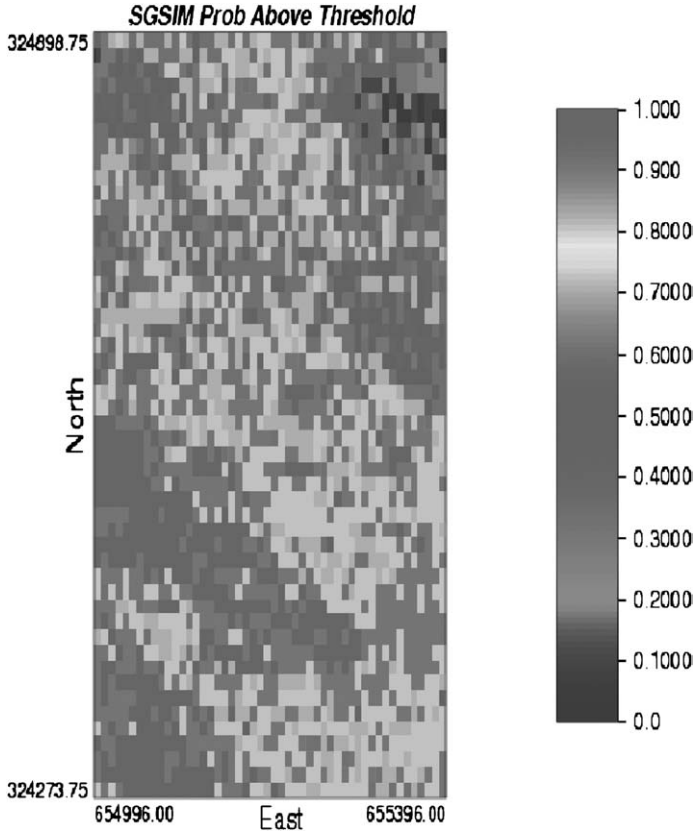

Figure 32.16. Sequential Gaussian simulation probability map using a non-Euclidean distance norm, showing the probability that the dioxin concentration is higher than the MDEQ threshold of 90 ppt.

half of the park needs remediation, with no remediation action required for the northeast corner of the park. This is in contrast to Fig. 32.14, where the simulated results show more randomness in the dioxin distribution.

When performing an actual remediation, usually the remediation unit is much larger than the model grid's resolution. Saito (Saito and Goovaerts, 2002) describes this in more detail. In general, once a geostatistical model is performed, the model is resampled into a lower resolution grid, by taking the arithmetic average of all the model grid cell values that fall within a reservoir unit. A probability threshold is determined by the regulatory agency, and if this probability is exceeded, cleanup will be performed in the reservoir unit.

One major problem with modeling this site arises because of the lack of sufficient data support. There were only 14 samples taken in Imerman Park, and this data was clustered in three primary areas. Given these data limitations, the data histogram, and hence the variogram model chosen and the normal score data transform, contain high degrees of uncertainty. In a remediation study, the author suggests using the resulting model as a guide to aid in developing a more detailed sampling plan, then updating the model as more data is collected. Deutsch (Deutsch and Journel, 1998) describes how to simply calculate statistical measures on multiple simulations. In general, statistics such as variances, medians, etc. can be calculated on a pixel-by-pixel basis using the multiple realizations.

32.3. Conclusions

The use of non-Euclidean distance norms for modeling the spatial variability in a random variable may lead to more realistic results and more informed decision-making when determining the remedial action in a contaminated site. This non-Euclidean norm used must be site and domain specific, to capture the objective and subjective knowledge regarding the processes that led to the distribution of the natural phenomena.

Ideally, more exploratory samples should be taken and the model iteratively updated and refined. There were not enough data samples to perform error analysis on this data set; one area of future work involves obtaining more exhaustive data sets to use with error analyses, and trying to get further samples taken at Imerman Park to validate the model results.

ACKNOWLEDGMENTS

I would like to thank the Michigan Department of Environmental Quality for providing me with the geo-referenced dioxin test results. Thanks also go to the editors and anonymous reviewers for their valuable and constructive comments. Lastly, I would also like to thank my wife and Central Michigan University's Geology Department for their support in this project.

REFERENCES

Barabas, N., Goovaerts, P., Adriens, P., 2001. Geostatistical assessment and validation of uncertainty for three-dimensional dioxin data from sediments in an estuarine river. Environ. Sci. Technol. 35(16), 3294–3301.

Clark, J.M., 2004. Health Risk Analysis of Tittabawassee Fish with Dioxin, EPA Open-File Report, July 30, 2004.

Cressie, N., Gotway, C.A., Grondona, M.O., 1990. Spatial prediction from networks. Chemometrics and Intelligent Laboratory Systems 7, 251–271.

Curriero, F.C., 1996. The use of non-Euclidean Distances in Geostatistics [Ph.D. thesis]. Manhattan, Kansas State University.

Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software Library and User's Guide, second ed.. Oxford University Press, New York, p. 369.

Dow, 2005. History of dioxins and furans in the Tittabawassee River, 2005. From http://www.dow.com/facilities/namerica/michigan/dioxin/what/history.htm web site.

Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York, p. 592.

Journel, A.G., Huijbregts, Ch.J., 1978. Mining Geostatistics. Academic Press, New York.

Krivoruchko, K., Gotway, C.A., 2003. Using Spatial Statistics in GIS. In: Proceedings, Modsim 2003 International Congress on Modeling and Simulation, Townsville, Queensland, Australia, pp. 713–718.

MDEQ, 1999. Michigan Department of Environmental Quality, 1999, Michigan soil background Data, from http://www.michigan.gov/deq/ web site.

MDEQ, 2002. Taylor, Allan B. and McCabe, John M., 2002, Baseline chemical characterization of Saginaw Bay Watershed sediments, from http://www.michigan.gov/deq/ web site.

MDEQ, 2005. Michigan Department of Environmental Quality, 2003, FINAL REPORT: Phase II Tittabawassee/Saginaw River Dioxin Flood Plain Sampling Study, from http://www.michigan.gov/deq/ web site.

Midland [2005] 2005. Interim Response Activities Decided: Dioxin Contamination Issues in Midland, from http://www.midland-mi.org/government/Dioxin/Updates.htm#Framework web site.

Rathbun, S.L., 1998. Spatial modeling in irregularly shaped regions: Kriging estuaries. Environmetrics 9, 109–129.

Rendu, J.M., Readdy, L., 1992. Geology and the Semivariogram—A critical relationship. In: T.B. Johnson, R.J. Barnes (Eds.), 17th Application of Computers and Operations Research in the Mineral Industry, American Institute of Mining, Metallurgical, and Petroleum Engineers, Inc., pp. 771–783.

Saito, H., Goovaerts, P., 2000. Geostatistical interpolation of positively skewed and censored data in a dioxin contaminated site. Environ. Sci. Technol. 34(19), 4228–4235.

Saito, H., Goovaerts, P., 2002. Accounting for measurement error in uncertainty modeling and decision making using indicator kriging and p-field simulation: Application to a dioxin contaminated site. Environmetrics 13, 555–567.