ArcGIS Geostatistical Analyst Application in Assessment of MTBE Contamination

Jie Y. He, and Xudong Jia

ABSTRACT

Methyl-Tert-Butyl-Ether (MTBE) is an oxygenated fuel addictive to reformulated gasoline. It is often considered as the leading groundwater contaminant that causes a threat to public health and living environment. This paper first discusses the background of MTBE contamination. It then uses the ArcGIS Geostatictical Analyst to explore sample data, analyze spatial distribution and autocorrelation of MTBE concentration data, and predicts MTBE concentrations in the groundwater of the City of Temecula, California. MTBE concentration data were collected from various sources, including the California Regional Water Quality Control Board and the Geotracker Web site. Finally The MTBE concentration spatial surface for the study area was generated, and the probability maps of the MTBE concentration that exceeds a critical threshold were created and cross-validated. The findings derived from this research strongly show that ArcGIS Geostatistical Analyst can greatly assist in the MTBE remediation if sample data are large enough.

INTRODUCTION

Methyl-*tert*-butyl ether (MTBE) is an oxygenated fuel addictive to reformulated gasoline (RFG). It is used to enhance gasoline combustion and to improve air quality. However, the U.S. Environmental Protection Agency (USEPA) has tentatively classified MTBE as a possible human carcinogen. MTBE is highly soluble in water and is chemically inert with low biodegradability. Once dissolved in water, it will persist and widely spread in the environment, becoming a significant contaminant of surface water and groundwater resources. The USEPA, Cal-EPA, and the California Department of Health Services have issued drinking water health advisory values for MTBE based on public health, cancer risk, taste, and odor considerations (California Department of Health Services, 2003).

Recent studies have shown that MTBE is a prevailing problem to the public. The United State Geological Survey has considered MTBE the second-most frequently detected groundwater contaminant in its National Water Quality Assessment Program. At the City of Denver, Colorado, samples from 79% shallow urban wells that were investigated were tested with the harmful concentrations of MTBE (USGS, 1995). In the City of Santa Monica, California, MTBE has been detected in groundwater with its concentrations of hundreds of parts per million. Because of the MTBE concentrations tens of thousands times higher than the regulatory drinking water advisory values, three of the five wells that provided water to 40% of the city's population were shut down. The city had to purchase additional water from the Metropolitan Water District of Southern California at a cost of $2 million per year (Regional Water Quality Control Board Santa Ana Region, 1999).

MTBE contamination has caused great concerns to the citizens, government agencies, scientists and the industry. In California, provisions on MTBE investigation and reporting requirements have been written into the law. The State Water Resources Control Board and its nine Regional Water Quality Control Boards are the lead State agencies responsible for MTBE investigation and cleanup. Under the mandates of California Water, Health and Safety Codes, the dischargers are required to test MTBE and report their results to the regional boards.

A good understanding of how MTBE is transported and distributed in groundwater is critical to the MTBE investigation and cleanup. This paper is aimed at the use of GIS and Geostatistical technologies to monitor and analyze the spatial patterns of MTBE in groundwater. The ArcGIS Geostatistical Analyst developed by the Environmental Systems Research Institute (ESRI) was used for this project. The Geostatistical Analyst consists of a great set of geo-statistical techniques including those for exploratory spatial data analysis, structural analysis, and surface prediction and assessment. The MTBE concentration maps developed from these techniques can provide valuable information for the groundwater cleanup and risk management decision-making.

STUDY AREA AND DATA COLLECTION

A groundwater basin located in the City of Temecula, CA was selected for this project (see Figure 1). Because severe MTBE contamination in the basin was found in the drinking water supply wells, the Rancho California Water District (RCWD) immediately shut down the wells that draw the water from the basin and investigated the extent of the MTBE impacts (San Diego Regional Water Quality Control Board, 2001).

This project used the sample data collected by the San Diego Regional Water Quality Control Board, Geotracker and other consulting firms for the above groundwater basin. As shown in Figure 2 (Please click the slide number in the left column to view the corresponding figure, so as for the other figures in context.), there are a total of 6 open gas stations in the vicinity of the highly contaminated production well (RCWD No. 118). Each gas station site has approximately 10 monitoring wells. Table 1 shows an example of the information collected from an MTBE monitoring well. The latitude and longitude of the monitoring site and the MTBE concentration is measured for each sample well.

Table 1 MTBE Concentrations for a Monitoring Well (RCWD Site 12)

	Field Name	Date Measured	MTBE Concentration (ug/L)	Latitude	Longitude	XY DATUM
1	MW-8	12/12/2002	200	33.4931764	-117.1485831	NAD83
2	MW-6	12/13/2002	2	33.4933428	-117.1481175	NAD83
3	MW-7	12/14/2002	590	33.4931728	-117.1483372	NAD83
4	MW-10	12/15/2002	710	33.4928986	-117.1484325	NAD83
5	MW-1	12/16/2002	34000	33.4930961	-117.1481717	NAD83
6	MW-11	12/17/2002	950	33.4927033	-117.1483147	NAD83
7	MW-9	12/18/2002	4	33.49229	-117.1483311	NAD83
8	MW-3	12/19/2002	530	33.4929806	-117.1481131	NAD83

9	MW-2	12/20/2002	19000	33.4930031	-117.1480161	NAD83
10	MW-4	12/21/2002	1500	33.4927975	-117.14812	NAD83
11	MW-5	12/22/2002	28000	33.4928844	-117.1478739	NAD83
12	MW-12	12/23/2002	2	33.4928111	-117.1477311	NAD83

(Source: GeoTracker No/Global ID T0606501130)

MTBE ASSESSMENT PROCESS

With the sample data collected for the study area, the project members have assessed the MTBE distributions in groundwater using the ArcGIS Geostatistical Analyst. The assessment process involves the following steps in order to create the prediction surface of MTBE concentration and the probability map for MTBE exceeding a threshold value in groundwater.

Step 1: **Exploratory Spatial Data Analysis**

The exploratory spatial data analysis uses various data analysis tools to explore the sample data, executes necessary transformation to reduce the skewness of the data, and predetermine the MTBE distribution pattern. It also identifies data global trend and calculates semivariogram values to discover spatial autocorrelation of the MTBE data. This process is useful in getting a deeper understanding of the MTBE sample data and establishing a solid foundation for further statistical spatial analysis and best-fit modeling of MTBE concentrations.

Step 2 **Structural Data Analysis**

The structural data analysis process uses kriging method to interpolate MTBE data, generates a predicting surface by weighting the spatial relationship among the sample data location and prediction location. Variography is used in quantifying spatial data structure, and a spherical semivariogram model is used in determining the best geo-statistical model for surface formation. The Anisotropy tool is used in identifying and adjusting directional influence for a more accurate prediction of MTBE distribution.

Step 3: **Cross Validation**

The cross validation process examines the accuracy of the surfaces generated in Step 2 by using a number of statistical criteria. It compares the appropriateness of different semivariogram models in predicting MTBE concentrations for a specific site. It enables the generation of probability maps that show the MTBE concentrations exceeding a pre-defined threshold at a given location.

These three steps are interrelated. Normally the exploratory spatial data analysis is performed first so that the spatial autocorrelation of MTBE sample data can be determined. In the stage of the structural data analysis, the kriging use IDW to further refine spatial autocorrelation. The cross validation process evaluates MTBE surfaces generated from different Kriging methods and selects a best surface that fits the sampled MTBE data.

EXPLORATORY SPATIAL DATA ANALYSIS

Once the sample data are collected and compiled from different sources, the project team has investigated the MTBE data by checking their histograms, directional trends and covariance. Histograms provide a univariate (one-variable) frequency distribution of the MTBE data and help the team members determine if a normal distribution is applied to the sampled MTBE data. When the raw MTBE data are not normally distributed, the ArcGIS Geostatistical Analyst provides log transformations for converting skewed distributions into normal distributions. Figures 3 and 4 show the transformation process of the MTBE data. The MTBE concentrations are positively skewed in Figure 3. After the log transformation, the MTBE concentrations are approximately normally distributed (see Figure 4).

The trend analysis tool from the ArcGIS Geostatistical Analyst provides a three-dimensional perspective of the MTBE directional trends. Figure 5 demonstrates that the MTBE data seem to exhibit a strong trend in the north-south direction but weaker in the east-west direction.

Semivariograms/covariances represent the spatial autocorrelation between the measured sample points. Given Z(Si) and Z(Sj) are the measured values at locations Si, and Sj. . The variogram value S(γ) is the average of $[(Z\ (Si) - Z\ (Sj)]^2$. If all of the location pairs (Si and Sj) are close each other, it is expected that Z(Si) and Z(Sj) will be similar in value so that the variogram value S(γ) should be small. The pairs of sample locations selected in the semivariogram are highlighted on the map (see Figure 5). Lines are interconnected between paired locations (see Figure 6). When the semivariogram values are large for two near locations, the semivariogram values indicate that the MTBE distribution in paired locations is impacted by abnormal factors.

STRUCTURAL DATA ANALYSIS – KRIGING

Kriging is a data interpolation method to create prediction surface. It weighs the surrounding measured values to derive a prediction for an unmeasured location. Weights are based on distance between the measured points, the prediction location, and the overall spatial arrangement among the measured points (McCoy and Johnston, 2001).

One of the essential features of Kriging is that the data points need to be sampled from a continuous space. The MTBE concentration data used in this project were randomly collected as points from a continuous MTBE plume. The Kriging is therefore suitable for the surface prediction of the MTBE concentrations.

The Kriging modeling process used in the project involves 1) the selection of a Kriging method for sampled MTBE data and 2) the mapping of the MTBE concentrations with the selected Kriging method. The selection of a Kriging method is based on the autocorrelation of MTBE concentrations between two points, which is formulated as follows:

$$Z(s) = \mu(s) + \varepsilon(s)$$

Z(s) consists of two parts: a deterministic trend μ(s) (i.e. flow direction) and a random autocorrelated error ε(s). The symbol *s* simply indicates the location of a point on a continuous MTBE plume. Because μ(*s*) is a deterministic trend, the selection of a Kriging method is based on whether a directional trend exists or not. If a directional trend is unknown, then an Ordinary Kriging method is appropriate. If a directional trend is known, then a Simple Kriging method should be selected (Johnston, Hoef, Krivoruchko, and Lucas 2001).

The Simple Kriging method was selected in this project to map the MTBE distribution because the Murrieta Creek watershed flows from Northwest to Southeast, about S41^{o}E.

There were three tasks involved in using the Simple Kriging method to model the MTBE surface:

Task 1.Determine the spatial structure of MTBE Concentration

This task uses the spatial dependence model developed from the semivariogram/covariance analyses to produce a prediction surface for MTBE concentrations. As discussed before, the overall spatial autocorrelation of the MTBE sample data are represented by a spherical model whose semivariogram values are best fitted to those semivariogram values generated from sampled MTBE data points. Figure 7 shows the spherical model in a solid line.

There are three primary parameters that describe the autocorrelation of the MTBE concentrations. These parameters are range, nugget and partial sill. As shown in Figure 7, the range is where the best-fit line starts to level off, in this case 26.28. Within the range, all data are correlated. Beyond the range where the best-fit line of the semivariogram flattens out (plateau or sill), there is little autocorrelation among the data. The nugget and partial sill will be further discussed in the Directional Influence section.

The color scale as shown in Figure 7 represents the calculated semivariogram value. It provides a direct link between the semivariogram values on the graph and those on the semivariogram surface. The value of each "cell" in the semivariogram surface is color-coded. The lower values are in blue and green. The higher values are in orange and red. The semivariogram surface indicates that the semivariogram value is small in the northwest to southeast direction because most cells are represented in blue or green. In that direction, MTBE data are highly correlated.

Task 2 Determine Directional Influences

Directional influences are critical to the accurate estimation of the MTBE surface. Caution should be taken when MTBE sample data are collected near a preferred pathway of groundwater. As the MTBE sample data for this project were collected near the Murrieta Creek, Figure 7 shows a strong directional influence on the semivariogram surface.

The Direction Search tool (also called Anisotropy) provided by the ArcGIS Geostatistical Analyst can statistically quantify directional influences. By clicking and dragging the arrow in the semivariogram surface (see Figure 7) to line up with the directional orientation of 320.8 degree (same as groundwater flow direction S41^{o}E), the anisotropy will calculate the nugget variation when the Nugget Error Modeling box is checked. After the directional influences are removed from the MTBE data, the semivariogram values are better correlated (see Figure 8). The best-fit line is slightly modified with a new range (35.04) and sill (62).

Task 3 Search Neighborhood

The Neighborhood Search method in the ArcGIS Geostatistical Analyst utilizes the Inverse Distance Weight (IDW) principle to consider autocorrelation of the adjacent MTBE points in predicting the MTBE concentration at a specific location.

In predicting the MTBE concentration of a specific point, the number of adjacent points, the searching radius, and the number of sectors of the circle (or ellipse) should be specified. As shown in Figure 9, eight neighboring points are considered. An ellipse shape was selected with its major/minor axis oriented along the identified directional trend. In order to avoid bias in a particular direction, the circle or ellipse can be divided into sectors from which an equal number of points are selected and all directional influences will be fairly considered.

The MTBE concentration of a specific location can be predicted by clicking on the location. Figure 9 illustrates the predicted MTBE concentration of 1058.3 at the specific location (x=-117.15037, y=33.503362), as well as the weights of each neighboring points.

CROSS VALIDATION

Before the final surface is produced for practical use, the cross-validation tool from the ArcGIS Geostatistical Analyst should be employed to examine "how well" the surface model predicts an unknown MTBE concentration at a specific location. The cross-validation tool compares measured MTBE values with predicted ones derived from the surface model, and uses statistical measures to assess the surface model's performance. The statistical measures serve as guidelines for the accuracy and reasonableness of the surface model and its prediction map.

Figure 10 provides a graphical comparison between measured and predicted values. The cross-validation tool sequentially omits a point, predicts the MTBE value for the missing points using the rest of the MTBE data, and then compares it with the measured value. Ideally, the predicted values should be the same as the measured ones, and all data points would form a 1:1 ratio line (the gray line in Figure 10). In reality, data points would scatter along this line due to natural variations and uncertainties, and the term "prediction error" is used to describe the difference between the prediction and the actual measured value. For a surface model that provides accurate predictions, the Mean error should be close to 0, the Root-Mean-Square error and the Average Standard error should be as small as possible, and the Root-Mean-Square Standardized error should be close to 1 (Johnston, 2001).

However, some of these criteria would be too stringent and unrealistic for MTBE sample data. The project has used the Mean, Root-Mean-Square, and Average Standard Errors criteria to cross validate the surface models generated from different Kriging methods and select a surface model that has the best prediction capabilities. Figure 11 compares the cross-validation results for the Ordinary Kriging and Simple Kriging models. By comparison, the Simple Kriging outperforms the Ordinary Kriging in every statistic category, except for the Mean Standardized value.

MTBE CONCENTRATION PREDICTION MAP

With the best surface map determined from the cross validation process, the MTBE concentration prediction map that shows the MTBE distribution in groundwater can be generated. Figure 12 shows is a MTBE prediction map in which the dark color represents high concentrations, the light color represents low concentrations.

As uncertainties and limitations are embedded in the MTBE prediction, care should be taken to use the predicted MTBE concentration map to make regulatory or remediation decisions. The ArcGIS Geostatistical Analyst provides a tool to help users determine a probability map that shows the probability of the MTBE concentration of a given point exceeding a predefined threshold.

When the Geostatistical Analyst creates a probability map, all MTBE concentrations are transformed to a series of 0s and 1s according to whether the values are below or above the predefined threshold. Figure 13 illustrates the data transformation, where the lower right corner contains a data table, the top chart shows that all points to the left of the blue line have a transformed value of 0, whereas all points to the right of the blue line have a transformed value of 1. In the prediction error section, it shows that the mean error is 0.214, the root mean square is 0.4858, which means the error is relatively small, the probability surface is fairly accurate, and the prediction is quite reliable.

Figure 13 is a probability map generated from the project. The colored contour lines indicate the probability of MTBE concentration exceeding the defined threshold of 20 ppb. Anywhere on the map, when a pointing device such as a mouse is clicked, the MTBE concentration and its probability of exceeding a defined threshold will be identified.

This probability map can visually aid regulators and the environmental consultants to make rational decisions. For example, if drinking water wells were to be located within the green contour lines, chances are that in more than 60% of the cases the MTBE concentration in the well would exceed the threshold of 20 ppb, and some actions (such as shut down or cleanup) need to be taken to solve the problem.

Another use of the probability map shown in Figure 14 is when monitoring wells are placed along the western boundary of the site in the high concentration zone (second darkest color zone in Figure 14). If concentrations in these monitoring wells are increasing with time, they will impact the water production well #118. The regulators may view this as a high priority task and ask the owner to clean up. The owner in turn may want to find out the costs and duration of the MTBE cleanup.

His consultant can use this map to estimate the total mass of MTBE within his site boundary, and calculate the costs and time based on the existing cleanup technologies and removal rates. The consultant may further use this map to design a remedial system to maximize the removal efficiency by placing extraction wells at appropriate locations. The consultants and the regulators also want to know how MTBE concentration changes in response to cleanup activities.

Additionally, environmental professionals would be able to assess the effectiveness of the remedial system and make necessary adjustments by generating sequential MTBE maps using new monitoring data. When the MTBE concentrations are reduced to certain level, the regulators can use the probability map to determine, with a predefined confidence level, when cleanup action is complete.

CONCLUSIONS AND FUTURE WORK

MTBE contamination threatens water resources and has become an emerging issue in the nation. The project described in this paper utilizes the ESRI's ArcGIS Geostatistical Analyst to analyze the spatial distribution patterns of MTBE in groundwater. The prediction maps generated in this project show an effective way to help investigate and remedy MTBE contamination. They provide visual representations of MTBE distributions in groundwater and help environmental professionals prioritize their limited budgets for MTBE cleanup, develop field investigation strategies and design remedial systems.

There are uncertainties and limitations found in this project in using the sample data for the determination of MTBE distributions in groundwater for the study area. First, a small set of samples was used in this project. The MTBE concentrations were significantly skewed and non-morally distributed. Prediction of MTBE distributions in groundwater using log transformation to convert the skewed datasets into morally distributed datasets could be not accurate. A larger size of samples should be needed if physical and economic constraints are not significant.

Second, MTBE concentrations in the groundwater can be influenced by soil and hydrologic condition of the area. MTBE, once released into the ground, might undergo some physical, chemical and biological reactions that could possibly alter concentration distribution. Human errors and laboratory uncertainties during sampling and data acquisition are also possible factors for prediction errors. These uncertainties and limitations should be considered in the assessment of MTBE contamination.

ACKNOWLEDGEMENT

We would like to thank Dr. Bill Lloyd for providing the software access. His encouragement and guidance was the foundation of this paper. We also thank Dr. John Carroll and Dr. Jonathan Taylor, for their suggestions and clarity for the paper. We also thank Dr. Charles Cheng for his assistance and perfecting for the paper.

REFERENCES

California Department of Health Services, 2003. Drinking Water Monitoring for MTBE, May 2003

ESRI, Using ArcGIS Geostatistical Analyst, ESRI, California, 2001

Johnson, Kevin M. 1992. Using Statistical Regression Analysis to Build Three Prototype GIS Wildlife Models. *GIS/LIS 92 Proceedings,* 1:374-386.

Johnston, K. ; Hoef, J.M.V. ; Krivoruchko, K. ; Lucas, N. 2001. *Using ArcGISTM Geostatistical Analyst.* ESRI. 380 New York Street. Redlands, CA 92373-8100, USA.

Johnson, M.L. 1999. Ecological Risk of MTBE in Surface Waters. John Muir Institute of the Environment, UC Davis.

Office of Environmental Health Hazard Assessment, 1998. Public Health Goal for Methyl Tertiary Butyl Ether (MTBE) in Drinking Water (Draft), April 1998.

San Diego Regional Water Quality Control Board, Staff Report: MTBE Detection in Temecula City, prepared by Berry Pulver, 2001.

San Diego Regional Water Quality Control Board, The Fourth Quarter Groundwater Monitoring Report, Temecula City, Dec 2002.

Santa Ana Regional Water Quality Control Board, Staff Report on MTBE Contamination, prepared by Ken Williams, 1999

Yong, T.M., and Couch, K. 1998. Leaking Underground Storage Tanks as Point Sources of MTBE to Groundwater and Related MTBE-UST Compatibility Issues. Department of Civil and Environmental Engineering, UC Davis.

U.S. EPA Region 9: Charnock MTBE Cleanup/ Regional Investigation

U.S. EPA, 1997. Drinking Water Advisory: Consumer Acceptability Advice and Health Effects Analysis on Methyl-tertiary-Butyl Ether (MTBE), Office of Water, Washington, D.C. 20460, December 1997.

U.S. Geological Survey, Fact Sheet FS-114-95, Occurrence of the Gasoline Additive MTBE in Shallow groundwater in Urban and Agricultural Areas, March 1995.

AUTHOR INFORMATION

Jie Y He, GIS Analyst
San Mateo County Building and Planning Division
455 County Center, Redwood City, CA 94063
Phone: 650-363-1827 Fax: 650-363-4849
Email Address: jhe@co.sanmateo.ca.us

Xudong Jia, Ph.D., P.E.
Associate Professor
Department of Civil Engineering
California State Polytechnic University, Pomona
Pomona, CA 91768
Phone: 909-869-4312 Fax: 909-869-4342
Email: xjia@csupomona.edu