J. ENVIRONMENTAL SYSTEMS, Vol. 30(1) 1-26, 2003-2004

LESSONS FOR ENVIRONMENTAL MANAGERS IN CHOOSING AND APPLYING MODELS FOR GROUNDWATER TRANSPORT AND REMEDIATION: A CASE STUDY

SARA SAYLER
Tinker Air Force Base

WILLIAM F. McTERNAN

SCOTT A. WARD

Oklahoma State University, Stillwater

ABSTRACT

A case study is presented which highlights the large discrepancy in projected environmental remediation requirements resulting exclusively from the selection and application of alternative modeling approaches for a contaminated groundwater system. Initially, a state of the practice numerical modeling system was employed to determine the final contaminant concentrations and plume configurations which could be expected in future years. These methods and models have become almost standard practice in these types of situations. In this particular application, however, the final model configuration was much more complex than was allowed by the collected data and by the aquifer system being modeled. Internal auto-calibration techniques were needed to calibrate critical field collected data with model results. The resultant model although calibrated projected contaminant plumes in excess of remediation goals well beyond regulatory property lines. Projected costs of remediation were in excess of 4 million USD. The projected concentrations and plume configurations following calibration did not, however, agree with time series data established during an extended monitoring program. These collected data suggested a much slower rate of plume migration. Inspection of the initial transport model suggested that the auto-calibration procedures resulted in parameter values far in excess of those measured by a variety of agencies and individuals. An alternative set of analytical procedures were employed to evaluate the statistically possible range of input parameters possible from aggregate measured data and to employ these data in various modeling configurations to define plume configurations and travel times beneath the study area. Geostatistical analysis and Monte Carlo simulations were employed together with relatively simple analytical transport codes in this effort. The results, statistically more robust than those derived in the initial effort, show a significantly reduced plume footprint in future years, with a corresponding reduction in projected remediation costs. The original 4 million USD estimate was reduced to approximately $155,000 USD. Most significantly, these results and projections are based upon techniques that are consistent with the measured data. This case study illustrates the types of technical concerns confronting environmental managers. By necessity, models frequently must be applied to help amplify complex conditions. If the wrong model is selected or if it is improperly configured, however, the resulting analysis may be woefully incorrect resulting in the expenditure of needless resources that may not provide adequate environmental protection. This article develops one set of methodologies available to address these concerns.

BACKGROUND

Environmental managers are routinely confronted with difficult decisions relative to how best to address potentially conflicting local and federal requirements for remediation and/or compliance. The position is part technical advisor, part contract manager, and part negotiator with concerned citizens and regulatory groups. The manager is responsible for determining the appropriate technical course to be undertaken in defining an acceptable level of residual contamination, identifying appropriate contractors, and assessing their final reports and work efforts. This is made more complicated by the myriad of technical tools available, each often championed by a specific contractor or regulatory agency. It becomes the responsibility of the military as well as the private sector environmental manager to identify which of these tools is most appropriate to the specific conditions found at the subject site.

This study initially involved a detailed evaluation of groundwater flow and transport modeling at Tinker AFB, Midwest City, Oklahoma (USA). Specifically, this effort focused upon the transport and subsequent remediation of Benzene and Trichloroethylene under the CG037 site, which is approximately the northeast quarter of the base. This site is underlain by the Hennessy and Garber-Wellington aquifers, included within the larger Central Oklahoma Aquifer. Sampling indicated that past practices caused a variety of hydrocarbon and chlorinated organic compounds to be introduced into the groundwater under the base. Critical questions include the overall area(s) of contamination, the rates of contaminant migration (particularly in the off-base direction), and identification of appropriate necessary remediation measures. This article details efforts initiated to address these concerns. A brief overview of an original modeling effort is presented. This work, not completed by these authors, resulted in projections of relatively large-scale groundwater contamination accompanied by significant restoration costs.

This first effort employed models generally held to be State of the Practice codes:

- USGS MODFLOW [1], the flow model and
- MT3D for contaminant transport [2].

Known as deterministic-numerical models, these very complex codes are fully capable of addressing the significant geologic and hydrogeologic heterogeneities found in the Central Oklahoma aquifer. Equations of flow and transport are respectively solved by numerical methods such as finite difference to address aquifers of significant spatial variation.

These codes, however, are also highly data intensive, often requiring thousands of input variables before modeling can be initiated. When data are missing, the codes can employ "auto-calibration" algorithms that allow the model to estimate input parameters on the basis of comparisons of outputs to measured values. Difficulties arise when multiple variables can either singly or in tandem interact to allow calibration for geologically inconsistent data sets. Environmental managers must address these discrepancies before model results are either released for publication or employed in the design of remediation methods. For example, the combined results from these codes projected excursions well beyond Tinker AFB boundaries for benzene and trichloroethylene (TCE) 30 years into the future.

These projections, based on these "auto-calibration" techniques, contradicted previously collected information that suggested a much slower rate of migration. While there is much that is not known regarding the age, location, direction of movement, and sources of these contaminants, recent monitoring data strongly indicate a much slower rate of migration. As these models were intended to aid in defining risk to exposed populations and to assist in designing remediation measures, this disparity had potentially significant consequences. Environmental managers at Tinker AFB were confronted with a situation where the analysis tools employed produced results that were counter-intuitive. A research effort was initiated to produce more robust analytical methods capable of utilizing the existing data to generate plume configurations and travel times consistent with underlying transport theory.

This second approach involved the application of statistical and geostatistical estimation and simulation to define similar events. Techniques such as kriging and Monte Carlo Simulation were utilized in conjunction with the application of simpler analytical groundwater transport models to identify future conditions at Tinker AFB. This article is primarily concerned with this research effort but will introduce data as necessary from the former to make comparisons.

Numerical Models

The modeling sequence employed in the original effort was typical of many other applications in that MODFLOW was used to identify flow patterns and predict local groundwater velocities. These and other input data were then used by MT3D to project contaminant plumes 30 years into the future. The resulting concerns, however, centered upon the MODFLOW portion of this initial effort, specifically with the estimation of hydraulic conductivities; in particular, the estimation of vertical conductivity between specific water bearing zones and the corresponding estimation of horizontal or lateral hydraulic conductivity in what is termed the Lower Saturated Zone (LSZ). Values of 30 and 244 feet per day (ft/day), respectively, were required in the MODFLOW model in order to achieve agreement (calibration) between measured potentiometric surfaces and those produced by the model. A cursory analysis indicated that these lateral hydraulic conductivities were significantly greater than those produced during various field efforts at Tinker AFB or by the United States Geological Survey (USGS) in regional studies in the same geologic unit.

The numeric values of vertical conductivity used in these models were based solely on auto-calibration features within the code which generated agreement between the modeled and measured water levels. No evidence was found, suggesting that this parameter had never even been evaluated with field study. It should be reemphasized that vertical hydraulic conductivities of this magnitude have never been observed following field testing in the Central Oklahoma aquifer. The median value as reported by Christenson and Carpenter [3] was 4.5 ft/day for the sandstone strata.

The hydraulic outputs from MODFLOW were then used in MT3D to define contaminant plume geometries and concentration profiles for the plus 30 year time period. It was this effort that resulted in the projections of benzene and TCE excursion beyond base boundaries in the LSZ.

In addition to those significant areas of uncertainty related to selection of hydraulic conductivity, there were other concerns with the original MODFLOW model. Specifically, a five layer numerical model was designed to emulate the underlying hydrogeology starting with the Hennessey formation and continuing through the Garber-Wellington's Upper Saturated Zone (USZ), the USZ/LSZ aquitard, into the LSZ and finishing with a Lower-Lower Saturated Zone (LLSZ). A total of 39,600 nodes per layer were needed to enclose the area of concern resulting in 198,000 total nodes requiring input data. Over 700 individual computer simulations were needed to achieve acceptable calibration levels. Initial efforts employed hydraulic conductivity values similar to those found in the open literature. When these inputs failed to produce an acceptable agreement with measured potentiometric surfaces, the model was placed into "auto-calibration" mode, which varied the input hydraulic conductivities until agreement between measured and modeled water surface was achieved. This required that conductivities far greater than those ever measured on or near Tinker AFB be employed.

The amount of measured data describing hydraulic conductivity throughout the base area was on the order of 100 wells or less per layer. Given the size and complexity of the model to be developed, the disparity between the amount of measured information and the amount of input data required seemed excessive.

The predicted contaminant plume easily escaped the base boundaries over the 30-year planning period. An associated estimate of the costs necessary to remediate this plume were between $1.2 and $4.2 million (USD), depending upon the technology selected.

At this point, the environmental manager must decide if the projections are appropriate. If so, an urgent remedial response to capture and detain this plume at significant costs would need to be initiated. If, however, the differences between the measured data and those used within the model are considered significant, it is possible that the plume will not migrate beyond the base boundaries. At this point, the environmental manager has four options:

- Proceed with the remediation as developed by the seemingly flawed model.
- Initiate a large scale groundwater monitoring program to increase the available data at the site to a level more compatible with the almost 200,000 node requirements of the original model.
- Remodel the site with the new data.
- Develop alternative approaches to address the original concern relative to plume migration.

This case study will detail efforts developed and applied in conjunction with option 4. A program was initiated to:

- Check the inputs to the original model; and
- Develop alternative plume configuration projections.

As the hypothesis is that the original model failed from a combination of excessive variation from the measured hydraulic conductivity by the "auto-calibration" procedure and from the complexity of the model relative to the available input information, these tools should more appropriately employ measured data sets in a manner which addresses spatial uncertainty.

CURRENT EFFORT

The current study addressed the evaluation of this problem in the following manner:

- Database Management
- Univariate statistical analyses
- Bivariate statistical analyses

- Monte Carlo Simulation: Stochastic time of travel analyses
- Analytical transport modeling

Database Management

Extensive effort was placed in an attempt to gather as much information as possible on numerical values for hydraulic conductivity within CG037 at Tinker AFB. This effort focused upon data provided by Tinker AFB personnel, Open-File Reports published by the USGS and project documents from the U.S. Corps of Engineers [3, 4]. These Open-File Reports consisted of, among other items, a statistical summary of hydraulic conductivities of the Central Oklahoma aquifer. The Central Oklahoma aquifer consists of the Garber Sandstone, Wellington Formation, alluvium, and terrace deposits, which form a narrow band on either side of the North Canadian River [5], and the underlying Chase, Council Grove, and Admire Groups. The aquifer spans Cleveland, Logan, Lincoln, Oklahoma, Payne, and Pottawatomie Counties. The findings published in these Open-File Reports are arbitrarily applied to the vast area in which the aquifer resides. However, these findings are generated from tests conducted on a limited fraction of the aquifer. Therefore, these results only truly represent the portion of the aquifer that was specifically tested. To consider how these data may reflect hydraulic conductivity at Tinker AFB, it should be noted that Central Oklahoma aquifer is $\cong 3000$ mi^2 as compared to Tinker AFB that is $\cong 8$ mi^2, or as compared to the area of CG037, which is $\cong 3.5$ mi^2 . This being the case, a spatially varying parameter such as hydraulic conductivity must be considered in relation to its location. Additionally, details of the specific studies, such as their exact location and depth of the wells that were used to compile these data are commonly left vague or are not included in these reports. Therefore, the specific values reported in the USGS Open-File Reports were found to be useful so that typical values could be recognized and distinguished from the extreme outliers. Table 1 presents hydraulic conductivity data for the Central Oklahoma aquifer reported in all USGS Open-File Reports to 1996. As mentioned, the data shown in Table 1 provide a general feel for a range of hydraulic conductivities within the 3000 mi^2 Central Oklahoma aquifer, but are, of course, very location- and well-specific.

In order to obtain more site-specific hydraulic conductivity data, numerous reports of studies conducted at Tinker AFB were reviewed. Specifically, the results of pump or slug tests conducted at distinct well locations on base were the focus of this data acquisition effort. The previous contractor authored the majority of the reports that were reviewed in this effort, and additionally authored the current model in question. In fact, 71% of the final hydraulic conductivity data sets used for this analysis were included in the data set prepared for the numerical modeling effort. Although moderate amounts of slug and, to a lesser extent, pump test data exist, only data that were deemed reliable were retained. Data that had individual test results presented so that their reliability could be assured were

Table 1. Hydraulic Conductivity for Wells Completed in the Central Oklahoma Aquifer [3]

K_{min} (ft/day)	25th percentile (ft/day)	Median (ft/day)	75th percentile (ft/day)	K_{max} (ft/day)	Sample size
.51	.64	.93	1.3	4.8	51
.48	1.6	2.2	2.8	4.8	18
.3	1.3	1.5	1.9	7.1	32
.09	1.5	3	5.5	60	251
.29	.62	1.3	4.9	29	14
.16	2.5	4.5	9.7	120	160

1 – Garber Sandstone/Wellington Formation
2 – Information not provided

retained. Additionally, data considered acceptable based on comments made in an original contractor report [6] were also retained. Although this data screening helped to ensure data quality, it also moderately limited the quantity of data available for analysis.

In addition to reviewing reports for data acquisition, several databases provided by Tinker AFB were also examined. These databases proved to be particularly useful for locating the state plane coordinates of wells and for determining the screened intervals so that slug or pump tests could be assigned to their respective geologic zones. For simplification, the zones were divided only into the USZ and LSZ.

Although over 800 wells were found to be located within CG037, less than 20 had slug or pump tests performed on them. In all, hydraulic data from only nine wells in the USZ and from five wells in the upper portion of the LSZ have been discovered within this area. Figures 1 and 2 locate the data gathered relative to CG037 for the USZ and LSZ respectively. Other well locations within Tinker AFB boundaries where slug or pump tests were performed were included in the database used in the original effort and in this work.

RESULTS

Univariate Statistical Analyses

While probability and cumulative density functions (PDF and CDF) were fitted to the data from all of the pertinent USZ and LSZ wells respectively, only those

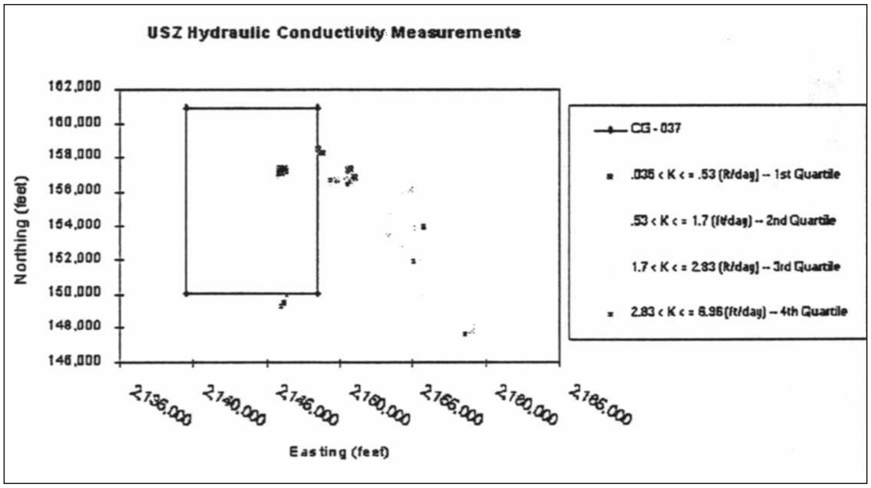


Figure 1. Location map for hydraulic conductivity—USZ.

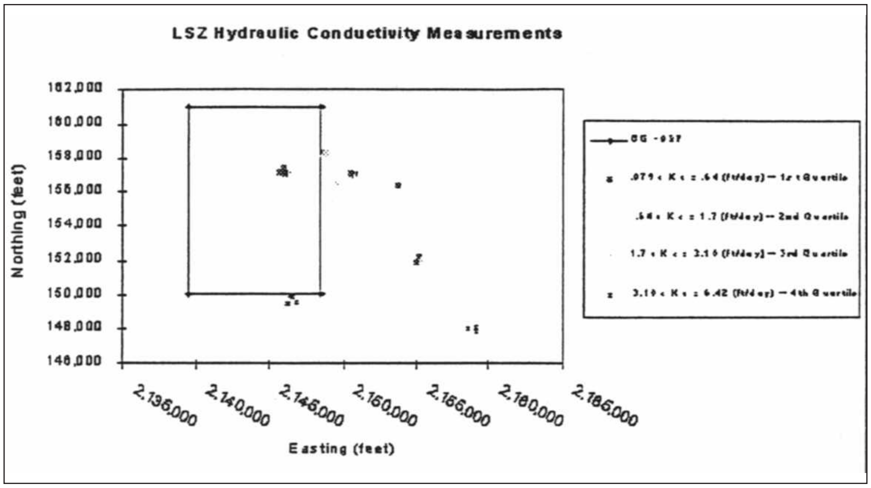


Figure 2. Location map for hydraulic conductivity—LSZ.

resulting from the lower water-bearing zone are included in this article, as it was the one identified in the original work as being the location for contaminant excursions off base. The software package "BestFit" was employed for this task. This software fits 26 distributions to the data. Chi Square, Kolmogorov-Smirnov, and Anderson-Darling techniques identified the most robust distributions for each of the respective water bearing zones [7].

Figure 3 presents the CDF plot for the Triangular distribution for the LSZ which was determined to be the most appropriate fit for these data by the various statistical tests previously identified. This figure shows that 90% of the expected hydraulic conductivity values for the Lower Saturated Zone (LSZ) would occur between 0.2474 and 5.2432 ft/day. Ninety-nine percent (99%) of the expected cases would be less than 7 ft/day (refer to Figure 4). It is functionally impossible to reach the 240 ft/day level achieved in auto calibration with this distribution.

Bivariate Statistical Analyses: Structural Analysis and Kriging

An additional method used to address the quality of the hydraulic conductivity data used in this current CG037 model was the geostatistical estimation technique known as ordinary kriging (OK) [8]. Kriging provides estimation based only on the available data and their spatial continuity. The estimation is independent of any complex models where variables may become determined on the basis of what calibrates the model and not the collected data. Ordinary kriging (OK) utilizes the "best linear unbiased estimator—BLUE" to determine parameter values at

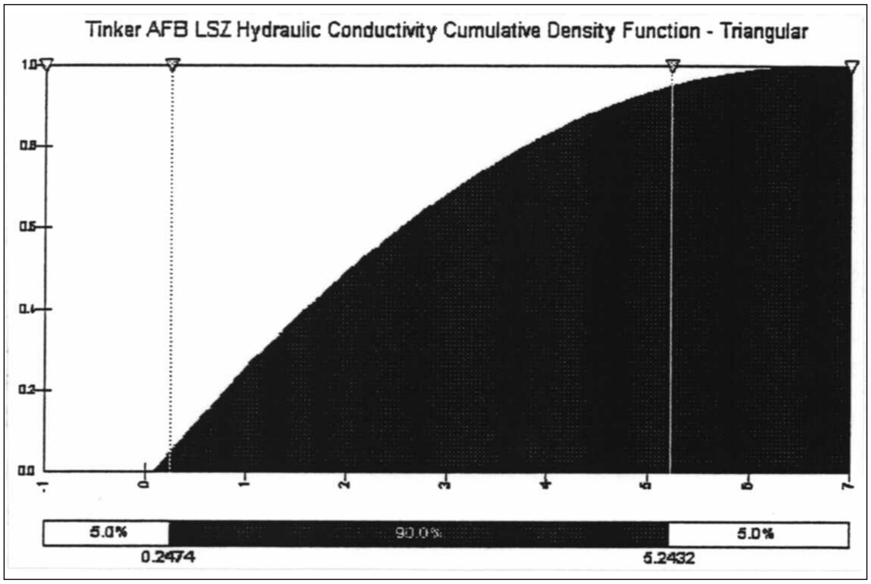


Figure 3. CDF for hydraulic conductivity for LSZ.

unmeasured locations. The hydraulic conductivity data available from Tinker AFB are somewhat limited, clustered in patches, and predominantly not present in CG037, or associated deep wells that penetrate the LLSZ. Kriging is based on probability theory and deals with making deductions with incomplete information and is a common technique frequently applied to estimate spatial variables such as hydraulic conductivity. As such, it is an appropriate technique to apply to CG037 and related databases.

The heart of a geostatistical analysis is the construction of the experimental semivariogram, which plots semivariance as a function of the distance between respective samples. While the calculated semivariogram can be used directly, more frequently a theoretical model is fitted to it. To aid in this effort, the Geostatistical Environmental Assessment software package GEO-EAS was employed [9]. Experimental semivariograms were constructed to estimate the spatial correlation of pairs of measured hydraulic conductivities as a function of separation distance and direction. As the total separation distance between any two measured data pairs was roughly 16,000 ft, one-half of this distance, 8,000 ft, was used as the maximum distance to correlate data. This was consistent with published guidance [10]. This working distance was further divided into 10 equal 800 ft increments called lags. Thus, the semivariance of the Tinker AFB data set was computed at 800 ft lags, and then plotted to provide the structure for the experimental semivariograms.

After completing an omni-directional (i.e., 360° window) experimental semivariograms for both the USZ and LSZ, directional-semivariograms were also developed in an attempt to address anisotropic conditions. The directional-experimental semivariograms were constructed from a 180 degree arc divided into 45 degree increments. By dividing the omni-directional experimental semivariogram into angles of 0, 45, 90, and 135 degrees, using an angle tolerance of 22.5 degrees, the omni-directional semivariogram was divided into four equal subsets. Therefore, data correlation at these angle increments could theoretically be identified. However, due to the limited and clustered data set, directional experimental semivariograms were considered invalid. These semivariograms were considered erroneous due to two main reasons. The first reason is that some of experimental semivariograms simply did not have enough data points to accurately define an appropriate model. In these instances, near and far lag distances, as defined by the respective semivariogram, may have been computed, while missing intermediate lag distances, and thereby giving an inadequate description of the hydraulic conductivity semivariance behavior over an intermediate range. The second reason for considering these semivariograms inaccurate was due to semivariance computations based on an inadequate amount of data pairs per semivariance data point. Although an adequate number of data pairs are somewhat subjective, for this analysis ≥ 20 data pairs per semivariance point was deemed sufficient. The main objective for using ≥ 20 data pairs per point was to ensure an unbiased estimate of the semivariance, and so that the variance of estimation during kriging would be minimized.

Goovaerts [11], and previously the American Society of Civil Engineers [12], discussed calculation of the experimental variogram and fitting a model to the variogram. Two moments of Z(x) (with Z(x) representing the random function of contaminant densities) are required for a linear geostatistical analysis. The first-order moment is the mean of Z(x), and the second-order moment includes:

variance: $\mathrm{Var}\,[Z(x)] = E\,\{[Z(x) - m]2\} = C(0)$
covariance: $C(\mathbf{h}) = E\,\{[Z(x + \mathbf{h})]\,[Z(x)]\} - m2$
variogram: $2\gamma(\mathbf{h}) = E\,\{[Z(x + \mathbf{h}) - Z(x)]2\} = C(0) - C(\mathbf{h})$
where: $m = E[Z(x)]$ = mean or expected value
$\mathbf{h} = xi - xi{+}1$ (vector)
$\gamma(\mathbf{h})$ = variogram in the form used most often

The semivariogram is the principal tool used in geostatistics because it can be applied with less restrictive assumptions than either the variance or covariance [10]. The true semivariogram, called the experimental semivariogram, is unknown and is estimated by:

$$\gamma^*(\mathbf{h}) = [2N(\mathbf{h})] - 1 \sum_{i=1}^{N(\mathbf{h})} [z(xi+\mathbf{h}) - z(xi)]2$$

where: $N(\mathbf{h})$ = the number of sample pairs separated by the vector **h**.

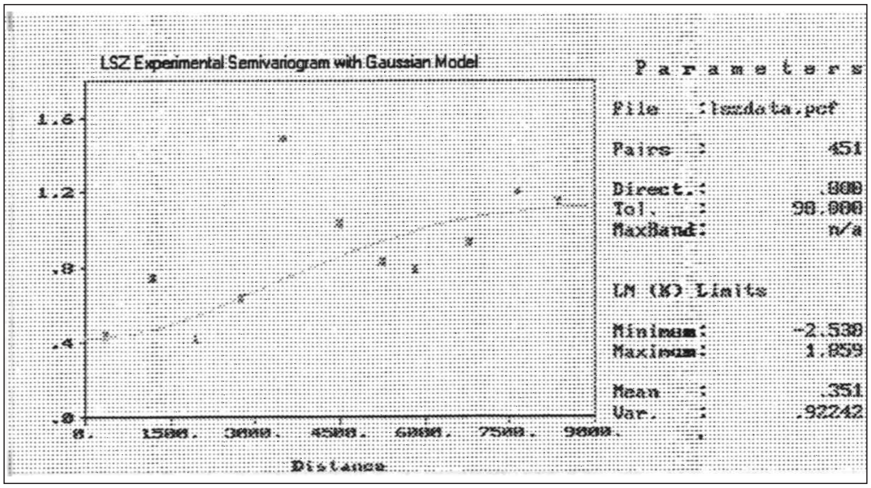
With semivariances computed and models fitted, kriging was performed next followed by cross-validation. The following sections describe each of these tasks. As before only the data describing the LSZ are included in this analysis.

Semivariance Results for the LSZ

Results from the experimental semivariograms constructed for the LSZ showed a predominately linearly increasing trend. When considering small inter-lag distances, the data display random behavior, however, when observed over the maximum lag distance of 8,000 ft, the linearly increasing trend was distinct. The LSZ experimental semivariograms were modeled with a natural-log transform gaussian model to provide a better fit to the original data. Figure 4 presents these semivariances with the gaussian model superimposed. The over and under semivariance points are essentially divided equally on either side of the model as the overriding trend of these data is supported by this model.

Kriging Results for the LSZ

Ordinary block centered kriging was performed to estimate hydraulic conductivities within CG037 for the LSZ. The CG037 area, which has a length of approximately 9,000 ft in the easting direction and 11,000 ft in the northing direction, was discritized into 8,200 blocks. Each block was constructed with dimensions of 111 ft by 111 ft.



```
Parameters
File     :lszdata.pcf
Pairs    :         451
Direct.  :        .000
Tol.     :      90.000
MaxBand  :         n/a

LN (X) Limits
Minimum  :      -2.530
Maximum  :       1.859
Mean     :        .351
Var.     :      .92242
```

Figure 4. Semivariogram of LSZ hydraulic conductivity from gaussian model.

Of the 39 LSZ wells that had hydraulic conductivity tests, only five were located within the boundaries of CG037. Because kriging is an estimation technique based on the optimal weighting of observations, data spatially located nearest to estimation points will receive the highest weight. Thus, the five slug tests performed within CG037 had a significant impact on the final kriging estimations. Additionally, it should be noted that the data collected within CG037 were located in the upper most part of the LSZ, and base wide, only four wells providing hydraulic conductivity data for the LLSZ were found to exist.

The results based on kriging of the LSZ based on a gaussian model fit to the natural-log transforms of hydraulic conductivity semivariance northern half was estimated to be less than 1.7 ft/day, while the southern region was interpolated to be less than 3.18 ft/day. Figure 5 presents the color-coded estimates of the kriging results for the LSZ based on the gaussian models fit to the natural-log transforms of hydraulic conductivity semivariance. By comparison, the conductivities used in the original numerical model exceeded 200 feet per day in a broad north-south orientation.

Cross-Validation for the LSZ

Cross-validation involves kriging estimates at specific well locations where known hydraulic conductivity values exist [13]. These kriging estimates are based on the hydraulic conductivity values of wells within a specified neighborhood surrounding a particular well, excluding the value of the well that is being

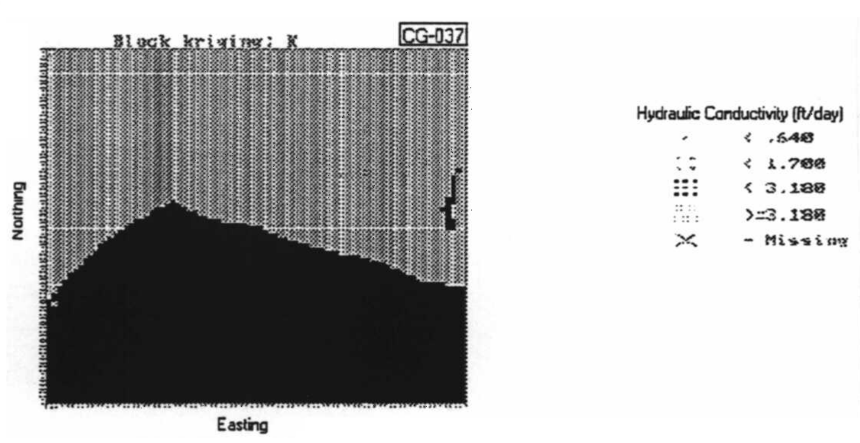


Figure 5. Estimates of hydraulic conductivity in LSZ from gaussian model.

estimated. Kriged estimates are then compared to the original measurements in order to quantify how well a particular model was able to reproduce the spatial variability of the measured locations. With this information, statistics and graphical depictions that conveniently summarize the important features of the kriged estimates can be obtained.

Cross-validation statistics for the gaussian model for the LSZ are presented in Table 2. These statistics, excluding the minimum and maximum, were fairly representative of the observed distribution. This table and Figure 5 show that kriging did fairly well in predicting the quartiles, and mean, but overestimated the minimum and underestimated the maximum, as is commonly the case.

Another useful way cross-validation can demonstrate how successfully kriging reproduced measured quantities is by plotting the estimated vs. observed values of the kriged variable. If kriging was capable of estimating the exact values of these measured quantities, this plot would produce a straight line with a slope of one-to-one. Figure 6 presents a plot of the cross-validation observed vs. kriged estimates for hydraulic conductivity in the USZ based on the gaussian model. We see that the gaussian model had the most difficulty estimating the upper end of the observed hydraulic conductivities.

Interpretation of this plot can be enhanced by recognizing that the x-value in any (x,y) pair shown on this graph would be the value of the measured hydraulic conductivity, and the y-value in any (x,y) pair would be the kriging estimated value. The box plots displayed above and to the right of each coordinate axis are used to graphically summarize the distribution of the observed and kriged data. The bottom and top of these boxes represent the 1st and 3rd quartiles, respectively. The median is shown as the horizontal (or vertical) line inside the box, and the mean is portrayed as the X inside the box. The overall range of data is shown by the lines that extend from each end of the box.

Table 2. Cross-Validation Results for the LSZ Gaussian Model

Kriging standard	Measured hydraulic conductivity	Kriged estimate
Minimum	.079	.43
25th percentile	.64	.83
Median	1.7	1.0
75th percentile	3.18	2.79
Maximum	6.42	3.306
N	39	39
Mean	2.01	1.94
Std. deviation	1.446	.964

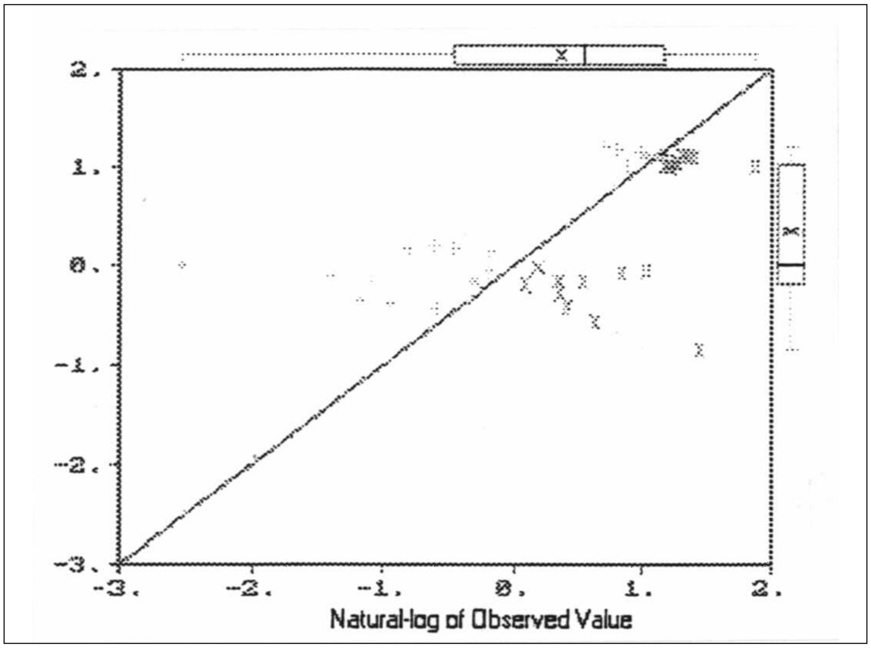


Figure 6. Cross-validation of kriged estimates from gaussian model.

Kriging Conclusions

Many features of what is currently known regarding hydraulic conductivity at Tinker AFB have become more clearly established by constructing the semivariogram and performing ordinary kriging. Hydraulic conductivity measurements were collected at only 19 wells within the CG037 model domain itself. Only five of these related to the LSZ. Across all of Tinker AFB only four wells that penetrated to depths of the LLSZ were found to have slug or pump test measurements. Although these data limitations are far from ideal, they did not prevent the successful application of geostatistical kriging. However, future data collection efforts centered within CG037 could greatly improve these estimations and provide greater resolution to many of the questions regarding the current model.

Kriging estimates for both the USZ and LSZ were heavily weighted by the relatively few hydraulic conductivity measurements within CG037, and overall showed a good deal of random behavior. This random behavior was shown by the kriging semivariograms and is indicative of the fact that the hydraulic conductivity measurements throughout the base are essentially the same. Data collection at intermediate distances however, may show stronger spatial continuity in near field measurements. Kriged estimates for the USZ ranged from less than approximately 6.5 ft/day to less than 1 ft/day. LSZ kriging showed values were no larger than 3.8 ft/day in the southern region and no larger than 1.7 ft/day in the northern portion of CG037. Based on the currently available data, kriging of CG037 has shown hydraulic conductivities that are representative of the entire base and are congruent with median values reported by the USGS for the Central Oklahoma aquifer.

Monte Carlo Simulation

Monte Carlo simulation was completed for the univariate statistical groupings and consisted of defining a time of travel to the Tinker AFB boundary from the areas of highest contaminant concentration. Hydraulic conductivities were randomly accessed as inputs to this time of travel equation. The results were pooled into a cumulative density function, which defines the probability associated with each travel time. The basic Monte Carlo algorithm follows:

If $z(\mathbf{u})$ represents the true concentration value at any node:
$z(\mathbf{u}) = g(\mathbf{X})$
where:
g = function representing the conditional simulation
$\mathbf{X}$ = vector of all simulation inputs [14].

Since the components of $\mathbf{X}$ contain the cumulative distribution function (CDF) $Fz(z)$ the goal of Monte Carlo analysis is to calculate the CDF $Fz(\mathbf{u})(zs(\mathbf{u}))$ given the probabilistic characterization of $\mathbf{X}$ [14]. $Fz(\mathbf{u})(zs(\mathbf{u}))$ is defined as:

$Fz(\mathbf{u})(zs(\mathbf{u})) = \text{Probability}\ (z(\mathbf{u}) \leq zs(\mathbf{u}))$
where:
$zs(\mathbf{u})$ = is the simulation output

Given a set of deterministic values for each of the input parameters, X1, X2, ..., Xn, Monte Carlo simulation computes the output value as:

$z(\mathbf{u}) = g(X1, X2, ..., Xn)$

Application of the Monte Carlo simulation procedure requires that at least one of the input variables, Xn, be uncertain and the uncertainty represented by a cumulative probability distribution. The simulation is then conducted numerous times to generate a series of $zs(\mathbf{u})$ values for each of the nodes within the two-dimensional grid simulated [15]. The simulated outputs are then statistically analyzed to yield the cumulative probability distribution of the simulated output. The steps involved in the application of the Monte Carlo technique include:

- Select the appropriate cumulative probability distribution function for describing uncertainty in the input variable(s).

- Select a random number from the distribution and use this as input to the model.
- Run the model using the random number taken from the input distribution to calculate the output.
- Repeat steps 2 and 3 for number (n) times.
- Determine the cumulative probability distribution function of the output step 3.
- Analyze the output distribution and utilize the statistics (i.e., mean and Upper Bound 95% Confidence Interval).

Stochastic Time of Travel Determinations

This analysis involved Monte Carlo simulation where the univariate description of the hydraulic conductivity (K) cumulative density function was randomly and repeatedly accessed as input for:

$$T_t = \frac{L\phi}{K(\partial h / \partial s)} R$$

where:
T_t = time of travel
L = flow length
ϕ = porosity
K = hydraulic conductivity
$\partial h/\partial s$ = hydraulic gradient
R = retardation factor

This approach contrasts to the numerical model employed by the original contractor where spatial variation in hydraulic conductivity was addressed by discretizing the groundwater domain under CG037 into 39,600 homogeneous blocks per layer. Differences from block to block were allowed, but a constant hydraulic conductivity within each block was assumed. Parenthetically, the approach that follows in Section 4 of this article ("Analytical Transport Modeling") assumes homogeneity across the entire section modeled. Numerical modeling is undoubtedly a superior way to construct a model where heterogeneities dominate the flow or transport domain. Difficulties arise, however, in that these models are highly data intensive, often requiring information not routinely collected. Although properly configured numerical models are capable of reducing "model uncertainty," they often suffer from "parameter uncertainty."

The approach selected for this phase of the study parallels that conducted at Love Canal, where the complex geology made numerical modeling tenuous [16]. Groundwater travel time in years to the Tinker AFB boundary is a function of the hydraulic conductivity, the flow length from the point of highest concentration to the boundary, and the hydraulic gradient. Contaminant travel time over the same distance is a function of the flow time, the effective porosity, and the retardation factor. Of these variables, only hydraulic conductivity was considered to be random. Porosity was fixed at 0.2, gradient at 0.003, and retardation at 1.0 (i.e., no sorption).

Each successive random access of a hydraulic conductivity from a previously defined statistical distribution results in a statistically feasible time of travel, when the preceding equation is applied. Once an adequate number of random hydraulic conductivities have been processed, a distribution of contaminant travel times can be generated. These distributions then define the probabilities associated with each travel time required. The determination of what defines an "adequate" sample size is frequently accomplished by specifying a tolerance error which is calculated from results generated. When the amount of difference between successive samples is less than this pre-established error, adequate sampling of the distribution has occurred and maximum precision has been reached. It must be added that this is a point of maximum precision, not accuracy. To achieve maximum accuracy, the underlying model must be correct. The model employed in this effort is Darcy's Law, rewritten for travel time. This assumes that flow and transport are through porous media.

Figure 7 presents the time of travel CDF for the LSZ. It shows that 90% of all simulated travel times were between 150 and 2770 years. Mean travel time is projected to be 794 years and there is a 99% chance that all travel times statistically possible with this data set will exceed 127 years. Given these

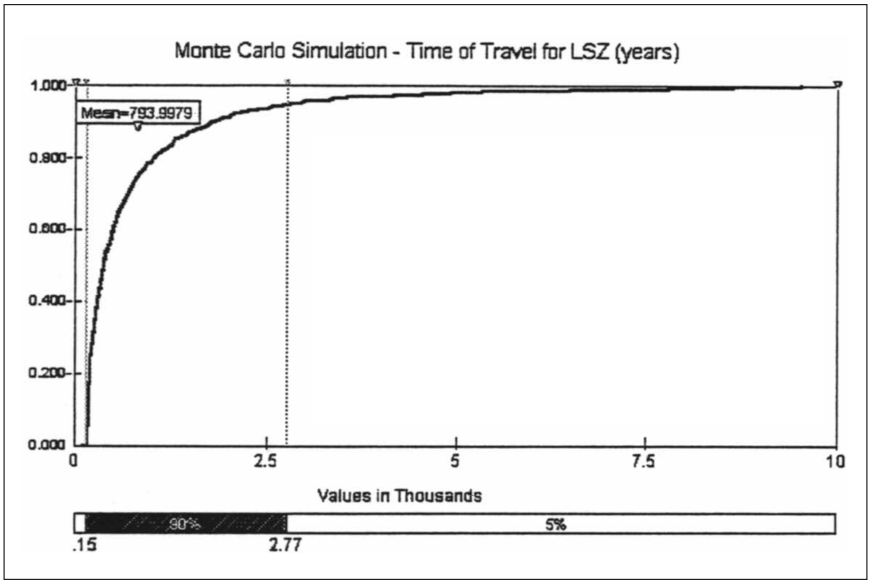


Figure 7. Stochastic time of travel determinations for LSZ.

analyses, there is virtually no chance that the contaminant plume could be found off base in only 30 years. This has implications for the environmental manager in that short-term regulatory response can be replaced with longer term environmental planning and remediation. Remediation options available to the environmental manager will also change dramatically with the different travel time estimates. Less intrusive methods such as natural attenuation can be considered for the smaller plumes resulting from the statistically defined hydraulic conductivity fields.

Analytical Transport Modeling

As a screening level comparison, a simple analytical transport model was prepared for the flow/transport path within the LSZ from the areas of highest contaminant concentration to the north property boundary of Tinker AFB. This area corresponded to that delineated in the original contractor report as being one where potential off-base excursion could occur. Recall that the earlier effort was based upon hydraulic conductivities of approximately 240 ft/day through the central corridor of this water bearing unit. Previously introduced univariate and bivariate statistical analyses do not support the use of this value. The transport model chosen for this application was the AT123D as incorporated in the American Petroleum Institutes (API) Decision Support System (DSS) [17, 18].

The basic equation used in this code that describes transport is:

$$\frac{\partial \mathrm{neC}}{\partial t} = \nabla\,(\mathrm{ne}\ D\ \nabla\ C) - \nabla\,(C\ \mathbf{q}) + M - (K\ \mathrm{ne}\ C) - (\lambda\ \mathrm{ne}\ C) - \frac{\partial(\rho b\ Cs)}{\partial t} + (\lambda\ \rho b\ Cs)$$

where:

$\mathbf{q}$ = Darcy velocity vector (L/T)
D = hydraulic dispersion coefficient tensor (L^2/T)
C = dissolved concentration of the solute (M/L^3)
Cs = absorbed concentration in the solid (M/M)
ρb = bulk density of the media (M/L^3)
M = rate of release of source ($M/(L^3{*}T)$)
n_e = effective porosity (L_o)
λ = first order decay constant (1/T)
K = degradation rate (1/T)

This fundamental advective-dispersion solute transport equation in three-dimensions can be simplified if the following assumptions are made:

- Groundwater characteristics are considered uniform.
- Sorption can be described with an instantaneous linear isothermal equilibrium.
- No waste flow exists across impervious boundaries.
- Flow through open boundaries occurs at infinity.
- There is a finite duration for contaminant release.

This equation can be reduced to:

$$C\,(x,y,z,t) = \int_{0}^{T} [M/(n_e Rd)]\ \mathrm{Fijk}\,(x,y,z,t;\tau)\, d\tau$$

where:
Fijk = integral of Green's function over the source space
M = instantaneous release of total mass
T = duration of waste release

When implemented, these assumptions allow these equations to be solved analytically rather than numerically. Application of the closed form of the integral is considerably easier than are the numerical forms as typified by MODFLOW. Disadvantages, however, include the inability to address heterogeneities within each simulation.

Input data for these models came from the reports developed by the original contractor. Variables such as the distance from the areas of contamination to Tinker boundaries as well as the size of these contaminated areas and the resultant plumes came directly from contractor reports. Transport variables such as partition coefficients and half-lives also came from these same reports. Whenever there was more than a single value presented, the most conservative choice was always selected. That is, the value that produced the most significant pollution event would be chosen. Similarly, the hydraulic conductivities utilized in this effort were those at the upper end of the distribution previously described. In this manner, the projected distance traveled could be considered to be a "worse-case" analysis. Table 3 presents the results from these simulations. These simulations support the findings of the previously presented geostatistical and stochastic time of travel efforts in that they show that the expected contaminant concentrations at locations approximately 600 feet from the presumed source after 30 years of transport would be in the hundredths of a part per billion range. That is,

Table 3. Analytical Contaminant Transport Modeling Results:
AT123D Code

Constituent	Model description	Maximum distance traveled in feet	Concentration at distance in mg/l
TCE	No biodecay/with sorption	600	2.65 (–6)
TCE	Biodecay and sorption	575	1.18 (–5)
Benzene	No biodecay/with sorption	660	1.85 (–8)
Benzene	Biodecay and sorption	575	1.54 (–9)

these concentrations would not only be far below every action level they would approach limits of detection.

Another model available from the U.S. Environmental Protection Agency, called BIOSCREEN [19], was also applied to this site [20]. Models with no biodecay and no sorption were prepared. These were considered to be the most conservative and should illustrate an approximate maximum distance that these plumes could travel at CG037. These simulations showed that a plume of concentration 3.7E-8 mg/l could travel approximately 800 feet for the given conditions. The distance from the location of maximum LSZ concentration through the center of the original contractor generated plume to the Tinker AFB boundary is approximately 4,200 feet. None of the analytical simulations performed for this current effort suggested contaminant migrations anywhere close to the base boundary. Given a most conservative hydraulic conductivity value consistent with all of the measurements found beneath Tinker AFB and assumptions of no sorption nor biodecay, the plume could only travel a bit more than one quarter of a mile from its most recently measured location. Given very conservative assumptions of sorption and biodecay, these plumes are projected to travel less than 200 yards in the next 30 years.

Figure 8 presents the plots for this no biodecay-no sorption model as well as a biodecay without sorption alternative. Also included in the legend for this figure is information for "instantaneous reaction" biodecay. This feature was not addressed for this effort and this portion of the legend can be ignored. The measured data taken from contours generated by the original contractor are also plotted on this figure to illustrate a level of calibration. Figures 9 and 10 illustrate these projected plume configurations.

These figures were generated for comparison purposes. Relative to this study, however, Figure 10 is more important in that it shows the maximum contaminant excursion distance for the conservative assumptions of no sorption nor biodecay.

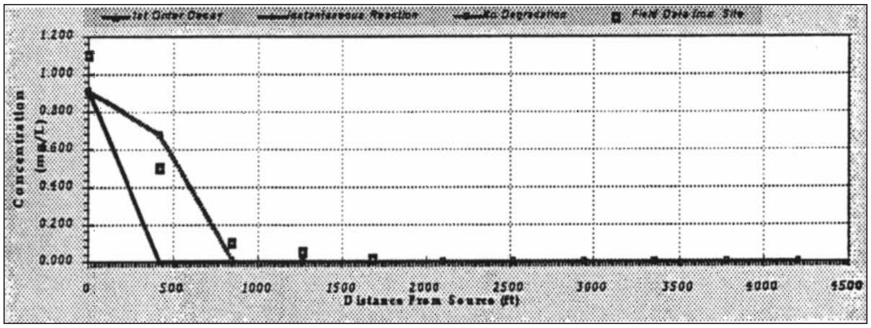


Figure 8. Plot of contaminant transport with biodecay and with no biodecay: both with no sorption. Original data included.

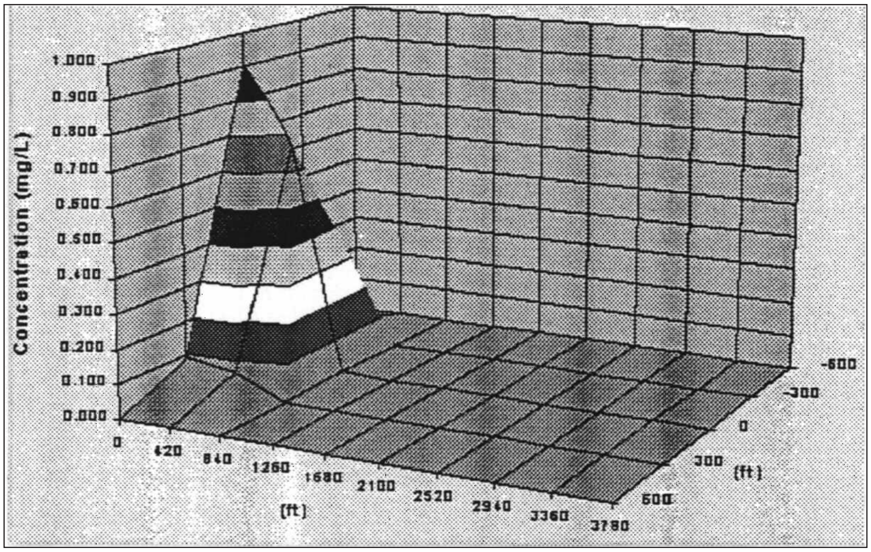


Figure 9. Representative TCE transport with no biodecay.

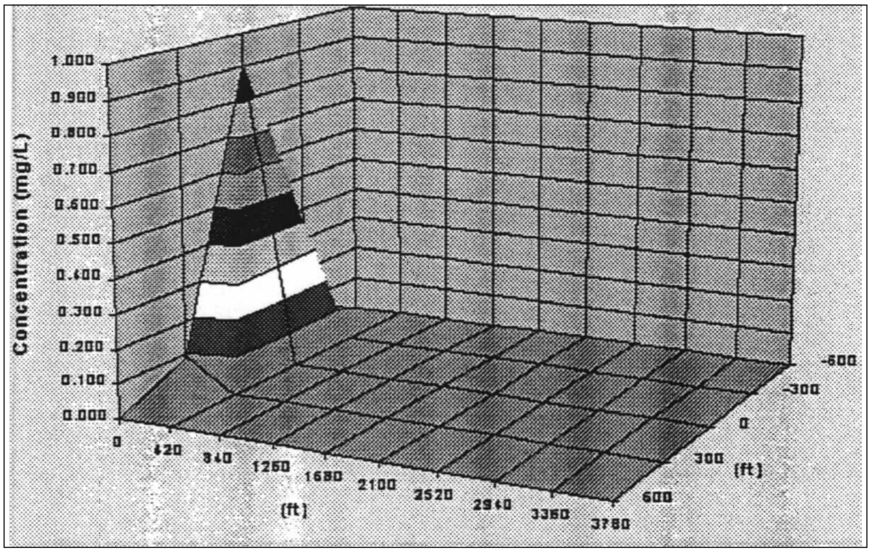


Figure 10. Representative TCE transport with biodecay.

These BIOSCREEN models support the AT123D analyses, which showed below detection limit concentrations occurring approximately 660 feet from the location of highest contaminant concentration after 30 years of travel time. Very conservative assumptions of biodecay reduce this excursion distance to between 420 and approximately 800 feet down-gradient of the location of the highest concentrations presented on the original contractor's maps.

Table 4 compares the plume sizes and approximated remediation costs for the plume generated by the numerical model with those determined by the analytical modeling presented in Table 3 and Figures 8, 9, and 10. Significant cost savings are projected to result from the smaller plume size generated by the analytical models and from alternative remediation options available for this smaller plume.

This analysis readily illustrates the differences in the two approaches. The more complex approach suffered from inconsistencies between model parameters and previously collected field measurements. The simpler approach utilized standard statistical methods when estimating similar transport parameters. The results from the simpler analytical model is more scientifically defensible and is more consistent with field data collection programs. Further, these results are more readily explainable to state and federal regulatory personnel. The natural attenuation remediation option that results from the smaller plume projection includes an extensive monitoring program. Should conditions arise that indicate that travel times to base boundaries are less than those predicted, these data programs will serve as an early warning system, allowing base environmental managers ample opportunity to implement protective measures.

SUMMARY AND CONCLUSIONS

An original modeling effort employing State of the Practice codes for groundwater flow and transport at the CG037 site on Tinker AFB projected excursions of TCE and Benzene off base within a 30-year period. These projections were

Table 4. Comparison of Projected Plume Size and Preliminary Remediation Cost Estimates

Technology	Affected area (ft^2) (approximate)	Estimate cost ($USD)
DUS	4,000,000[a]	4,160,000
Permanganate oxidation	4,000,000[a]	1,210,000
Natural attenuation	750,000[b]	155,000

[a] Projected plume size; not expected remediation area.
[b] Projected plume and remediation area.

strongly dependent upon the selection of the vertical and lateral hydraulic conductivity values used in the Lower Saturated Zone (LSZ) simulations. Specifically, conductivities of 30 and 244 ft/day respectively were required to calibrate the MODFLOW model. These values were determined by auto-calibration features within the code and were inconsistent with previous field measurements. The resulting models projected significant plume migration beyond base boundaries. Subsequent cost estimates to remediate the projected plume ranged from approximately one to over four million USD.

Given the inconsistencies between this model output and measured data, environmental managers at Tinker AFB requested a statistical evaluation of these findings. Geostatistical estimation of hydraulic conductivity based on structural analyses of the available hydraulic conductivity data set projected a range from less than 1 ft/day in the northern half of the study area to less than 3.1 ft/day in the south. This contradicted the original work, which utilized increasing hydraulic conductivities as the northern boundary of the AFB was approached.

A discussion of the data requirements for each modeling effort is warranted. Numerical modeling is highly data intensive. For the original effort, approximately 200,000 computational nodes were created to describe a five-layer aquifer system. Each node required extensive input data. While geostatistical modeling is also data intensive, in order to generate an appropriate semivariogram the level of information required is frequently orders of magnitude less than that needed for a complex numerical model. Similarly, the amount of data needed to complete the stochastic time of travel is less than that needed to construct the semivariogram and even less is needed for the analytical models. Figures 1 and 2 simultaneously underscore the minimal amount of data available to construct the original numerical model while providing a measure of confidence that sufficient data existed to complete the kriging, stochastic time of travel, and analytical modeling efforts. This strengthened the quality of these estimates.

Simulation techniques confirmed the observations made on the measured and estimated data. There is less than a 1% chance that the hydraulic conductivity of the LSZ is greater than 7 ft/day given the data reported in the original work and utilized in this study. Subsequent stochastic time of travel analyses and analytical transport modeling supported the data analyses and the simulation by showing that it was statistically unreasonable to expect off-base excursions of these contaminants in the next 30 years, given the data available for these determinations. The stochastic projections showed that there was less than a 1% chance that groundwater contaminants would migrate off base in less than 127 years and the analytical transport models projected travel distances of approximately 575 to 660 feet within the 30-year period.

The tools developed and the methods employed in this effort work together to indicate that the rate of advance for the plumes at CG037 is far slower than previously indicated by the complex numerical model. This means that Environmental Managers at Tinker AFB have significantly more time to develop properly configured remediation goals and plans rather than responding to short term alerts. Additionally, cost projections suggest that alternative, less costly remediation options may be available for the smaller, slower moving plume.

The methods developed for this study are considerably less complex than those of the linked numerical models originally employed. It is argued that given the level of information available and the complexity of the site, these methods are far superior to a more complex model with insufficient data. There is a tendency in environmental management to believe all outputs from computer models. The case presented here, however, serves to show that even the most complex codes and resulting model configurations must still be evaluated by common sense. Auto-calibration features of modern codes potentially eliminate much labor intensive effort. Even though the outputs from these calibration exercises are mathematically correct, they must be constantly evaluated to insure that they are also hydrogeologically correct.

REFERENCES

1. A. W. Harbaugh, *A Generalized Finite-Difference Formulation for the U.S. Geological Survey Modular Finite Difference Ground-Water Flow Model,* United States Geological Survey, Open-File Report 91-494, Reston, Virginia, 1992.
2. C. Zheng, *A Modular Three-Dimensional Transport Model for Simulation of Advection, Dispersion and Chemical Reactions of Contaminants in Groundwater Systems,* United States Environmental Protection Agency, Ada, Oklahoma, 1990.
3. S. Christenson and L. Carpenter, *Ground-Water Quality of the Central Oklahoma (Garber-Wellington) Aquifer,* proceedings of the United States Geological Survey, Open-File Report 92-116, February 20, 1992.
4. D. L. Parkhurst, S. Christenson, and G. N. Briet, *Ground-Water-Quality Assessment of the Central Oklahoma Aquifer, Oklahoma: Geochemical and Geohydrologic Investigations,* United States Geological Survey, Open-File Report 92-642, 1992.
5. J. S. Havens, *Geohydrology of the Alluvial and Terrace Deposits of the North Canadian River from Oklahoma City of Eufaula Lake, Central Oklahoma,* United States Geological Survey, Water-Resources Investigation Report 88-4234, 1989.
6. N. J. Gantos, T. Fox, N. J. Gupta, M. Kelley, and G. Yu, *Work Plan and Data Compilation for the Base-Wide Model at Tinker Air Force Base (AFB),* Contract No. DAAL03-91-C-0034, Battelle Memorial Institute, Columbus, Ohio, 1994.
7. Palisades Corporation, *@RISK. Risk Analysis and Simulation Add-In for Microsoft Excell,* Newfield, New York, 2000.
8. A. G. Journel and C. Huijbregts, *Mining Geostatistics,* Academic Press, London, 1978.
9. E. Englund and A. Sparks, *Geostatistical Environmental Assessment Software, User's Guide,* U.S. Environmental Protection Agency, EPA 600/8-91/008, Las Vegas, Nevada, 1991.
10. R. M. Cooper and J. D. Istok, Geostatistics Applied to Groundwater Contamination I: Methodology, *Journal of Environmental Engineering, 111*:2, 1988.
11. P. Goovaerts, *Geostatistics for Natural Resources Evaluation,* Oxford University Press, New York, 1997.

12. American Society of Civil Engineers, Review of Geostatistics in Geohydrology. I: Basic Concepts. II: Applications, *Journal of Hydraulic Engineering, 116*:5, pp. 612-658, 1988.
13. E. H. Isaaks and R. M. Srivastava, *An Introduction to Applied Geostatistics,* Oxford University Press, New York, 1989.
14. J. D. Dean, P. S. Huyakorn, A. S. Donigian, K. S. Voos, R. W. Schanz, Y. J. Meeks, and R. F. Carsel, *Risk of Unsaturated/Saturated Transport and Transformation of Chemical Concentrations (RUSTIC). Volume I: Theory and Code Verification,* U.S. Environmental Protection Agency, EPA 600/3-89/048a, Athens, Georgia, 1989.
15. L. Smith, R. A. Freeze, and J. Massmann, Geostatistical Approach to Site Characterization and Risk Assessment Related to Groundwater Contamination at Hazardous Waste Sites, in *Risk Assessment for Groundwater Pollution Control,* W. F. McTernan and E. Kaplan (eds.), American Society of Civil Engineers, New York, 1990.
16. J. W. Mercer, L. R. Silka, and C. R. Faust, Modeling Ground-Water Flow at Love Canal, New York, *Journal of Environmental Engineering, 109*:4, pp. 924-942.
17. G. T. Yeh, *AT123D: Analytical Transient One- Two- and Three-Dimensional Simulation of Waste Transport in the Aquifer System,* Environmental Sciences Division Publication No. 1439, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1981.
18. American Petroleum Institute, *Risk/Exposure Assessment Decision Support System (DSS),* Washington, D.C., 1994.
19. U.S. Environmental Protection Agencym, *Bioscreen: User Manual,* Subsurface Protection and Remediation Division, Ada, Oklahoma, 1997.
20. C. J. Newell, R. K. McLeod, and J. R. Gonzales, *BIOSCREEN Natural Attenuation Decision Support System: User's Manual Versions 1.3 and 1.4,* U.S. Environmental Protection Agency, Ada, Oklahoma, 1996.

Direct reprint requests to:

William F. McTernan
Oklahoma State University
207 Engineering South
Stillwater, OK 74078
e-mail: wmctern@okstate.edu