Search strategy for groundwater contaminant plume delineation

William A. McGrath and George F. Pinder
Department of Civil and Environmental Engineering, University of Vermont, Burlington, Vermont, USA

Received 5 August 2002; revised 9 April 2003; accepted 14 May 2003; published 23 October 2003.

[1] The definition of the boundary of a groundwater contaminant plume can be best viewed as a search strategy objective. To this end, a new approach to plume boundary identification based upon the goal of a stepwise reduction in the overall uncertainty in the location and geometry of the boundary can be used. Through the sequential selection of investigatory well locations based on the criteria of maximizing plume boundary uncertainty reduction and on updated field information, a least cost plume boundary investigation can be realized. The effectiveness of this approach can be investigated using a realistic hypothetical example and demonstrated via a field application.

INDEX TERMS:
1803 Hydrology: Anthropogenic effects; 1829 Hydrology: Groundwater hydrology; 1831 Hydrology: Groundwater quality; 1894 Hydrology: Instruments and techniques; KEYWORDS: search, contamination, groundwater, Kalman filter, Monte Carlo


1. Introduction

[2] While considerable effort has been expended in the pursuit of groundwater contaminant monitoring strategies [see Loaiciga et al., 1992], relatively little work has been dedicated to locating and defining the plume perimeter. In contrast to the monitoring problem, the definition of the geometry of the contaminant plume boundary can be best viewed as a problem in search strategy.

[3] Information pertaining to defining a plume boundary using search theory is scarce. Heuristic approaches have been described by Karatzas et al. [1992], McGrath and Pinder [1994] and Imhoff et al. [1988]. In addition, Borchers et al. [1997] described an approach to delineate the location of residual dense nonaqueous phase liquids which is marginally related. However, their methodology, while employing a random field strategy, is fundamentally different in goal and application to that which we are about to describe. In our paper we seek to formulate an approach to the plume boundary search problem that is based on search theory, a concept that becomes evident in the course of the following discourse.

2. Theoretical Development

2.1. Threshold Concentration

[4] To address the plume delineation problem one must a priori define what is meant by a plume boundary. Herein we denote a plume boundary via a threshold concentration level (TCL). Locations where this concentration is exceeded are assumed to reside within the plume perimeter. The engineering design goal is to cost effectively find the locus of points in three dimensions defining the TCL surface.

[5] In the definition of the plume boundary location:

\[ \delta_c(x) = \begin{cases} 1 & \text{if } E[c(x)] - Z_{\alpha_1}S_c(x) \leq TCL \leq E[c(x)] + Z_{\alpha_2}S_c(x) \\ 0 & \text{otherwise} \end{cases} \]  

where \( E[c(x)] \) is the expectation of the contaminant concentration, \( S_c(x) \) is the standard deviation of this concentration, and \( Z_{\alpha_1} \) and \( Z_{\alpha_2} \) are the standardized z-scores for lower and upper confidence limits in \( \alpha_1 \) and \( \alpha_2 \) respectively, the moments of concentration are used. In practice, and in the field application to follow, log-concentrations are used because concentration and its error typically follow a lognormal distribution. Despite this, the theoretical development herein will continue to use concentration space for simplicity of presentation.

[6] Equation (1) states that the delta function \( \delta_c(x) \) is unity when the location \( x \) is within a band defined by the confidence interval enclosing the location of the threshold concentration level, TCL. Outside of this band \( \delta_c(x) \) vanishes. The size of this confidence interval, which is, in fact, a volume in three space dimensions, can be stated as

\[ U^0 = \int_\Omega \delta_c(x) \, d\Omega \]  

where \( \Omega \) is the entire domain of interest and \( U^0 \) represents that portion of \( \Omega \) wherein the concentrations identified with the TCL are going to occur with probability \( \alpha \). We will denote this as the confidence volume, \( CV \). In fact in a two spatial dimensional problem, \( U^0 \) will be an area.

[7] The concept presented in equations (1) and (2) is illustrated in Figure 1. In this areal, two-dimensional example groundwater flows from west to east in a homogenous aquifer. The north and south boundaries of this model are specified as no fluid or dispersive mass flux. The head on the west side is 5.0 m and on the east side 1.0 m. Constant...
concentration values of zero are specified on the west and east sides of the model. The concentration at the source is set to be \( c/c_0 = 1 \) where \( c_0 \) is the reference concentration. The model has a north-south dimension of 100.0 m and an east-west dimension of 400.0 m. In this example, the isotropic and homogeneous hydraulic conductivity is 0.01 m/sec, the porosity is 0.25, the longitudinal dispersivity is 0.05 m and the transverse is 0.005 m. The time of observation is 104.1 hours after introduction of the contaminant at the source.

The thick line found in Figure 1a is the outer boundary of the 95 percent confidence interval. The analogous line in Figure 1b represents the inner boundary of the 95% confidence interval. The area between these two curves represented by Figure 1c is the area \( U^0 \) defined in equation (2).

**2.2. Worth of Sample Data Function (WSD)**

The goal of our research is to reduce the volume \( U^0 \) as much as possible with each addition of geohydrological or geochemical information. The function that measures the degree to which this area reduction takes place with the addition of water quality information we define as the worth of sample data function \( WSD_{c_i} \), i.e.,

\[
WSD_{c_i}(x) = \frac{U^0 - U^{c_i}(x)}{U^0} \tag{3}
\]

where \( U^{c_i}(x) \) is the CV after the information from a water quality sample \( c(x_i) \) is added to the data set. If there are \( M \) possible sampling locations, clearly there will be \( M \) values for the function \( WSD_{c_i}(x) \), one for each of the \( M \) sampling locations. Upon calculation, these values can then be contoured to show the relative worth of additional data for any location within the area of interest \( \Omega \) found in equation (2) and represented by the area of \( 400 \times 100 = 4.00 \times 10^4 \) m² in Figure 1. Selection of the best well location is then easily achieved via the \( WSD_{c_i} \) surface in combination with practical engineering considerations regarding well site accessibility.

The \( WSD_{c_i} \) surface for the measurement of concentration is presented in Figure 2a. Figure 2b illustrates the \( WSD_{c_i} \) which is appropriate for the selection of a location.
for the measurement of hydraulic conductivity. It is noted that the asymmetry of Figure 2 about the longitudinal axis is a function of the numerical calculation of these fields.

The addition of information on either concentration or hydraulic conductivity will impact the \( WSD_c \) surface. By adding data points sequentially with each new point augmenting the existing data set, an interactive search algorithm is generated.

Concentration measurements contain errors which we wish to incorporate in our analysis. In this spirit, when samples are extracted from the reference plume (the plume from which concentration data would be obtained in the field), they are perturbed with an error. The method to perturb the concentration drawn from the reference plume to obtain observed concentrations is to draw a random number from the standard normal distribution \( N[0, 1] \) and scale it by the standard deviation in the measurement error, that is

\[
e = r \sqrt{R_c}
\]

where \( e \) is the random error, \( R_c \) is the measurement error variance, and \( r \) is the random number drawn from the standard normal distribution. This error is added to the concentration of the reference plume to obtain the observed concentration. If the perturbed observations are negative, they are truncated to a value of zero.

Figure 3 illustrates this process using concentration data. The first sample is selected at the point \( x_1 \), which is at one of the two most beneficial locations according to Figure 2. The information collected from point \( x_1 \) is used to modify the respective \( WSD_c \) function as seen through a comparison of Figure 2a and Figure 3a. The same procedure is used to select the second and third sampling locations as illustrated in Figures 3b and 3c. Notice in the sequence of panels in Figure 3 how the \( WSD \) function responds to selection of each additional sampling location, i.e., locations \( x_1, x_2, \) and \( x_3 \).

2.3. Monte Carlo Simulation

While the above protocol is deceptively simple in concept, understanding its implementation requires additional background. In this spirit, consider generation of the confidence volume CV shown in Figure 1. The determination of this initial volume \( U^o \) is achieved by calculating a random field of concentration values using a Monte Carlo procedure with samples drawn from a random field of hydraulic conductivity.

Assuming a logarithmic distribution of hydraulic conductivity and adding correlated random perturbations to the log conductivity field by the method of sequential
Gaussian simulations [see Deutsch and Journel, 1992] one can obtain the appropriate random field for hydraulic conductivity for this analysis. In the above example the correlation structure of the random field is assumed to have an exponential variogram with a correlation length of 30 m, a variance of 1.0 m/sec, and a nugget effect of zero. The analysis was done using five hundred Monte Carlo simulations, a number deemed adequate based upon an analysis of the trace of the covariance matrix versus the number of simulations.

The Monte Carlo simulations provide the realizations from which one can calculate the confidence intervals associated with the WSD analyses. By examining the concentrations associated with all realizations at a given position \( x_i \), one can calculate a standard deviation of the concentration at that point, that is \( S_c(x_i) \). Combining information from each nodal location in the finite element representation of the contaminant transport model associated with our example, one can obtain, with the help of equation (1), the information provided in Figure 1. Equation (2) can now be used with equation (3) to create the worth of data function provided in Figure 2.

2.4. Kalman Filter

Updating the worth of data function with new concentration information requires a different strategy than that required for new hydraulic-conductivity data. Consider first the use of the Kalman filter to condition modeled expectations and covariances using measurements of concentration [see Zou and Parr, 1995]. In essence the Kalman filter is used to update the prior estimate \( \tilde{c}^* \) and its covariance \( \tilde{C} \) using the measured observations \( c_{\text{obs}} \) whose measurement error covariance is \( R \). Because the Kalman filter updating technique is documented in the literature it is not presented here. Details of the technique are given by Zou and Parr [1995], and its application to this particular problem are given by McGrath [1997].

2.5. Concentration Field Update

When hydraulic conductivity is the measured quantity, the methodology used to update the state variable, in this case concentration, is called first-order uncertainty analysis. First introduced in the groundwater literature by Dettinger and Wilson [1981], the approach relates the vector of state variables, say the concentration, with the random parameters, in our case the hydraulic conductivity. We formally express this relationship as:

\[
\mathbf{c} = \mathbf{c}(\mathbf{K}),
\]

Expansion of \( \mathbf{c}(\mathbf{K}) \) as a first order Taylor series yields:

\[
\mathbf{c}(\mathbf{K}) \approx \mathbf{c}(\mathbf{K}) + \frac{\partial \mathbf{c}}{\partial \mathbf{K}} (\mathbf{K} - \mathbf{K})
\]
where $\mathbf{K}$ is the expected value of the hydraulic conductivity, i.e., $\mathbf{K} = E[\mathbf{K}]$ and $\mathbf{e}(\mathbf{K}) = \mathbf{c} = E[\mathbf{e}]$ is the expected value of the concentration. The function $\frac{\partial \mathbf{e}}{\partial \mathbf{K}}$ is the sensitivity of the concentration with respect to the hydraulic conductivity and is a matrix quantity.

Consider now the following approximate representation of the covariance of the concentration, i.e.,

$$
\mathbf{C}_{cc} = E[(\mathbf{c} - \mathbf{\bar{c}})(\mathbf{c} - \mathbf{\bar{c}})^T]
$$

$$
\approx E\left[\mathbf{e} + \frac{\partial \mathbf{e}}{\partial \mathbf{K}} (\mathbf{K} - \mathbf{\bar{K}}) - \mathbf{\bar{c}}\right] \left[\left[\mathbf{e} + \frac{\partial \mathbf{e}}{\partial \mathbf{K}} (\mathbf{K} - \mathbf{\bar{K}}) - \mathbf{\bar{c}}\right]^T \right]
$$

$$
\approx \frac{\partial \mathbf{e}}{\partial \mathbf{K}} E\left[\left(\mathbf{K} - \mathbf{\bar{K}}\right)\left(\mathbf{K} - \mathbf{\bar{K}}\right)^T\right] \frac{\partial \mathbf{e}}{\partial \mathbf{K}}^T
$$

$$
\approx \frac{\partial \mathbf{e}}{\partial \mathbf{K}} \mathbf{C}_{KK} \frac{\partial \mathbf{e}}{\partial \mathbf{K}}^T
$$

(7)

where $\mathbf{C}_{KK}$ is the covariance of the hydraulic conductivity. Equation (7) demonstrates that, to first order, the uncertainty in state variables depends upon the uncertainty in the parameters and on the sensitivity of the states to changes in the parameters.

An alternative strategy for obtaining $\mathbf{C}_{cc}$ is based upon the use of the cross-covariance between the parameters, i.e., hydraulic conductivity, and the states, i.e., the concentration. From the definition of the covariance we have:

$$
\mathbf{C}_{cc} \approx E\left[\left[\mathbf{c} - \mathbf{\bar{c}}\right]\left[\left(\mathbf{K} - \mathbf{\bar{K}}\right)\right]^T\right]
$$

$$
\approx \frac{\partial \mathbf{e}}{\partial \mathbf{K}} E\left[\left(\mathbf{K} - \mathbf{\bar{K}}\right)^T\right] \frac{\partial \mathbf{e}}{\partial \mathbf{K}}^T
$$

$$
\approx \frac{\partial \mathbf{e}}{\partial \mathbf{K}} \mathbf{C}_{KK} \frac{\partial \mathbf{e}}{\partial \mathbf{K}}^T
$$

(8)

Equation (8) is equivalent to the first two terms of the right hand side of equation (7), which allows one to rewrite equation (7) as:

$$
\mathbf{C}_{cc} \approx \mathbf{C}_{cc} \frac{\partial \mathbf{e}}{\partial \mathbf{K}} \frac{\partial \mathbf{e}}{\partial \mathbf{K}}^T
$$

(9)

which, in combination with equation (8) yields:

$$
\frac{\partial \mathbf{e}}{\partial \mathbf{K}} \approx \mathbf{C}_{cc} \mathbf{C}_{KK}^{-1}
$$

(10)

From equations (9) and (10) we have:

$$
\mathbf{C}_{cc} \approx \mathbf{C}_{cc} \left(\mathbf{C}_{cc} \mathbf{C}_{KK}^{-1}\right)^T
$$

(11)

which is an alternative representation of the concentration covariance matrix.

### 2.6. Sampling Strategy

It is important to note that the confidence interval for the plume boundary depends both on the expected value of the concentration and on its variance. Thus, in showing the impact of sample collection on the plume boundary one must assume concentration information is obtained at each sampling event. In the case of the variance of the concentration, however, this is not the case. A hypothetical sampling location will yield a reduction in the uncertainty at that point to that of the sampling uncertainty, irrespective of the concentration of contaminant present at that point. Consequently, the impact of sampling on the covariance of concentration can be determined in the absence of actual samples being collected.

The impact of collecting a measurement of concentration on the location of the plume boundary is illustrated in Figure 4. To provide this information a target plume had to be defined. Since no field data actually existed for this example, as a surrogate for the real plume one was selected at random from the 500 realizations available from the Monte Carlo analysis. This plume is provided in Figure 5. Had another plume been selected, the evolution of the estimated plume topology as the plume was updated would have been different.

In examining Figure 5, it is helpful to refer back to Figure 2 which illustrates the worth of sample data function for this example. Note that the sample location selected for the first water quality sample is consistent with the recommendations for Figure 5.

The location of the target plume superimposed on the computed confidence interval after four samples have been selected is shown in Figure 6. The dashed curves represent the upper and lower confidence intervals for the plume boundary that is indicated by the thick line surrounding the shaded area. As is evident, the true boundary is contained within the confidence interval area. In Figures 7 and 8 one observes the impact of increasing and decreasing the sampling error, respectively. A summary of the strategy to be used in the search algorithm is provided in Figure 9.

### 3. Field Application

The field site chosen for this demonstration is site 73 on the Pease Air Force Base in Portsmouth, New Hampshire. The site is located near the center of a peninsula on New Hampshire’s coast, which is bounded to the north and east by the Piscataqua River and to the west by Little Bay and Great Bay. Site 73 was characterized in September 1996 by The Johnson Company, Inc. of Montpelier, Vermont, under contract with the Department of the Air Force. The results of that study are reported by Pitkin [1996]. Site 73 was operated as a liquid oxygen plant from 1959 to 1978 and it is believed that these operations involved the use of trichloroethene (TCE). During the removal of two underground petroleum storage tanks between 1989 and 1991, contamination of soils by petroleum hydrocarbons, as well as TCE and its degradation products cis-1,2-dichloroethene (DCE), and vinyl chloride (VC), were observed. This led to an investigation into the extent of the hydrocarbon contamination in 1992 and 1993 [Roy F. Weston, Inc., 1993].

The investigation consisted of soil borings and soil sampling at 26 locations, HydroPunch groundwater sampling at 25 locations and an additional investigation was performed in 1996 by The Johnson Company. This case study focuses on the 1996 investigation. Unconsolidated deposits underlie site 73 and consist of fine to very fine sand from 30 to 60 feet.
A thin layer of poorly sorted till is located below the sand and above a highly transmissive saprolite. The saprolite is underlain by competent phyllite bedrock. The water table occurs at depths of 5 to 10 feet below ground surface. The upper portion of the unconsolidated deposits are characterized by hydraulic conductivities less than one foot per day, while the deeper sediments have hydraulic conductivities on the order of 20 feet per day. The horizontal components of the hydraulic gradients range from 0.004 in the upper aquifer to 0.012 in the lower portion of the aquifer. The vertical component of the hydraulic gradient is downward from the unconsolidated deposits to bedrock at approximately 0.015 [Roy F. Weston, Inc., 1994]. It is presumed that these vertical gradients result from a combination of surface recharge and exfiltration through fractured bedrock underlying the aquifer.

The sampling program was undertaken in two steps. In step one, 13 locations along three transects oriented approximately normal to the suspected principle axis of the plume were identified a priori (see Figure 10). At each of these locations, samples of groundwater quality were collected from the water table to refusal at a vertical spacing of 1.5 feet.

The objective of the second phase was to locate the leading, or down-gradient, edge of the plume. Ten additional locations were profiled, primarily along three transects. As in the first stage the vertical spacing between samples was 1.5 feet. In this instance sampling began approximately 360 feet down-gradient from the source area.

For this demonstration, all 13 profiles from the first phase and two of the profiles from the second phase are used to create a concealed data set. The remainder of the locations profiled in the second phase are withheld from the data set because they contained chlorinated solvents that come from an unknown source located to the east of this study area. The compounds from this unknown source are identical to those emanating from site 73. Thus the data collected at these profiles can not be used to assess the contamination that has come from site 73. The data, as it pertains to site 73 is therefore not useful.

Contaminant concentrations from each of the 15 profiles that comprise the concealed data set for this demonstration are reported by Pitkin [1996] These locations are provided in Figure 10. The primary contaminant in the plume was originally trichloroethene (TCE). However, this has been largely degraded to cis-1,2-dichloroethene (DCE).
and vinyl chloride (VC). This degradation is the result of bacterial cometabolism resulting from the degradation of petroleum hydrocarbons detected within the plume. Because this example application is not concerned with the mechanics of the degradation process but rather with identifying optimal locations to collect groundwater quality samples to delineate the chlorinated solvent plume, the multiconstituent plume is expressed as a single-constituent plume.

**Figure 5.** Reference realization. (a) Hydraulic head. (b) The contaminant plume such that the outer edge of the filled area is the boundary whose location is sought (the contour with a concentration of 0.1, our TCL for this example).

**Figure 6.** Reference plume superimposed upon the TCL contours for the upper and lower confidence limits in concentration after four samples have been taken with $R_{l}I = 1.0 \times 10^{-4}$. 
Therefore, when DCE and VC are encountered, their measured values are converted to equivalences of TCE, and all three constituents are summed together to obtain a single, representative measurement of the original chlorinated solvent. The plume illustrated in Figure 10 is the original TCE plume identified in the field.

The decision space for this demonstration is coincident with the nodes of the finite element mesh for the rudimentary flow and transport model. The mesh for site 73 is displayed in Figure 11. The locations of the profiles reported by The Johnson Company are presented in Figure 10 for reference.

This mesh was constructed using Argus ONE (Argus Holdings Ltd., 1994). The boundaries are specified so that the regional groundwater flow patterns around site 73 are captured. This includes no-flow boundaries on the east, west and southern edges which are perpendicular to the hydraulic gradients measured by Roy F. Weston, Inc. [1994]. The mesh is also large enough so that groundwater flow patterns in the contaminated region of site 73 are not influenced by the boundaries when random hydraulic conductivity fields are presented to the rudimentary model to conduct Monte Carlo simulations. Finally, the mesh is locally refined around site 73 so that sharp fronts can be captured when transport simulations are conducted.

The random parameters used in this demonstration to conduct Monte Carlo simulations for evaluating the plume boundary uncertainty, \( U_{BL} \), are the hydraulic conductivity, \( K(x) \) (20 ft/day) and the infiltration flux rate, \( q_z \) (−0.003 ft³/(ft² × day)). These are used because they define the flow path that the unknown plume follows. They thus have the most influence upon the plume boundary location. Both of these parameters are assumed to be lognormally distributed (i.e., their logarithms are normally distributed) with variances of one and expected values of \( K(x) = 20 \) ft/day and \( q_z = -0.003 \) ft³/(ft² × day). Distributed values of log conductivity are assumed to be correlated in space with a statistical structure that follows an exponential variogram with a variance of 1.0 ft²/day² and a correlation length of 250 feet.

Realizations of the infiltration flux rate are obtained by adding a zero-mean, unit variance, normally distributed \( \mathcal{N}(0,1) \) random perturbation, \( \delta \) (ln \( q_z \)), to the logarithm of the expected flux,

\[
q_i = \exp[\ln q_i + \delta(\ln q_i)], \quad i = 1, n_{sim}
\]
No correlation is assumed between the random realizations of hydraulic conductivity and the net infiltration flux rate.

While we have considered only the hydraulic conductivity and infiltration flux rate as random parameters, the approach can be extended to consider random boundary conditions and has been used to accommodate random source magnitude and strength in the work by McGrath [1997].

A total of 250 random parameter sets are computed. These are entered into the flow and transport model to simulate 250 chlorinated solvent plumes. This number of realizations is simulated based upon an analysis of the convergence in nodal concentration variances. These statistical moments of log-concentrations are used to quantify the uncertainty of the plume boundary location, \( U_{BL} \). Log-concentration analysis space is used in this demonstration because of the large disparity in the concentrations. These range from the source concentration \( (C_0 = 50,000 \, \mu\text{g/L}) \) to the TCL (\( TCL = 10 \, \mu\text{g/L} \)) which are four orders of magnitude apart.

The locations of the TCL contours from the upper and lower limits of the statistical confidence interval in the log-concentrations of the chlorinated solvent are shown in Figure 10. The area of space in which the plume boundary is expected to reside is shaded grey and is the area in-between these two contours. At this point in the investigation, the plume boundary location is estimated in terms of the region of space where it is expected to reside. This estimate encompasses the understanding, and the uncertainty in the regional groundwater flow patterns, as well as the professional judgment that was used in constructing the model. The uncertainty in the plume boundary location is large, but it is constrained to the region bounded by the two contours in Figure 10.

The worth of sample data function, \( WSD_c(x) \), is now calculated. Examination of the resulting surface indicates that of the 15 potential well locations, the location marked as one in Figure 12 is optimal. When the data from profile one, which was heretofore hidden, is revealed to provide the equivalent of a field exploration data measurement, it is found to contain a maximum averaged chlorinated solvent concentration of 3201 \( \mu\text{g/L} \) as TCE. As in the earlier theoretical examples, this value is now considered as input into the data set and the investigation is continued.

Having incorporated the “measured” data, the worth of sample data, \( WSD_c(x) \), is used to identify the second profile location. Location 2 in Figure 12 is selected. It is interesting to note that the \( WSD_c(x) \) function is negligible along the main axis of the suspected plume up-gradient of profile one. This arises because profile one has identified concentrations of chlorinated solvents above the TCL. This implies that all locations in-between it and the source are also contaminated. Sampling in this region will therefore not decrease the size of the plume boundary area and have no utility for this investigation.

The worth of sample data function, \( WSD_c(x) \), is now recomputed and is the result presented in Figure 13.
The best location for selecting the next sample is indicated by the 0.150 contour on the west side of the project area. As expected, the superior candidate here is the runner-up from the previous analysis. The data collected at profile location three is now revealed to this analysis. It also contained no TCE contamination (i.e., 0.0 μg/L). The limits of the plume boundary estimate after these three profiles have been assimilated into the algorithm are shown in Figure 12.

The resulting worth of sample data function, $WSD_y(x)$, was used to identify a fourth profile location. Unfortunately, there is no hidden data to reveal at that location. This is because the profiles conducted in that region during phase two of The Johnson Company’s investigation contained cross-contamination arising from another indistinguishable source of chlorinated solvents. Thus these data have been omitted from the concealed data set and the analysis terminated.

The remainder of the concealed data set is now revealed to the algorithm and the final plume boundary results are assessed. At this point, detailed information is available to estimate the plume boundary location and the distributed chlorinated solvent concentrations. The resulting confidence area is displayed in Figure 14.

A comparison of Figures 10, 12, 14, and 15 reveals that the greatest return on investment in terms of identifying the boundary of the contaminant plume is with the first three samples. The difference in the area associated with the boundary estimate after these three profiles have been assimilated into the algorithm are shown in Figure 12.

The triangular finite element mesh of site 73 containing 1493 nodes and 2919 elements. No-flow boundaries for the rudimentary groundwater flow and transport model are displayed along with northing and easting coordinates. Profile locations are shown as squares for reference.

Figure 12. Estimate of area containing the plume perimeter after three samples have been taken. The data profile locations that were selected from the hidden sample set by the algorithm are indicated in the sequence in which they were selected by the numbers assigned to them.
initial plume boundary location estimate and that determined using three samples is 33,965 ft². However the additional improvement achieved using all the samples was only 4,878 ft².

[47] One also sees that the TCE plume as estimated by the groundwater professional is smaller than predicted by the confidence interval. In addition to the realization that the true plume remains unknown, the reader is reminded that the analysis done here is on TCE and its daughter products and the plume presented as the target is for TCE alone.

4. Conclusions

[48] The definition of the boundary of a contaminant plume when identified with a specified threshold concentration level can be effectively identified through the use of search theory. Using groundwater flow and transport models formulated for random parameter fields (in this example hydraulic conductivity), in combination with the linear Kalman filter and sensitivity theory, an interactive computer assisted technique has been developed. On the basis of the utilization of these mathematical tools a worth of data function is calculated that provides a visual representation of the benefits attached to sampling either hydraulic conductivity or water quality in a specified location. The worth of sample data surfaces provide effective guidance for the selection of aquifer test and water quality sampling locations.

[49] The proposed approach calculates the location of the 95% confidence interval for a target contaminant plume and, in our sample problem, the true plume is found to reside within the indicated confidence interval. The error identified with the sampling protocol, including the effects of laboratory uncertainty, influence the geometry of the

Figure 13. WSD function contours obtained after two samples have been selected.

Figure 14. Estimate of area containing the plume perimeter after all 15 hidden samples are used in the algorithm. Samples have been taken.

Figure 15. Confidence area $U_{BL}$ versus number of samples taken. Numbers indicate confidence area in square feet.
confidence interval; large sampling errors give rise to confidence intervals with larger errors.

Application of the procedure to a field situation shows that relatively few samples were required to effectively define the plume boundary. While the first three samples reduced the area of the confidence interval by 48%, 12 additional samples reduced the area by only three percent.

While no detailed record was kept of the time required to execute the analyses presented above, it was on the order of less than a minute using current generation personal computers. In general the use of Monte Carlo simulation and Kalman filter prediction requires hundreds of groundwater flow and transport simulations per optimal search design. Thus the efficiency of the approach depends upon the computational speed of the flow and transport algorithm. In this instance a very efficient two-dimensional finite element model is used that, through a splitting algorithm is readily extended to three dimensions. The increase in computational effort increases approximately linearly with the number of layers used in the vertical dimension using this model [see Pinder, 2002].

Acknowledgments. The authors wish to acknowledge the financial support of the Department of Defense, U.S. Army Corps of Engineers, via contract DACA39-95-K-0001. We also wish to thank David E. Dougherty for his assistance in the formulation and preparation of this manuscript.

References

W. A. McGrath and G. F. Pinder, Department of Civil and Environmental Engineering, University of Vermont, Burlington, VT 05405, USA. (pinder@emba.uvm.edu)