Stochastic Inverse Modeling for Capture Zone Analysis

N. L. Jones[1], J. I. Green[2], and J. R. Walker[3]

Abstract

In this paper, we describe an approach to probabilistic capture zone analysis that utilizes a stochastic inverse modeling method. Indicator Kriging or transition probability geostatistics is used to generate N equal probability realizations of indicator distributions where each realization is conditioned to a set of boreholes or scatter points. Aquifer properties are then inherited from the indicators, thus resulting in N different MODFLOW 2000 flow models. Each of these flow models is subsequently run in inverse mode to optimize the aquifer properties with respect to observation well and flow data. The end result is N flow solutions, each of which has an associated global error norm indicating goodness of fit between the computed and observed values for the optimal set of parameters. After creating a set of flow fields, we can create probabilistic capture zones using simple advective particle tracking. The contribution to the probabilistic capture zone from each of the N model runs is weighted by the error norm for that particular run.

Introduction

Most attempts at probabilistic capture zone delineation utilize a simple Monte Carlo approach to generate the multiple flow fields. The modeler typically generates a calibrated flow model and the stochastic runs are made by perturbing the parameters from a mean value defined by the optimized values from the parameter estimation process. One of the problems with this approach is that it may give too much weight to the optimal set of parameter values and it does not include the uncertainty associated with the aquifer zonation used to define the calibrated flow model.

[1] Environmental Modeling Research Laboratory, 242 Clyde Building, Brigham Young University, Provo, Utah 84604; PH (801) 422-2812; e-mail: njones@byu.edu

[2] Environmental Modeling Research Laboratory, 242 Clyde Building, Brigham Young University, Provo, Utah 84604; PH (801) 422-2812; e-mail: green@byu.edu

[3] Environmental Modeling Environmental Research Laboratory, 242 Clyde Building, Brigham Young University, Provo, Utah 84604; PH (801) 422-5846; e-mail: jawalker@byu.edu

Furthermore, many of the model instances generated during the parameter estimation process may result in a poor fit with the field observation data.

In this paper, we describe an approach to probabilistic capture zone analysis that utilizes a stochastic inverse modeling method. MODFLOW 2000 is used to develop the flow fields, and MODPATH particle tracking yields the capture zones. Indicator Kriging or transition probability geostatistics is used to generate N equal probability realizations of indicator distributions where each realization is conditioned to a set of boreholes or scatter points. Aquifer properties are then inherited from the indicators, resulting in N different MODFLOW 2000 flow models. Each of these flow models is subsequently run in parameter estimation mode to optimize the aquifer properties with respect to observation well and flow data. The end result is N flow solutions, each of which has an associated global error norm indicating goodness of fit between the computed and observed values for the optimal set of parameters.

After creating a set of flow fields, we can create probabilistic capture zones using simple advective particle tracking. Multiple approaches can be used to generate a 2D capture probability map from a 3D capture zone. A weighting factor can also be applied to the contribution from each of the N calibrated flow models where the weighting factor is a function of the error norm.

The problem of stochastic inverse modeling has been studied by several researchers (Van Leeuwen, et.al., 2000; Guagagnini & Franzetti, 1999; Copty & Findikakis, 2000; Evers & Lerner, 1998;Varljen & Shafer, 1991; Vassolo et al. 1998). However, most of these approaches are based on either analytical models or inverse algorithms applied to polygonal material property zones. Our approach is based on transition probability geostatistics, thus allowing us to condition the results not only to head observations, but also to borehole data.

Indicator Simulations

The first step in the capture zone delineation process is to run an indicator simulation to generate N equally probable realizations of the aquifer heterogeneity. The indicator simulation approach described in this paper is based on the T-PROGS software (Carle, 1997a; Walker, 2002). The T-PROGS software utilizes a transition probability-based geostatistical approach to model spatial variability by 3-D Markov Chains (Carle & Fogg, 1997), set up indicator cokriging equations (Carle & Fogg, 1996), and formulate the objective function for simulated annealing (Carle, 1997b).

The transition probability approach has several advantages over traditional indicator kriging methods. First, the transition probability approach considers asymmetric juxtapositional tendencies, such as fining-upwards sequences. Second, the transition probability approach has a conceptual framework for incorporating geologic interpretations into the development of cross-correlated spatial variability. Furthermore, the transition probability approach does not exclusively rely on empirical curve fitting to develop the indicator (cross-) variogram model. This is advantageous because geologic data are typically only adequate to develop a model of spatial variability in the vertical direction.

The transition probability approach provides a conceptual framework to geologic insight into a simple and compact mathematical model, the Markov chain. This is accomplished by linking fundamental observable attributes – mean lengths, material proportions, anisotropy, and juxtapositioning – with Markov chain model parameters.

The first step in performing a transition probability analysis is to compute a set of transition probability curves as a function of lag distance for each category for a given sampling interval. Each curve represents the transition probability from material j to material k. The next step is to develop a Markov Chain model for the vertical direction that fits the observed vertical transition probability data. Mathematically, a Markov chain model applied to one-dimensional categorical data in a direction ϕ assumes a matrix exponential form:

$$T(h_\phi) = \exp(\mathbf{R}_\phi h_\phi) \tag{1}$$

where h_ϕ denotes a lag in the direction ϕ, and $\mathbf{R}_\phi$ denotes a transition rate matrix

$$\mathbf{R}_\phi = \begin{bmatrix} r_{11,\phi} & \cdots & r_{1k,\phi} \\ \vdots & \ddots & \vdots \\ r_{k1,\phi} & \cdots & r_{kk,\phi} \end{bmatrix} \tag{2}$$

with entries $r_{jk,\phi}$ representing the rate of change from category j to category k (conditional to the presence of j) per unit length in the direction ϕ. The transition rates are adjusted to ensure a good fit between the Markov Chain model and the observed transition probability data.

Once the Markov chain is developed for the z direction from the borehole data, a model of spatial variability is developed for the x and y directions. The x, y, and z Markov chains are converted into a continuous 3D Markov chain using the MCMOD utility within T-PROGS. In the final phase of setting up a transition probability analysis using T-PROGS, the modeler creates a grid, specifies the number of model instances (N), and launches the TSIM utility. The TSIM code uses the 3D Markov chain to formulate both indicator cokriging equations and an objective function for simulated annealing. It generates stochastic simulations using a combination of modified versions of the GSLIB codes SISIM and ANNEAL (Deutsch & Journel, 1992).

The output from the TSIM code is a set of N arrays of indicators, where each entry in the arrays is the material id for the corresponding MODFLOW grid cell. These indicators can be used to define parameter zones in MODFLOW 2000. Each indicator type becomes a parameter and the hydraulic properties (k_h, k_z) for each of the cells are inherited from the list of parameters. A sample MODFLOW grid generated via transition probability geostatistics is shown in Figure 1.

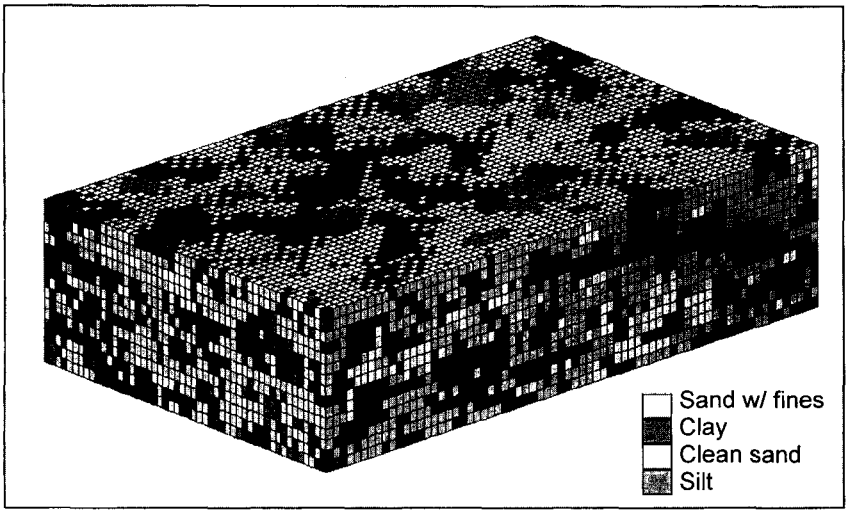

Figure 1. MODFLOW grid with hydraulic property data populated by T-PROGS.

Inverse Models

Once the N MODFLOW models are established by the indicator simulation, the next step is to calibrate each of the models to observed heads and flows. This can be accomplished using the new PES process in MODFLOW 2000 or using a separate parameter estimation utility such as PEST or UCODE (Doherty, 2000; Poeter & Hill, 1998). A set of material properties (K_h, K_v, etc) is associated with each of the indicators. These properties are marked as parameters and each of the N models is solved in inverse mode. This is a time-consuming process but it can be done in batch mode and can be solved using a distributed set of computers. During the inversion process, some of the models may not converge to a solution. These models are discarded prior to continuing to the next step.

Capture Zone Delineation

Once the entire set of models is calibrated, the final step is to delineate the probabilistic capture zones. To develop a capture zone risk map, we generate a "capture frequency" array with one entry for each cell and we initialize all of the array values to zero. We then perform a particle tracking analysis with MODPATH using the results from each of the MODFLOW solutions from the stochastic run. For each MODPATH run, one or more particles is placed in each cell and is tracked forward in time. If a particle is captured by a well, we increment the entry in the capture frequency array corresponding to the cell where the particle originated. We then divide the number of "captures" for each cell by the total number of MODPATH

runs to give a percentage representing the probability of capture. These percentages are contoured as a capture zone risk map.

As part of this research, we developed three methods for delineating capture zone using this basic approach. The three methods differ in terms of how the particles are distributed throughout the grid prior to the particle tracking step and how the frequency array is converted to a risk map. The three methods involve placing particles at the water table, placing particles at the cell centers, and deriving two dimensional projection views using results from the cell center method.

Water Table Approach. One method for particle placement is to place a particle at the center of the top-most active cell in each vertical column (i.e., at the water table). Each particle is then tracked forward in time to determine which particles reach the well in question. The results from each material set are then combined to generate a data set of capture probability as described above. The two-dimensional distribution represents the area at the top of the aquifer that could be captured by the well. It is important to note that this two-dimensional capture zone does not describe the total extent of the capture zone in the projected x and y direction, only the area where the capture zone intersects the water table. In theory, this capture zone is the most critical capture zone in terms of wellhead protection. However, it may not be sufficiently conservative since it is highly dependent on assumption made in developing the model (number of model layers, location of the well screen, etc.) and it ignores the possibility that contaminants leaking from the surface downward may come in contact with the full 3D plume by traveling through a preferential flow path (fissure, abandoned borehole, etc.) that is not represented in the model.

Three-Dimensional Approach. A more conservative approach is to generate a full three-dimensional capture zone. This is accomplished by placing the particles at the center of all cells in the grid. Once again, the particles are tracked forward in time and a capture probability data set is generated. This yields a three-dimensional probability array that can be used to develop probability iso-surfaces.

Two-Dimensional Projection. The most conservative method for defining probabilistic capture zones is to combine aspects from both the cell-centered and water table approach. A two-dimensional probability array is developed from three-dimensional capture zones created from each simulation. The capture zone from each model run represents the projection of the three-dimensional capture zone onto a 2D surface. In developing a 2D projection from 3D results, we looked at two methods to "flatten" the data. The first method, called the **any point** method and viewed as the most conservative, is to place particles at the cell centers of each active cell in the grid and track particles forward in time. For a given vertical column of cells, if any of the particles for any of the cells reached the well in question, then the cell is marked in the corresponding 2D grid. This means that a column where only one particle in one cell reached the well would receive the same weight as a column where almost every particle in every cell reached the well. The second method is called the **maximum probability** method. With this method, we weight each vertical column according to the maximum probability found in any cell in the column.

Weighting Factor

In all of the capture zone methods, the algorithms used to synthesize the probability dataset can be weighted using observation data. This makes it possible to give more weight to model instances with smaller calibration error when calculating the capture zone probabilities. The weighted head and flow observations can be compared to the computed values to come up with a global error norm, E, for each model run. This error norm can be based on the root mean squared (RMS) error, the sum of the weighted residuals, or any other measure selected by the modeler. The error norm from each MODFLOW run is used to compute a weight for the given solution using the following equation:

$$W_i = \alpha^{\left[\frac{ME-E_i}{SD}\right]} \tag{3}$$

Where w_i is the weight applied to solution i, α is a user-defined factor, ME is the mean of the error values from all solutions, E_i is the error for solution i, and SD is the standard deviation of error values from all solutions. The weights are also normalized as follows

$$W_{final} = \frac{W_i}{\sum W_i} \tag{4}$$

so that the weights sum to unity. Equation 3 was developed to give the greater emphasis to the lower error values and to allow the user to control the relative emphasis given to low vs. high values simply by adjusting the α value. The equation also avoids problems when one of the error values is zero, since a zero error value does not result in an infinite weight. We also wanted the equation to scale the weights according to the data being examined. This is done by subtracting the individual RMS from the mean error and dividing by the standard deviation.

Equation 3 centers the weights on the mean error. The relative weight given to values differing from the mean is biased by the factor. This makes it possible to bias the resulting weight using knowledge of the site and the quality of the observation data.

Figure 2 shows how the α factor in Equation 3 affects the weight applied to a given error. An α factor of 1.2 makes the contribution of each RMS almost linear, whereas an α factor of ten gives most of the weight to the lowest 5-10 percent while discounting the other error values. We typically use an α value of 2.0.

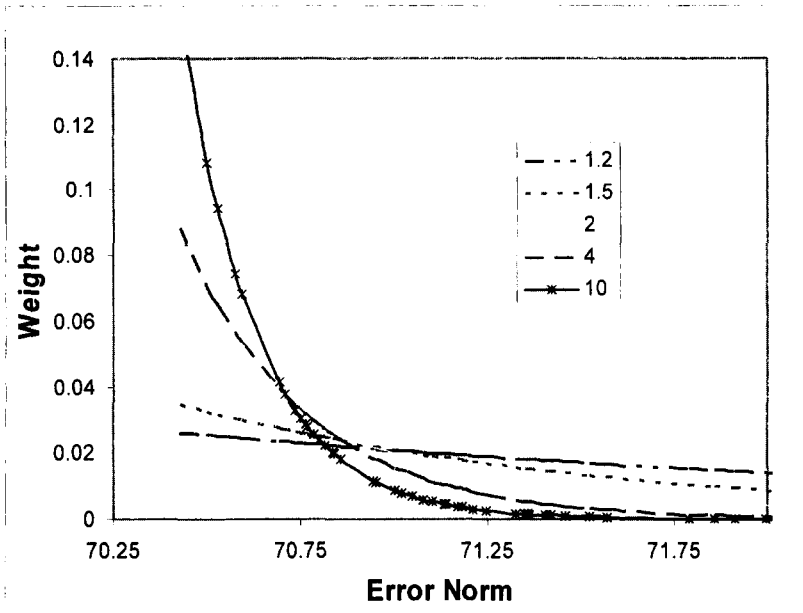

Figure 2. Graph showing how different values for the factor a result in different weights for a given RMS.

Sample Application

The following case study illustrates the stochastic inverse method and the capture zone delineation techniques described above. The Longhorn Army Ammunition Plant (LHAAP) is an inactive installation in eastern Texas. Before receiving inactive status, the LHAAP was responsible for production, distribution, and decommissioning of munitions that resulted in the contamination of surface and groundwater systems. This contamination ranges from solvents and oxidizers to explosives (U.S. Army, 2001). In March of 2001, the United States Army Engineer Research and Development Center – Waterways Experiment Station (ERDC-WES) released a study of the groundwater and surface water of the area that attempted to simulate the hydrogeologic system in the area to predict the transport of chemicals from the LHAAP site. We selected a small subset of this site to illustrate our new capture zone delineation method. The local study area contained 74 complex boreholes with 61different materials described using the Unified Soil Classification system. We reviewed the borehole data and simplified the material groupings into four basic materials: clean sand, sand with fines, silt, and clay.

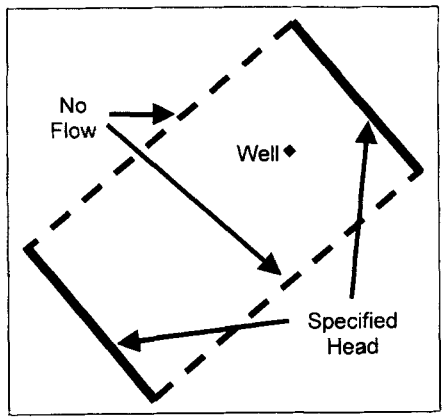

Figure 3. Conceptual model of the LHAAP site.

We created 100 indicator realizations for a ten layer grid using the T-PROGS software. The grid had 70 cells in the x direction and 50 cells in the y direction for a total of 35,000 cells. Once the indicator sets were generated, we constructed the MODFLOW model based on the conceptual model shown in Figure 3. Because this is a local scale model, there are parallel flow boundaries on the east and west, and the remaining north and south boundaries are no-flow. In the center right of the site is one well that is screened in the fifth layer. The top boundary of the MODFLOW grid was taken from the top of the boreholes. The bottom grid boundary was inferred from geological estimates provided by the ERDC-WES model documentation. The remaining layer boundaries were divided so that each layer has an equal thickness.

Basic Stochastic Approach. We first generated a set of capture zones using a basic stochastic approach without the inversion step. The time it took to complete all the model runs was approximately 4.25 hours. The capture zone was first delineated using the water table approach. Four particles were placed in each cell at the water table boundary and traced forward in time as described above. The results are shown in Figure 4. Next, we utilized the three dimensional approach where one particle was placed at the center of each of the cells in the three-dimensional grid and tracked forward in time. The resulting probabilistic capture zone is shown in Figure 5. Note that the probability contours in this case are represented as iso-surfaces corresponding to different levels of capture probability. We then produced a set of 2D capture zones by projecting the results of the 3D approach using the *any point in column* method and the *maximum probability in column* method

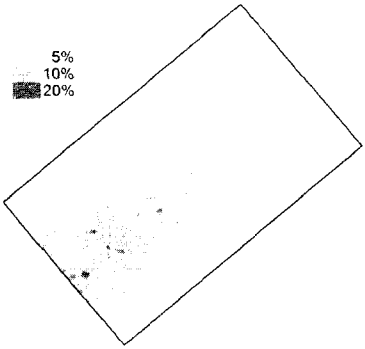

Figure 4. Capture probability map using the "water table" approach.

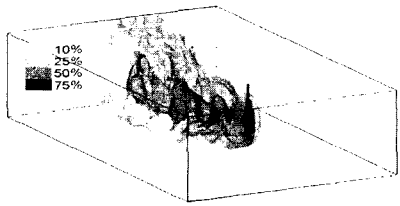

Figure 5. Three-dimensional capture probability map.

(Green, 2002). The *any point in column* method (Figure 6) is the more conservative of the two methods. With this method, when analyzing the contribution from a single model to the capture probability for a cell in the 2D grid, the cell is considered to be part of the capture zone for a well if any of the particles originating in the corresponding vertical column of cells in the 3D grid is captured by the well. With the *maximum probability in column* method (Figure 7), each of the cells in a vertical column of cells in the 3D grid is analyzed and the probability for the cell with the maximum probability of capture is assigned to the corresponding cell in the 2D grid.

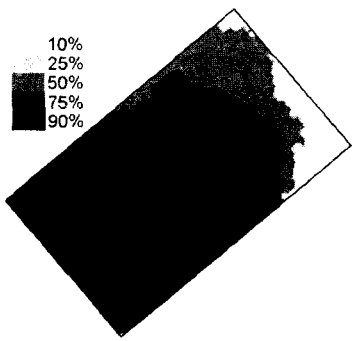

Figure 6. A 2D projection of a 3D probabilistic capture zone created using the *any point* method.

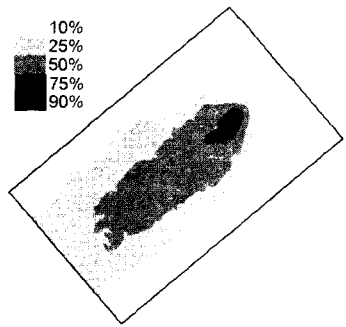

Figure 7. A 2D projection of a 3D probabilistic capture created by using the *maximum probability* method.

Stochastic Inverse Approach. Next, we developed capture zones using the stochastic inverse approach. Inverse modeling is more time intensive than pure stochastic modeling because when each stochastic simulation is inverted, MODFLOW can be run two times the number of parameters for each inversion iteration. To reduce the run time for stochastic inverse modeling of the LHAAP site, we chose to reduce the number of layers from twenty to one. This was accomplished by simplifying the boreholes to a single material represented by the predominant material found in each borehole and re-running T-PROGS with a one-layer grid. Each instance of the one layer model was run in forward mode to determine if the model would converge. Each model that converged was then inverted to attempt to match the heads at 49 observation wells. We inverted the horizontal hydraulic conductivity for each of the four materials. Running all of the stochastic inverse simulations took almost three hours. The probabilistic capture zone resulting from the stochastic inverse solutions is shown in Figure 8. When generating the capture zone map, we weighted the results from each of the individual models according to the RMS error norm using Equation 3.

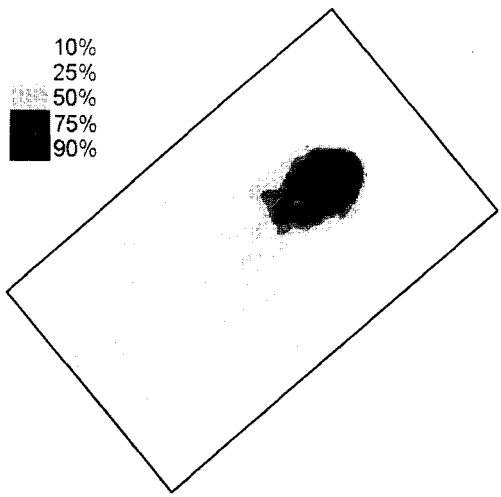

Figure 8. Probabilistic capture zone generated using the stochastic inverse method.

Conclusion

The technique described in this paper can be used to successfully develop probabilistic capture zone maps. A unique feature of the stochastic inverse approach described in this paper is that each of the candidate model solutions used to develop the capture zones is conditioned to both observed heads/flows and to borehole data.