

Application of Transition Probability Geostatistics for Indicator
Simulations Involving the MODFLOW Model

by

Justin R. Walker

A thesis submitted to the faculty of

Brigham Young University

In partial fulfillment of the requirements for the degree of

Master of Science

Department of Civil Engineering

Brigham Young University

March 2002

Copyright © 2002 Justin R. Walker

All Rights Reserved

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Justin R. Walker

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date

Norman Jones, Chair

Date

Kyle Rollins

Date

Wayne Downs

Date

Terry Lee

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Justin R. Walker in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Norman L. Jones
Chair, Graduate Committee

Accepted for the Department

Wayne C. Downs
Graduate Coordinator

Accepted for the College

Russell J. Hunt
Dean, College of Engineering and Technology

ABSTRACT

APPLICATION OF TRANSITION PROBABILITY GEOSTATISTICS FOR INDICATOR SIMULATIONS INVOLVING THE MODFLOW MODEL

Justin R. Walker

Department of Civil Engineering

Master of Science

This thesis describes a technique for utilizing the transition probability geostatistics method for stochastic simulations using the MODFLOW model. Transition probability geostatistics has numerous advantages over traditional indicator Kriging methods including a simpler and more intuitive framework for interpreting geologic relationships and the ability to simulate juxtapositional tendencies such as fining upwards sequences. The indicator arrays generated by the transition probability simulation can be

converted into material identification arrays for use in the Layer Property Flow (LPF) package in MODFLOW 2000. Furthermore, material sets can be used to assign three dimensional borehole data to a one-layer model. The indicator arrays generated by the transition probability simulation can also be converted into layer elevation and thickness arrays for use with the new Hydrogeologic Unit Flow (HUF) package in MODFLOW 2000. This makes it possible to preserve complex heterogeneity while using reasonably sized grids. A case study is presented illustrating all three types of applications on a real site in Texas.

ACKNOWLEDGMENTS

I would like to express appreciation to my wife, Aubrey for her enduring support and patience in this endeavor. Second, I would like express appreciation to Dr. Norman Jones, my committee chair. He aided in the research as well as drafting this paper. I could not have completed this work without his help. Finally I would express gratitude to the Civil Engineering Department, the College of Engineering and Technology, and Brigham Young University for establishing the framework of cooperative research between student and faculty.

TABLE OF CONTENTS

1	Introduction.....	1
1.1	Indicator vs. Monte Carlo Approach	2
1.2	Transition Probability Geostatistics.....	4
1.3	The MODFLOW Model	4
1.4	Overview of Research.....	4
2	Literature Review.....	6
2.1	Normal Processes.....	6
2.2	p-Field Simulation	7
2.3	Boolean Algorithms	7
2.4	Simulated Annealing.....	8
2.5	Kriging	8
2.5.1	Simple Kriging.....	9
2.5.2	Ordinary Kriging.....	10
2.5.3	Trend Model Kriging	10
2.5.4	CoKriging	11
2.5.5	Indicator Kriging.....	11
3	Transition Probability Geostatistics.....	15
3.1	Markov Chains.....	17
3.2	Options for Creating Markov Chains	19

3.2.1	Rates.....	19
3.2.2	Embedded Transition Probabilities.....	23
3.2.3	Embedded Transition Frequencies.....	27
3.2.4	“Independent” or “Maximum Entropy” (Disorder) Transition Frequencies 28	
3.3	Conceptual Development.....	33
3.3.1	Background Material	33
3.3.2	Material Proportions	34
3.3.3	Mean Length	34
3.3.4	Asymmetry.....	35
3.3.5	Sparse or No Data	36
3.4	Multidimensional Markov Chains	36
3.5	Eigensystem Analysis	40
3.6	Developing a Simulation.....	40
4	Indicator Arrays to HUF arrays	48
4.1	Hydrogeologic Unit Flow (HUF) Package	49
4.2	T-PROGS to HUF Conversion Algorithm.....	50
4.2.1	Create a MODFLOW Grid	51
4.2.2	Create a Background Grid	51
4.2.3	Run T-PROGS	52
4.2.4	Convert the T-PROGS Output	52
5	Single Layer MODFLOW Models	56
5.1	Description of Algorithm.....	57
6	Case Studies	61
6.1	Site Description.....	61

6.2	Description of Model	62
6.2.1	Transition Trends	66
6.3	3-D Orthogonal Grid.....	67
6.4	Single Layer Model.....	76
6.5	HUF Arrays.....	79
7	Conclusion	87
Appendix A	References.....	91
Appendix B	GMS/T-PROGS Interface.....	93
B.1	Building a Grid	94
B.2	General Options	94
B.3	Vertical (Z) Markov Chains.....	98
B.3.1	Plot Section	98
B.3.2	Markov Chains Section.....	102
B.3.3	Spreadsheet Section	110
B.4	Strike (X) Markov Chains.....	110
B.5	Dip (Y) Markov Chains	112
B.6	Running TSIM	112
B.7	Post-Processing.....	115

LIST OF FIGURES

Figure 3.1 Matrix of vertical-direction transition probabilities for LLNL core data: measurements (dots) and Markov chain model (solid lines). Intersection of dashed line (tangent) with lag axis indicates mean length; dotted line indicates proportions (Carle, 1996)	22
Figure 3.2 Embedded occurrences of a three-category system with 1 = white, 2 = gray, 3 = black. (Carle, 1999)	23
Figure 3.3 Markov chain model fit to matrix transition probability data with the embedded transition probability (solid line) and the maximum entropy (dashed line) approaches. Actual measured data is illustrated with a dotted line. (Carle, 1999)...	26
Figure 3.4 Transition probability curves from measured data and Markov chain generated from embedded transition frequencies.....	28
Figure 3.5 Transition probability curves generated from measured data and maximum entropy factors.	31
Figure 3.6 Transition probability curves generated from measured data and maximum entropy Markov chains.	32
Figure 3.7 Sample parameter file for GAMEAS. (Carle, 1999).....	42
Figure 3.8 Sample data file. (Carle, 1999).....	42
Figure 3.9 Sample parameter file for MCMOD (Carle, 1999)	44
Figure 3.10 Vertical (z)-direction transition probabilities for models of spatial variability with SIS only and with SIS and quenching. (Carle, 1996).....	46
Figure 3.12 Sample TSIM parameter file. (Carle, 1999).....	47
Figure 4.1 Illustration of T-PROGS --> HUF.....	53
Figure 4.2 Sample HUF arrays for a four-layer grid.	54

Figure 5.1 Data file for a one-layer model.....	58
Figure 5.2 3D borehole condensed into a one-layer model using the 2D approach.	60
Figure 6.1 LHAAP and "Local study area". The dots represent boreholes.....	62
Figure 6.2 Plan view layout of the grid. Dots represent borehole locations.....	65
Figure 6.3 Vertical transition probability curves.....	67
Figure 6.4 Oblique view of the realization #1 with cells filled with material colors.	69
Figure 6.5 Oblique view of the realization #2 with cells filled with material colors.	69
Figure 6.6 Head (black) and particle flow line (blue) solution for realization #1.	70
Figure 6.7 Head (black) and particle flow line (blue) solution for realization #2.	71
Figure 6.8 Side view of flow lines and water table.....	72
Figure 6.9 Side view of flow lines and water table.....	72
Figure 6.10 3D capture zone for a probability shell equal to 10%.	73
Figure 6.11 3D capture zone for a probability shell equal to 30%.	74
Figure 6.12 3D capture zone for a probability shell equal to 50%.	74
Figure 6.13 3D capture zone for a probability shell equal to 70%.	75
Figure 6.14 3D cutaway of the interior of the capture zone with iso-surfaces capped.....	75
Figure 6.15 3D cutaway of the interior of the capture zone with iso-surfaces not capped.	76
Figure 6.16 Realization #1 for 2D approach with MODFLOW and MODPATH solutions.	77
Figure 6.17 Realization #2 for 2D approach with MODFLOW and MODPATH solutions.	78
Figure 6.18 2D probabilistic capture zone for the well.	79
Figure 6.19 Realization #1 for HUF approach with MODFLOW solutions (side view).	80
Figure 6.20 Realization #2 for HUF approach with MODFLOW solutions (side view).	81
Figure 6.21 Realization #1 for HUF approach with MODFLOW and MODPATH solutions (plan view).....	81

Figure 6.22 Realization #2 for HUF approach with MODFLOW and MODPATH solutions (plan view).....	82
Figure 6.23 HUF capture zone for a probability shell equal to 10%.....	83
Figure 6.24 HUF capture zone for a probability shell equal to 20%.....	83
Figure 6.25 HUF capture zone for a probability shell equal to 30%.....	84
Figure 6.26 HUF capture zone for a probability shell equal to 40%.....	84
Figure 6.27 HUF cutaway of the interior of the capture zone with iso-surfaces capped..	85
Figure A T-PROGS menu commands.	94
Figure B Markov chain general options dialog without borehole data.....	95
Figure C Markov chain general options dialog with borehole data.....	96
Figure D Definition of azimuth in T-PROGS algorithm	97
Figure E Grid angle orientation.	97
Figure F Vertical Markov chain dialog.....	98
Figure G Maximized plot.....	99
Figure H Curves with equal probability.....	100
Figure I GAMEAS executable inside GMS interface.....	101
Figure J Vertical (Z) Markov Chain dialog with option #1 selected.	104
Figure K Vertical (Z) Markov Chain dialog with option #2 selected.	105
Figure L Vertical (Z) Markov Chain dialog with option #3 selected.	106
Figure M Vertical (Z) Markov Chain dialog with option #4 selected.	107
Figure N Vertical (Z) Markov Chain dialog with option #5 selected.	109
Figure O Strike/Dip Markov chain dialog.	111
Figure P General options for simulation generation.	113
Figure Q MCMOD executable inside GMS interface.	114
Figure R TSIM executable inside GMS interface.....	115

Figure S Data tree used to organize HUF and Material sets..... 116

Figure T Grid populated with data from Material Set 2 2. 117

LIST OF TABLES

Table 3.1 LLNL site characterization from core data (Carle, 1996).	21
Table 6.1 Original materials in site boreholes.	63
Table 6.2 Materials in the local study area with corresponding proportions and hydraulic conductivity.....	64
Table 6.3 Strike/dip and vertical lens lengths for each material.....	65

1 INTRODUCTION

Ground water model algorithms compute hydraulic heads corresponding to hydraulic properties like conductivity and transmissivity. Typically, site investigations utilizing boreholes, ground penetrating radar, or other methods delineate the spatial distribution of hydraulic properties. Unfortunately, site data commonly represent at best one percent of the actual study area. Therefore, homogeneity is assumed to create polygonal zones with uniform hydraulic properties that are applied to the remainder of the study area. However, soil properties like hydraulic conductivity can vary as much as two orders of magnitude within a ten-foot radius. As a result, although overall regional aquifer behavior and analysis can be represented with uniform hydraulic properties, local conditions can deviate drastically from the model representation. The impact of soil heterogeneity can be magnified by the presence of small lenses, which dominate flow conditions. This situation can be significant when analyzing the transport of contaminants or when a local analysis requires precision and accuracy.

One approach for dealing with model heterogeneity is stochastic simulations based on multiple equally plausible candidate realizations of the site heterogeneity. Ideally, such an approach would enable the generation of variability in subsurface soil stratigraphy based on interpretable geologic parameters such as lens width, material proportions, juxta-positioning tendencies and anisotropy. Multiple realizations that

are conditioned to borehole data provide modelers with a rational approach for dealing with uncertainty associated with site characterization. Stochastic simulations can be applied to regional representations of the aquifer behavior in addition to local scale simulations. Stochastic simulations are particularly well-suited to local scale models since the resulting complex heterogeneity is more representative of actual stratigraphic deposition. This heterogeneity makes realistic and potentially more accurate contaminant transport simulation possible by simulating the preferential flow channels resulting from thin lenses of clays, sands, or other materials. The ultimate result of a stochastic approach is multiple simulations of hydraulic parameters that create a probabilistic solution. Such a solution has more credence and provides a better understanding of actual site conditions.

1.1 Indicator vs. Monte Carlo Approach

There are multiple methods of generating stochastic simulations of hydraulic parameters. The most common methods are the indicator approach and the Monte Carlo approach. The indicator approach operates on discrete parameters like geologic facies or material types. Indicator geostatistics generates multiple distributions of material zones (each of which is called a “realization”) that have hydraulic characteristics assigned to each material. This approach presumes that a relationship between geologic facies and hydraulic properties can be reasonably estimated. An indicator approach is defensible because hydraulic parameters and geologic data are categorical.

The Monte Carlo approach involves the generation of variability in the hydraulic

properties associated with a pre-defined material distribution. The variability for a simulation is created with a random number generator from a mean and standard deviation for the parameter being analyzed. The random number generator can be designed such that the resulting parameters exhibit either a linear (purely random), normal, or log-normal distribution. Each model instance is then generated by producing a random value for each of the selected parameters. The model is then run once for each combination of parameter values. The resulting multiple solutions can be used to assess the probability of some model outcome. The greater the number of model runs, the greater the confidence in the results. The Latin Hypercube approach is a variation on the Monte Carlo approach that ensures that the parameters values are combined in a fashion that allows for an equivalent confidence with fewer model runs.

The research described in this thesis utilizes the indicator approach. The main advantage of the indicator approach over the Monte Carlo approach is that it is conducive to stochastic inverse modeling. Stochastic inverse modeling involves using an automated parameter estimation engine such as PEST (Doherty, 2000) to optimize the parameter values associated with the material zones generated for each model instance. The end result of this process is that each of the N model runs is calibrated to field observed heads and flows. This provides a greater confidence in the probabilistic model predictions.

Another advantage of the indicator geostatistics approach is that it allows for geologic interpretation. A random number generator with only mathematical interpretation produces Monte Carlo realizations. Conversely, indicator geostatistics functions on the existence of geologic depositional knowledge. In addition, indicator

simulations can be conditioned to existing site data.

1.2 Transition Probability Geostatistics

The transition probability approach of indicator geostatistics is a relatively new method for representing heterogeneity in 3-D soil stratigraphy. Steven F. Carle and other researchers have demonstrated that this approach has some unique advantages over traditional indicator methods like cross-variograms. The transition probability approach is rational and interpretable as well as mathematically efficient due to probability constraints. Carle has developed a software package called T-PROGS that operates on the basis of transition probabilities (Carle, 1999). The research in this thesis utilizes the T-PROGS software and the groundwater modeling tools of the Department of Defense Groundwater Modeling System (GMS) to generate 3-D multiple realizations of soil heterogeneity.

1.3 The MODFLOW Model

MODFLOW (Harbaugh, 2000) is the most widely used groundwater model. MODFLOW is a 3D saturated flow model that can be used for steady state or transient simulations and it can represent a variety of sources and sinks. Since it is a saturated flow code, it is a relatively efficient model. Model runs generally converge rapidly and model instability is rare. This makes MODFLOW an ideal candidate for stochastic simulations where a large number of runs are required.

1.4 Overview of Research

The objective of my research was to link transition probability geostatistics with the MODFLOW model for stochastic simulations. This research has resulted in the development of two new algorithms for building MODFLOW-compatible datasets from indicator arrays generated by transition probability geostatistics. Both tools operate with the aid of the T-PROGS software developed by Steven Carle. The first tool enables the generation of multiple realizations of HUF data, which can be incorporated into the new HUF package in MODFLOW (Anderman, 2000). The second tool condenses three-dimensional borehole data into a two-dimensional MODFLOW grid based on the dominant material in each borehole data. It then generates stochastic two-dimensional realizations of material types that honor the borehole data.

In the remainder of this thesis, I will first discuss the theory behind the transition probability approach of indicator geostatistics and Steven Carle's software package, T-PROGS. Next, I will describe the interface that enables the implementation of the T-PROGS algorithm inside *GMS*. Thirdly, the two new tools of stochastic HUF arrays and 2D material sets will be outlined. Last, sample applications of these two new tools will be described.

2 LITERATURE REVIEW

Various methods exist to generate stochastic realizations of subsurface stratigraphy. Stochastic simulation involves generating alternate, equally probable, high-resolution models of the spatial distribution of aquifer zones. A simulation is considered “conditioned” if the resulting realization honors known data values at their respective locations. Methods of generating stochastic simulations were reviewed in order to determine what methodology would be most effective and compatible with MODFLOW simulations. These methods can be grouped into five general types: Normal processes, Indicator-Based Algorithms, p-Field Simulation Algorithms, Boolean Algorithms, and Simulated Annealing (Deutsch, 1998). In addition, Kriging algorithms will be reviewed.

2.1 Normal Processes

Normal, or Gaussian-related, methods are the algorithms of choice for most continuous random variables. Gaussian random function models are unique because of their analytical simplicity. The basis of normal processes is the Central Limit Theorem. Normal processes are particularly useful because many natural systems can be modeled with random variables that are approximately normally distributed. A stochastic process $\{X(t), t \in \mathbf{T}\}$ is said to be normal if for any integer n and any subset $\{t_1, t_2, \dots, t_n\}$ of \mathbf{T} random variables $X(t_1), \dots, X(t_n)$ are jointly normally distributed (Parzen, 1964).

HMMM.. THIS STILL LEAVES SOMETHING TO BE DESIRED.

2.2 p-Field Simulation

In contrast to sequential simulation methods, which condition to the original data as well as previously simulated values, p-Field simulation condition only to the original data. As a result, conditional cumulative distribution functions need only be calculated once instead of iteratively calculated for each realization. This advantage streamlines time requirements, which, is one of the advantages of the p-Field method. The conditional cumulative distribution functions (ccdf's) can be obtained through multi-Gaussian Kriging of the z continuous data or through indicator Kriging performed on indicator data.

2.3 Boolean Algorithms

Boolean algorithms cover a large range of categorical simulation algorithms. Boolean processes are generated by the distribution of geometric objects in space according to some probability laws. The major problem with geologic application of Boolean algorithms is that geological lithofacies are rarely a simple parametric shape. Furthermore, they are rarely distributed uniformly within a study area. As a result, the determination of a Boolean model is a trial-and-error process until the final stochastic image is visually satisfactory to the modeler. Calibration of the Boolean model is also a matter of art rather than statistical inference. Hence, Boolean methods are usually custom-built.

2.4 Simulated Annealing

Simulated annealing is a relatively new approach that is capable of generating conditional stochastic images of continuous or categorical variables. The basic idea of simulated annealing is to iteratively perturb an initial image until it satisfies predefined characteristics included in an objective function. The objective function is minimized during the perturbations. If the perturbation improves the objection function, it is accepted. However, some negative perturbations are accepted to avoid local minimums.

Simulated annealing can be CPU-intensive if the quality of the perturbation cannot be determined accurately and quickly. The key to this technique is to quickly and accurately ascertain the quality of the image between perturbations. A popular application of simulated annealing is to improve an initial stochastic image generated with a Gaussian or indicator algorithm. An initial configuration can be determined during the simulated annealing process by assigning the conditioning data to the nearest grid nodes and then assigning the remaining nodes randomly from the user-specified histogram.

2.5 Kriging

Kriging is the fundamental mathematical algorithm employed by many stochastic simulator processes. A concept integral to geostatistics and Kriging is a regionalized variable. In contrast to random variables, regionalized variables have continuity from point to point, but the correlation is so complex that is cannot be described by any mathematical function. Therefore, spatial correlations for regional variables must be

drawn over short distances. Regionalized variables typically describe natural phenomena that have geographic distributions like depositional tendencies. Geostatistics is the estimation of the form of a regionalized variable in one, two, or three dimensions (Statistics and Data Analysis in Geology, Davis, 1986). The spatial continuity of a regional variable can be measured by a variogram, which is a plot of semivariances γ_h given by

$$\gamma_h = \sum_i^{n-h} (X_i - X_{i+h})^2 / 2n \dots\dots\dots \text{Eq. 2.2}$$

where X_i is a measurement of a regionalized variable taken at a location i , X_{i+h} is another measurement taken h intervals away, and n is the number of points. If sample of the data have been taken and the form of the variogram is known, it is possible to estimate the unsampled points anywhere along the surface. This estimating process is called Kriging. Unlike most interpolation algorithms, Kriging establishes a measure of the error or uncertainty of a estimated surface. Traditionally, Kriging is used to provide estimates for unsampled locations that minimize the error variance at these locations. However recently, Kriging has been utilized to build probabilistic models of uncertainty about these unknown locations. The Kriging principle has been presented in numerous papers and books (GSLIB Geostatistical Software Library and User's Guide, Deutsch, 1998). There are several variations of Kriging. I will address five of the approaches: simple Kriging (SK), ordinary Kriging (OK), Kriging with various trend models (KT), coKriging, and indicator Kriging (IK).

2.5.1 Simple Kriging

The simple Kriging estimator is defined as

$$[Z_{SK}^*(\mathbf{u}) - m(\mathbf{u})] = \sum \lambda_{\alpha}(\mathbf{u})[Z(\mathbf{u}_{\alpha}) - m(\mathbf{u}_{\alpha})] \dots\dots\dots \text{Eq. 2.3}$$

where $Z(\mathbf{u})$ is the random variable model at location \mathbf{u} , the \mathbf{u}_{α} 's are the n data locations, $m(\mathbf{u}) = E\{Z(\mathbf{u})\}$ is the location-dependent expected value of the random value $Z(\mathbf{u})$, the λ_{α} are the weights, and $Z_{SK}^*(\mathbf{u})$ is the linear regression estimator, also called the “simple Kriging” estimator. Simple Kriging does not adapt well to local trends because it assumes that the mean remains constant.

2.5.2 Ordinary Kriging

Indicator random functions are binary and are therefore ideally suited for simulating categorical variables. The binary indicator variable is set to 1 if a certain category is present at a given location, or 0 if not. The ordinary Kriging estimator is given by

$$Z_{OK}^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha}^{(OK)}(\mathbf{u})Z(\mathbf{u}_{\alpha}) \dots\dots\dots \text{Eq. 2.4}$$

where $\lambda_{\alpha}^{(OK)}(\mathbf{u})$'s are the ordinary Kriging weights. Ordinary Kriging is more conducive to local trends because it allows for a location-dependent mean. Therefore, it is the algorithm of choice for geostatistics.

2.5.3 Trend Model Kriging

Kriging with a trend, or “universal” Kriging, allows for the mean to vary

according to a particular trend given by

$$Z_{KT}^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha}^{(KT)}(\mathbf{u})Z(\mathbf{u}_{\alpha}) \dots\dots\dots \text{Eq. 2.5}$$

where $\lambda_{\alpha}^{(KT)}(\mathbf{u})$'s are the Kriging trend weights. Forms of Kriging with a trend are Kriging with an external drift, Bayesian Kriging, and factorial Kriging.

2.5.4 CoKriging

While Kriging is traditionally linear regression using data on the same attribute as that being estimated, cokriging is linear regression that also uses data defined on different attributes. For example, cokriging could be applied to hydraulic conductivity as the primary sample, and porosity (which can be related to hydraulic conductivity) as the secondary sample. The ordinary cokriging estimator of $Z(\mathbf{u})$ is

$$Z_{COK}^*(\mathbf{u}) = \sum_{\alpha1=1}^{n1} \lambda_{\alpha q}(\mathbf{u})Z(\mathbf{u}_{\alpha1}) + \sum_{\alpha2=1}^{n2} \lambda'_{\alpha2}(\mathbf{u})Y(\mathbf{u}_{\alpha2}) \dots\dots\dots \text{Eq. 2.6}$$

where the $\lambda_{\alpha1}$'s are the weights applied to the n_1 z samples and the $\lambda'_{\alpha1}$'s are the weights applied to the n_2 y samples.

2.5.5 Indicator Kriging

An indicator is a random variable that represents a discrete category at a given location. Indicator variables are mutually exclusive and are defined over a region **D** by

$$I_k(\mathbf{x}) = \begin{cases} 1, \text{ if category } k \text{ occurs at location } \mathbf{x} \\ 0, \text{ otherwise} \end{cases} \dots\dots\dots \text{Eq. 2.7}$$

where $\mathbf{x} \in \mathbf{D}$ and category $k = 1, \dots, K$. An indicator approach is fundamentally different than a continuous variable approach. A continuous random variable can assume any value within a given range. A geologic example of a continuous variable is hydraulic conductivity. The hydraulic conductivity at a given location can have any value inside some rational range. Indicator Kriging produces a least-square estimate of the conditional cumulative distribution function (ccdf) at cutoff z_k :

$$\begin{aligned} [i(\mathbf{u}, z_k)]^* &= E\{I(\mathbf{u}, z_k | (n))\}^* \\ [i(\mathbf{u}, z_k)]^* &= Prob^*\{Z(\mathbf{u}) \leq z_k | (n)\} \end{aligned} \dots\dots\dots \text{Eq. 2.8}$$

where (n) represents the conditioning information available in the neighborhood of location u . The indicator Kriging process is repeated for a series of K cutoff values $z_k, k = 1, \dots, K$, which discretize the interval of variability of the *continuous* attribute z (Deutsch, 1998). The conditional cumulative distribution function represents a probabilistic model of uncertainty for the unsampled value $z(u)$. If $z(u)$ is categorical, then the direct Kriging of $z(u)$ provides a model for the probability that $z(u)$ is equal to one, or in other words that a particular category prevails at that location u . If $z(u)$ is continuous, then the selection of the cutoff values, z_k becomes important. Too many cutoff values produce computational inefficiency. Too few cutoff values result in the loss of detail in the model.

One type of indicator Kriging uses a transition probability approach. The transition probability method is a modified form of indicator Kriging. The transition probability approach couples geologic knowledge and mathematical manipulations to

overcome many of the shortcomings of the traditional indicator geostatistical methods. Conceptually, the transition probability approach assumes that the future depends entirely on the present, not the past. When applying it to spatial distribution terms, the time lag is replaced with a distance lag and the local occurrence of a category is entirely dependent on the nearest occurrence of another or the same category. Mathematically, the transition probability $t_{jk}(h_\phi)$ is given by

$$t_{jk}(h_\phi) = \Pr\{\text{category } k \text{ occurs at } \mathbf{x} + h_\phi \mid \text{category } j \text{ occurs at } \mathbf{x}\} \text{ Eq. 2.9}$$

where h_ϕ represents a positive lag separation in the direction ϕ . The transition probability approach is gaining popularity in geologic applications because it is intuitive and mathematically efficient.

The transition probability approach to indicator geostatistics was incorporated into the research described in this thesis. Certain characteristics made it the method of choice. First, the indicator approach is conducive to geologic applications because hydrogeology can be categorized into discrete geologic facies. Transition probabilities can account for asymmetric juxtapositional tendencies like fining-upwards, which are typical in deposition patterns. Furthermore, the transition probability approach provides a conceptual framework to incorporate geologic knowledge such as mean lengths, material proportions, anisotropy, and juxtapositioning. The transition probability approach helps incorporate geologic interpretation into the development of Markov chain models of spatial variability, providing means for quantifying subjective insights on spatial variability that a geologist might otherwise infuse directly into a hand-drawn

geologic cross-section. A main advantage of the transition probability indicator approach is that it is conducive to stochastic inverse modeling. Stochastic inverse modeling involves using an automated parameter estimation engine such as PEST (Doherty, 2000) to optimize the parameter values associated with the material zones generated for each model instance. The end result of this process is that each of the N model runs is calibrated to field observed heads and flows. This provides a greater confidence in the probabilistic model predictions.

3 TRANSITION PROBABILITY GEOSTATISTICS

The research described in this thesis is based upon the transition probability geostatistics approach for generating multiple realizations of model heterogeneity for use in indicator based stochastic simulations. The transition probability method is a modified form of indicator Kriging, but it has several advantages over traditional indicator Kriging. First, traditional indicator geostatistics does not provide consideration for asymmetric tendencies because indicator cross-variograms inherently assume symmetry. Asymmetry implies that

$$p_{mk}(h) \neq p_{mk}(-h) \dots\dots\dots \text{Eq. 3.1}$$

or

$$p_{mk}(h) \neq p_{km}(h) \dots\dots\dots \text{Eq. 3.2}$$

where p_{mk} denotes the joint probability and h denotes a lag separation vector. The cross-variograms $\gamma_{mk}(h)$ uses an averaging technique illustrated by

$$\gamma_{mk}(h) = p_{mk}(0) - [p_{mk}(h) + p_{mk}(-h)]/2 \dots\dots\dots \text{Eq. 3.3}$$

Therefore, any asymmetry such as fining upwards is lost. Second, traditional indicator geostatistics does not have a conceptual framework for incorporating geologic interpretations into the development of cross-correlated spatial variability. Third,

traditional indicator geostatistics lack the ability to consider local variable anisotropy directions like radial morphology of alluvial fans or meandering of fluvial depositional units and structure resulting from deformation (Carle, 1998). Furthermore, traditional models of spatial variability are generated by empirical curve fitting a sample indicator (cross-) variogram with a mathematical function such as a spherical or exponential model (Carle, 1996). However, geologic data is typically only adequate to develop a variogram in the vertical direction. The scarcity of site data makes curve fitting impractical in the strike and dip direction. It is not uncommon to be unable to generate a realistic variogram even in the vertical direction.

The transition probability approach couples geologic knowledge and mathematical manipulations to overcome many of the shortcomings of the traditional indicator geostatistical methods. First, transition probabilities can account for asymmetric juxtapositional tendencies like fining-upwards. Furthermore, the transition probability approach provides the conceptual framework to incorporate geologic knowledge such as mean lengths, material proportions, anisotropy, and juxtapositioning. The transition probability approach helps incorporate geologic interpretation into the development of Markov chain models of spatial variability, providing means for quantifying subjective insights on spatial variability that a geologist might otherwise infuse directly into a hand-drawn geologic cross-section. The remainder of the chapter will outline the transition probability approach as developed by Steven F. Carle as described in various publications including Carle (1996), Carle (1999), Carle et al (1998), Carle & Graham (1995), Carle & Graham (1997).

3.1 Markov Chains

Conceptually, when applied to time, a 1-D Markov chain assumes that the future depends entirely on the present, not the past. When applying it to spatial distribution, time is replaced with a distance lag and the local occurrence of a category is entirely dependent on the nearest occurrence of another or the same category. “Mathematically, a continuous Markov chain is a transition probability model described by a matrix exponential function, which provides a solution to a first-order stochastic differential equation.” (Carle, 1996) The 1-D transition probability matrix $T(h_\phi)$ has the following definition:

$$T(h_\phi) = \exp(R_\phi h_\phi) \dots\dots\dots \text{Eq. 3.4}$$

where

$$T(h_\phi) = \begin{bmatrix} t_{11}(h_\phi) & \cdots & t_{1K}(h_\phi) \\ \vdots & \ddots & \vdots \\ t_{K1}(h_\phi) & \cdots & t_{KK}(h_\phi) \end{bmatrix}$$

Each matrix entry $t_{jk}(h_\phi)$ is defined by

$$t_{jk}(h_\phi) = \Pr\{(\text{category } k \text{ occurs at } \mathbf{x} + h_\phi) | (\text{category } j \text{ occurs at } \mathbf{x})\} \text{ Eq. 3.5}$$

where h_ϕ represents a positive lag separation in the direction ϕ . R_ϕ is a transition rate matrix

$$R_\phi = \begin{bmatrix} r_{11,\phi} & \cdots & r_{1K,\phi} \\ \vdots & \ddots & \vdots \\ r_{K1,\phi} & \cdots & r_{KK,\phi} \end{bmatrix}$$

and will be described later in the text. Markov chains can be used to represent random systems of spatial variability as well as structured systems of spatial variability. Markov chains have found most of their application in one-dimensional systems. However, one-dimensional Markov chains can be extended to three dimensions by manipulation of one-dimensional Markov chains in the three primary directions: strike, dip, and vertical. The resulting 3-D Markov chain models can then be applied to indicator estimation and simulation techniques like indicator cokriging, sequential indicator simulation, and simulated annealing. Markov chains can be defined by a matrix of transition rates that delineate the probability of changing from one category to another based on a lag distance. These transition rate matrices have uniquely simple properties that enhance their usefulness. First, indicator variables are mutually exclusive, such as geologic units, and are defined over a region \mathbf{D} by

$$I_k(\mathbf{x}) = \begin{cases} 1, & \text{if category } k \text{ occurs at location } \mathbf{x} \\ 0, & \text{otherwise} \end{cases} \dots\dots\dots \text{Eq. 3.6}$$

where $\mathbf{x} \in \mathbf{D}$ and category $k = 1, \dots, K$. According to probability rules, the proportions of each geologic unit obey

$$\sum_{k=1}^K p_k = 1 \dots\dots\dots \text{Eq. 3.7}$$

The row sums in $T(h_\phi)$ obey

$$\sum_{k=1}^K t_{jk}(h_\phi) = 1 \quad \forall j \dots\dots\dots \text{Eq. 3.8}$$

The columns sums obey

$$\sum_{j=1}^K p_j t_{jk}(h_\phi) = p \quad \forall k \dots\dots\dots \text{Eq. 3.9}$$

where

$$0 \leq t_{jk}(h_\phi) \leq 1 \quad \forall j, k \dots\dots\dots \text{Eq. 3.10}$$

3.2 Options for Creating Markov Chains

I will discuss four alternatives for generating the Markov chains fundamental to the transition probability approach: definition of transition rates, embedded transition probabilities, embedded transition frequencies, and maximum entropy factors.

3.2.1 Rates

First, Markov chains can be defined by simply defining the matrix of transition rates. An N by N matrix of transition rates is required to define the Markov chains for a model of variability with N geologic units. A transition rate is defined as:

$$r_{jk,\phi} = \frac{\partial t_{jk}(0)}{\partial h_\phi} \quad \forall j, k \dots\dots\dots \text{Eq. 3.11}$$

Mathematical manipulations of the transition rates are accomplished in matrix form:

$$R_\phi = \begin{bmatrix} r_{11,\phi} & \cdots & r_{1K,\phi} \\ \vdots & \ddots & \vdots \\ r_{K1,\phi} & \cdots & r_{KK,\phi} \end{bmatrix}$$

Conceptually, the transition rate describes the change in the transition probability over the change in lag distance. The diagonal terms of the transition rate matrix are defined as:

$$r_{kk,\phi} = -\frac{1}{\bar{L}_{k,\phi}} \dots\dots\dots \text{Eq. 3.12}$$

where $\bar{L}_{k,\phi}$ is the average mean length of material k in the direction ϕ .

The row sums of R_ϕ must obey

$$\sum_{k=1}^K r_{jk,\phi} = 0 \quad \forall j \dots\dots\dots \text{Eq. 3.13}$$

The column sums obey

$$\sum_{j=1}^K p_j r_{jk,\phi} = 0 \quad \forall k \dots\dots\dots \text{Eq. 3.14}$$

The application of these two constraints eliminates the need to define the row and column transition rates for one entire category in the system. This category will be referred to in the remainder of the paper as the background category and will receive attention later in the report. The transition probabilities for each geologic unit with respect to each other can be directly calculated from the transition rate matrix and visa versa. An example transition rate matrix developed for a site near the southwest portion of the Lawrence Livermore National Laboratory (LLNL) in the Livermore Valley of the Coast Range in California is

$$R_z = \begin{bmatrix} -0.875 & 0.706 & 0.104 & 0.064 \\ 0.088 & -0.447 & 0.150 & 0.209 \\ 0.024 & 1.080 & -1.227 & 0.123 \\ 0.040 & 0.000 & 0.766 & -0.806 \end{bmatrix} m^{-1}.$$

The site is an alluvial fan deposit with a large proportion of fine-grained materials. The site is composed of the materials listed in Table 3.1.

Table 3.1 LLNL site characterization from core data (Carle, 1996).

#	facies	texture	%	Avg. vertical lens length
1	Debris flow	Poorly-sorted clay gravel	7	1.14
2	Flood plain	Clay and silt	56	2.24
3	Levee	Silty or clayey fine sand	19	0.82
4	Channel	Sand and gravel (rounded)	18	1.24

The average vertical lens length for each facie can be confirmed with Eq. 3.12. The correlating transition probability curves generated from this site data are represented in Figure 3.1.

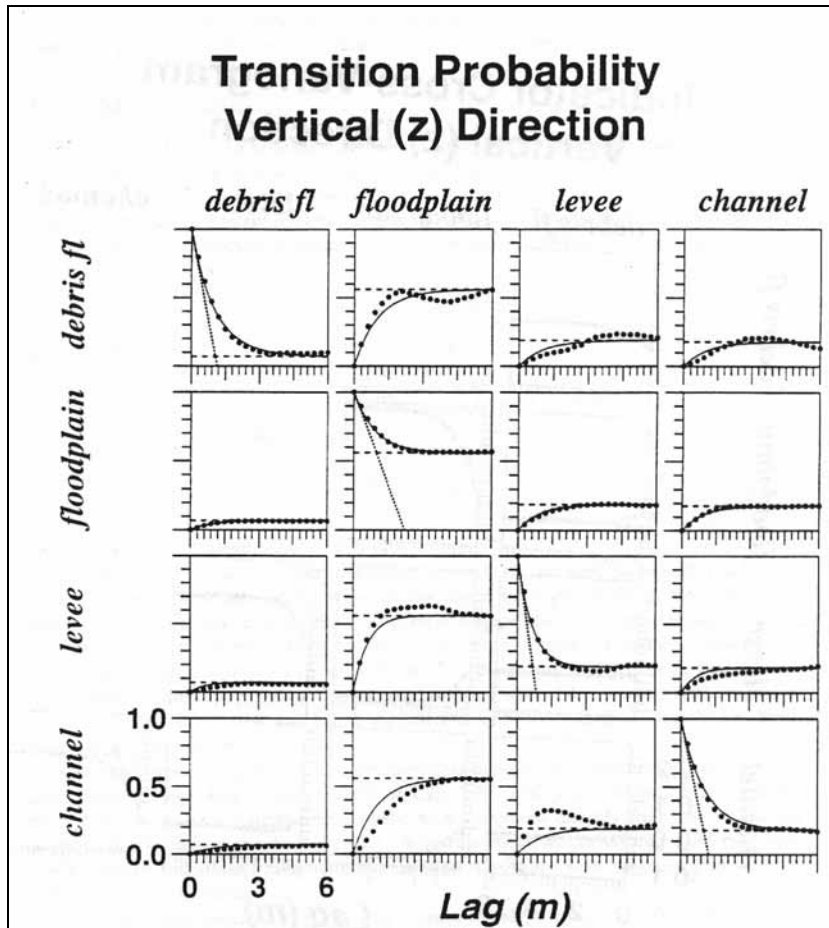


Figure 3.1 Matrix of vertical-direction transition probabilities for LLNL core data: measurements (dots) and Markov chain model (solid lines). Intersection of dashed line (tangent) with lag axis indicates mean length; dotted line indicates proportions (Carle, 1996)

This figure illustrates the unique characteristics and interpretability of the transition probability approach. First, the intersection of the dashed line (tangent to the Markov chain at lag = 0) and the lag axis indicates the average vertical mean length of the corresponding geologic unit. The dotted line indicates the proportion of each geologic unit. The dotted curve indicates the actual measurements from the borehole data, and the solid line represents the Markov chain computed from the transition rates. Conceptually, we see a much greater probability of every material transitioning to flood

plain due to the high proportion of that material. We also see a greater probability of debris flow transitioning to levee than levee transitioning to debris flow. Similar comparisons can be made for each material.

3.2.2 Embedded Transition Probabilities

Embedded transition probabilities provide a more interpretive method of defining the directional Markov chains. Figure 3.2 illustrates embedded occurrences of geologic facies.

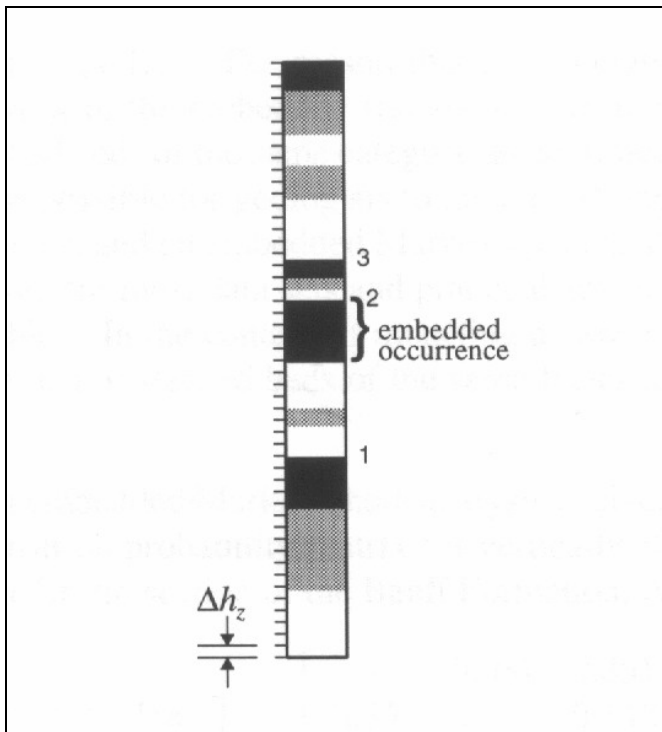


Figure 3.2 Embedded occurrences of a three-category system with 1 = white, 2 = gray, 3 = black. (Carle, 1999)

Implementation of an embedded Markov chain analysis requires minor

computations on the borehole data.

1. Forget about lag or spatial dependency and relative thickness of the beds.
2. Record the succession of “embedded occurrences,” that is, simply log each occurrence of every geologic unit in borehole data, which might look something like *ABCABACABCABABC*.
3. Tally up the transition count matrix, which for the succession above would be

$$\begin{bmatrix} - & 5 & 1 \\ 2 & - & 3 \\ 3 & 0 & - \end{bmatrix}$$

The diagonal elements are blank because “self-transitions,” e.g. from category 1 to category 1, are unobservable. That is, stacked beds of the same category are assumed not distinguishable from a single bed. The “embedded occurrence” term refers to a discrete occurrence of category 1, which may consist of either a single bed or stacked bed.

4. Divide each row by the row sum to obtain the embedded transition probabilities

$$\begin{bmatrix} - & 0.833 & 0.167 \\ 0.40 & - & 0.60 \\ 1.0 & 0 & - \end{bmatrix}$$

5. Divide each entry by the total embedded occurrences in the matrix to obtain the transition probability frequency

$$\begin{bmatrix} - & 0.357 & 0.071 \\ 0.143 & - & 0.214 \\ 0.214 & 0.00 & - \end{bmatrix}$$

Once again, an N by N matrix of embedded transition probabilities is required in each primary direction. Conceptually, an embedded transition probability describes the frequency of the occurrence of material A following material B in the borehole data or the marginal probability of material A occurring above material B.

$$\pi_{BA,z} = \Pr\{\text{material } A \text{ occurs above} | \text{material } B \text{ occurs below}\}.. \text{Eq.}$$

3.15

The embedded transition probabilities are independent of the actual length of occurrence of a given material. And the off-diagonal entries of the embedded transition probability matrix satisfy

$$\sum_{k=1}^K \pi_{jk,z} = 1 \dots\dots\dots \text{Eq. 3.16}$$

Once again, using the LLNL site with *debris flow, floodplain, levee, and channel* facies, an embedded transition probability matrix can be constructed as

$$\pi_z = \begin{bmatrix} - & 0.803 & 0.124 & 0.073 \\ 0.176 & - & 0.390 & 0.434 \\ 0.026 & 0.846 & - & 0.128 \\ 0.045 & 0.058 & 0.896 & - \end{bmatrix}$$

The embedded transition probability approach provides a more conceptual approach to generating transition probabilities because embedded occurrences can be easily tallied from the borehole data. For example, $\pi_{43,z} \gg \pi_{41,z}$ and $\pi_{43,z} \gg \pi_{42,z}$, which indicates that *levee* tends to occur above *channel*. Once again, application of the probability properties excludes the need to define the row and column entries for the background

category. Embedded transition probabilities can easily be converted to transition rates with

$$r_{jk,z} = \frac{\pi_{jk,z}}{L_{j,z}} \dots\dots\dots \text{Eq. 3.17}$$

The transition probability curves associated with the embedded transition probabilities above are illustrated in Figure 3.3.

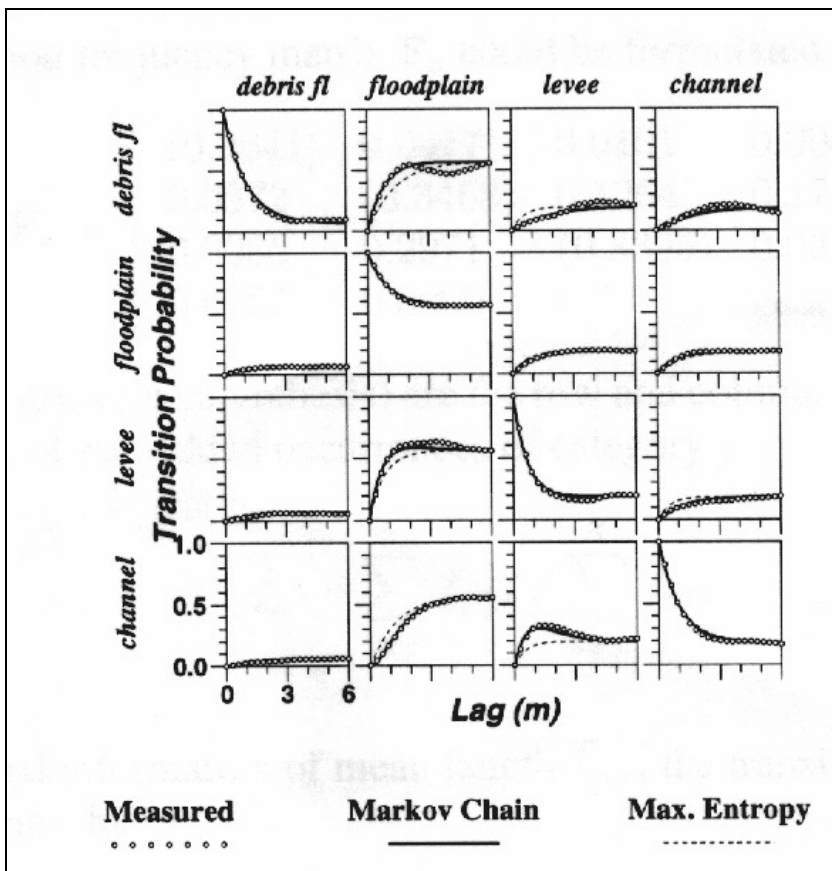


Figure 3.3 Markov chain model fit to matrix transition probability data with the embedded transition probability (solid line) and the maximum entropy (dashed line) approaches. Actual measured data is illustrated with a dotted line. (Carle, 1999)

The embedded transition probability approach provides a conceptual and accurate method of generating 1-D Markov chains.

3.2.3 Embedded Transition Frequencies

An embedded Markov chain analysis can also be performed based on embedded transition frequencies, which are defined as

$$f_{AB,z} = \Pr\{\text{material } B \text{ occurs above and material } A \text{ occurs below}\} \dots \text{Eq. 3.18}$$

An embedded transition frequency matrix for the LLNL data can be formulated as

$$F_z = \begin{bmatrix} (0.0841) & 0.0677 & 0.0101 & 0.0063 \\ 0.0672 & (0.3468) & 0.1264 & 0.1713 \\ 0.0085 & 0.2971 & (0.3395) & 0.0340 \\ 0.0085 & 0.000 & 0.2031 & (0.2115) \end{bmatrix}$$

The diagonal terms of the matrix correspond to the marginal frequencies of embedded occurrence of category j such that:

$$f_{j,z} = \sum_{k \neq j}^K f_{jk,z} = \sum_{k \neq j}^K f_{kj,z} \dots \text{Eq. 3.19}$$

Similar to the preceding methods, the probabilistic properties of the matrix make the definition of one row and column not required in the matrix. Furthermore, transition frequencies can be converted to transition rates by:

$$r_{jk,z} = \frac{f_{jk,z}}{f_{j,z} L_{j,z}} \quad \forall k \neq j \dots \text{Eq. 3.20}$$

The transition probability curves using the embedded transition frequencies, F_z , are

illustrated in Figure 3.4.

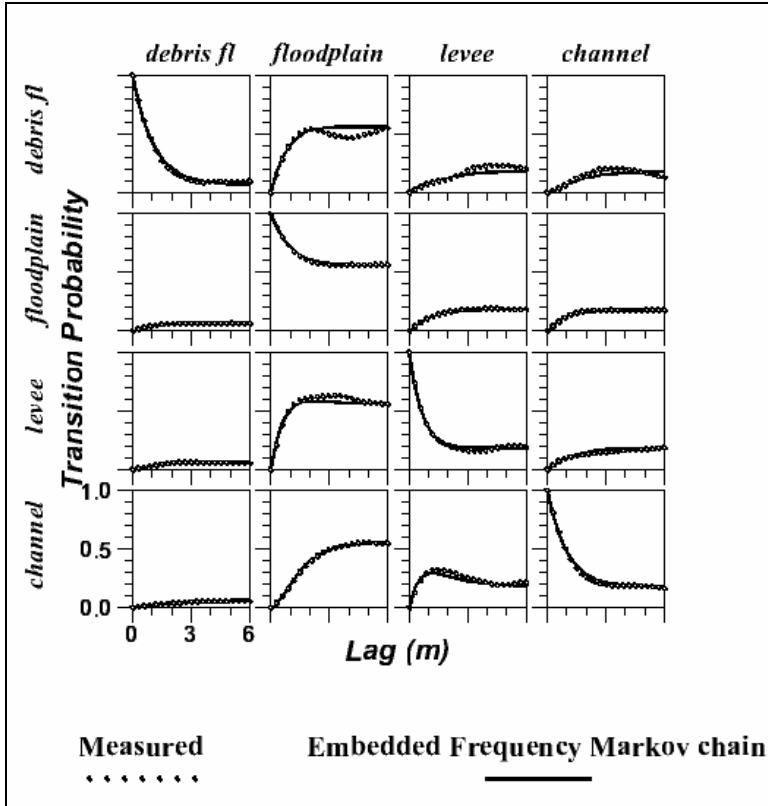


Figure 3.4 Transition probability curves from measured data and Markov chain generated from embedded transition frequencies.

3.2.4 “Independent” or “Maximum Entropy” (Disorder) Transition Frequencies

Geologic deposition typically presents some juxtapositional tendencies in the bedding sequence. This order or disorder in a direction ϕ can be defined by entropy,

S_ϕ , of facie to facie transition frequency $f_{jk,\phi}$. S_ϕ is defined as:

$$S_\phi = -\sum_J \sum_k f_{jk,\phi} \ln f_{jk,\phi} \dots \dots \dots \text{Eq. 3.21}$$

However, consider an area that displays complete disorder in the bedding sequence.

$$F_z^{(S \max)} = \begin{bmatrix} (0.0841) & 0.0377 & 0.0311 & 0.0152 \\ 0.0377 & (0.3671) & 0.2196 & 0.1075 \\ 0.0311 & 0.2196 & (0.3396) & 0.0888 \\ 0.0152 & 0.1075 & 0.0888 & (0.2115) \end{bmatrix}$$

Maximum disorder would be void of any juxtapositional tendencies and would only depend on mean lengths and proportions. As a result, $F_z^{(S \max)}$ is symmetric with the diagonal entries equal to the row/column off-diagonal totals or marginal frequencies. By comparison to this case, one could determine to what extent a site exhibits juxtapositional tendencies. For example, if an observed transition frequency $f_{43,z}$ is greater than $f_{43,z}^{(S \max)}$, then one might determine that there is a tendency for material 3 to occur above material 4. One can develop a transition rate matrix from a maximum transition frequency matrix, $F_z^{(S \max)}$, with three steps. First, one would determine the mean lengths and proportions for each geologic unit in the corresponding direction. Second, one would compute the maximum entropy frequency matrix $F_z^{(S \max)}$. Third, one would establish the off-diagonal transition rates relative to the maximum entropy transition rates $R_z^{(S \max)}$, which can be calculated from $F_z^{(S \max)}$ by

$$r_{jk,z}^{(S \max)} = \frac{f_{jk,z}^{(S \max)}}{L_{j,z} \sum_{k \neq j} f_{jk,z}^{(S \max)}} \dots \dots \dots \text{Eq. 3.22}$$

In practice, maximum entropy application involves maximum entropy coefficients.

Markov chains are quantified with a mean length for the diagonal entries and a

coefficient equal to the actual transition rate divided by the maximum entropy transition rate. These coefficients represent maximum entropy factors.

$$R_z^{(S \max)} = \begin{bmatrix} \frac{-1}{L_{1,z} = 1.15} & 1.80r_{12,z}^{(S \max)} & \mathbf{0.32}r_{13,z}^{(S \max)} & 0.41r_{14,z}^{(S \max)} \\ 1.78r_{21,z}^{(S \max)} & \frac{-1}{L_{1,z} = 2.27} & 0.58r_{23,z}^{(S \max)} & 1.59r_{24,z}^{(S \max)} \\ 0.27r_{31,z}^{(S \max)} & \mathbf{2.35}r_{32,z}^{(S \max)} & \frac{-1}{L_{1,z} = 0.82} & 0.45r_{34,z}^{(S \max)} \\ \mathbf{0.11}r_{41,z}^{(S \max)} & 0.00r_{42,z}^{(S \max)} & 2.29r_{43,z}^{(S \max)} & \frac{-1}{L_{1,z} = 1.24} \end{bmatrix}$$

A maximum entropy factor of 1.0 displays maximum disorder in depositional tendencies. A factor greater than unity indicates that the two categories tend to occur next to each other. A factor less than unity would infer the opposite. The transition probability curves generated from $R_z^{(S \max)}$ are shown in Figure 3.5.

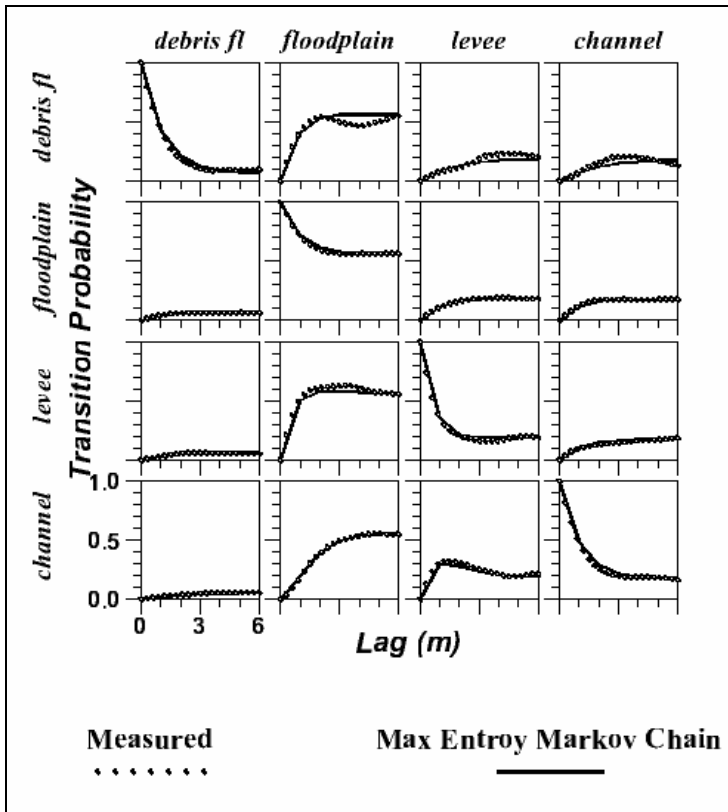


Figure 3.5 Transition probability curves generated from measured data and maximum entropy factors.

Two sets of transition probability curves were generated from two slightly different maximum entropy factor matrices to demonstrate the application of these factors. The maximum entropy factors that are bolded and italicized in $R_z^{(S \max)}$ were adjusted slightly as follows:

$$R_z^{(S \max)} = \begin{bmatrix} \frac{-1}{L_{1,z} = 1.15} & 1.80r_{12,z}^{(S \max)} & 0.5r_{13,z}^{(S \max)} & 0.41r_{14,z}^{(S \max)} \\ 1.78r_{21,z}^{(S \max)} & \frac{-1}{L_{1,z} = 2.27} & 0.58r_{23,z}^{(S \max)} & 1.59r_{24,z}^{(S \max)} \\ 0.27r_{31,z}^{(S \max)} & 0.35r_{32,z}^{(S \max)} & \frac{-1}{L_{1,z} = 0.82} & 0.45r_{34,z}^{(S \max)} \\ 1.56r_{41,z}^{(S \max)} & 0.00r_{42,z}^{(S \max)} & 2.29r_{43,z}^{(S \max)} & \frac{-1}{L_{1,z} = 1.24} \end{bmatrix}$$

The transition probability curves for *channel->floodplain* and *channel->levee* in Figure 3.6 display a slight change from their counterparts in Figure 3.5 due to the variation in the maximum entropy factors.

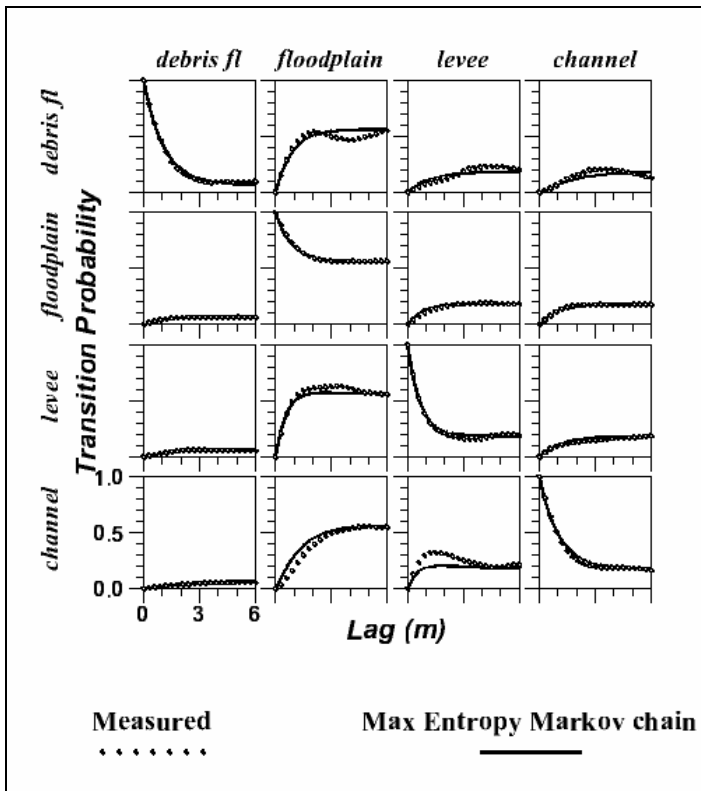


Figure 3.6 Transition probability curves generated from measured data and maximum entropy Markov chains.

This concept is particularly useful in interpreting the juxtapositional tendencies of an existing site by comparing the actual transition rates with those produced with the maximum entropy method.

3.3 Conceptual Development

One of the unique characteristics of the transition probability approach is the conceptual framework that enhances Markov chain development. The conceptual framework of this approach refers to the ability to logically incorporate intuitive, geologic parameters such as mean length and proportion into this mathematical algorithm. It is this characteristic that has elevated interest and sponsored further research. Furthermore, this characteristic has made the transition probability approach more attractive than traditional indicator methods like cross-variograms because they require abundant data to enable curve-fitting. The conceptual framework that the transition probability approach is built on enables users to generate relationships between different materials based on geologic features of the study site. Some of these features include a background material, mean lens lengths, material proportions, asymmetry, juxtapositional tendencies, and 3-D Markov chains.

3.3.1 Background Material

Application of the transition probability approach involves the designation of a background material. The probabilistic constraints of the Markov chains make it unnecessary to quantify data for one category. Not only is it unnecessary, but it is futile

to do so because values will be overwritten in order to satisfy constraints. Conceptually, the background material can be described as the material that “fills” in the remaining areas not occupied by other units. For example, in a fluvial depositional system, a floodplain unit would tend to occupy area not filled with higher-energy depositional units and would therefore be a logical choice for the background material. The simplification of a background material can be illustrated with a three-unit system. A nine-entry system reduces to a four-entry system with the application the second material as the background material.

$$\begin{bmatrix} \# & b & \# \\ b & b & b \\ \# & b & \# \end{bmatrix}$$

Mathematically, any material could be selected as the background material.

3.3.2 Material Proportions

Material proportions are quantified as the proportion of a given material at a particular site. This site characteristic is integral in the transition probability approach. Graphically, the material proportions correlate to the sill on the transition probability curve as illustrated in Figure 3.1. The plots in each column have approximately the same sill, which correlates to the proportion of the corresponding material. As $h_z \rightarrow \infty$, the transition probability approaches the proportion.

3.3.3 Mean Length

Conceptually, the mean length, $\overline{L_{k,\phi}}$ of the category k in the direction ϕ is defined as the total length of the category k in direction ϕ divided by the number of embedded occurrences.

$$\overline{L_{k,\phi}} = \frac{\text{total length of } k \text{ in direction } \phi}{\text{number of embedded occurrences of } k} \dots\dots\dots \text{Eq. 3.23}$$

In other words, the mean length, $\overline{L_{k,\phi}}$ corresponds to the “mean thickness” in any given direction. Mathematically, $\overline{L_{k,\phi}}$ relates to a diagonal transition probability $t_{kk}(h_\phi)$ by

$$-\frac{\partial t_{kk}(0)}{\partial h_\phi} = \frac{1}{\overline{L_{k,\phi}}} \dots\dots\dots \text{Eq. 3.24}$$

or

$$r_{kk,\phi} = -\frac{1}{\overline{L_{k,\phi}}} \dots\dots\dots \text{Eq. 3.25}$$

As illustrated in Figure 3.1, the mean length will correspond to the slope of the transition probability curve at a $h_\phi = 0$. Mean lengths can be used in model development by establishing the diagonal terms of the rate matrix given the knowledge of material mean thickness. Alternatively, one can interpret the mean length from the transition probability model.

3.3.4 Asymmetry

One functional advantage of the transition probability approach over traditional approaches is the ability to model asymmetry. Cross-variograms are inherently

symmetric such that $\gamma_{jk}(h) = \gamma_{jk}(-h)$ whereas transition probabilities allow for $t_{jk}(h) \neq t_{jk}(-h)$. This capability is particularly useful when modeling systems with vertical juxtapositional tendencies like fining-upward tendencies in fluvial deposits.

3.3.5 Sparse or No Data

Typical data sets yield noisy transition probabilities, especially in the lateral directions. The traditional indicator approaches that utilize empirical curve-fitting can be more complicated than the data warrants. However, the transition probability approach addresses mathematical and probability theory while providing an interpretive framework for model development. The transition probability approach is inherently simpler because it assumes that spatial variability depends only on the nearest location. Furthermore, in the absence of borehole data one can still develop Markov chains from basic geologic data: by estimating the proportion, mean length, and juxtapositional tendencies of the materials.

Markov chains are even conducive to the development of models without any data at all. Markov chains can be developed from purely conceptual geologic data and will result in an unconditioned realization of spatial variability. This approach is useful when conducting hypothetical simulations where some type of model heterogeneity is desired. The Markov chain methods particularly conducive to these situations are the embedded transition probability, transition rate, and maximum entropy methods.

3.4 Multidimensional Markov Chains

Typical site stratigraphic data is conducive to developing vertical spatial variability. However, rarely is the quantity of data adequate to develop an accurate model in the lateral directions. The combination of Walther's Law and the transition probability approach allows for a logical method of developing lateral spatial variability from vertical spatial variability. Walther's Law states that vertical successions of deposited facies represent the lateral succession of environments of deposition. Therefore, a logical method of generating a 3-D model of spatial variability would be first to develop a 1-D Markov chain in the vertical direction based on site data, assuming there is such data. Second, using Walther's Law and geologic knowledge, one can develop lateral, strike and dip, Markov chains of spatial variability. However, one issue that arises when applying vertical transition trends to lateral directions is how to cope with asymmetric vertical trends like fining upwards. For example, if in the vertical direction, sand tends to deposit on gravel as would be typical in a fluvial deposition, there will be a transition rate associated with the transition of sand->gravel and gravel->sand. The transition of gravel->sand will be greater than sand->gravel because of the fining upward trend. However, in the lateral direction, which of these transition rates should be applied? Although the trend of sand next to gravel remains in the lateral direction, the transition rates of sand->gravel and gravel->sand should be equivalent or symmetric as defined by

$$r_{jk} = \frac{p_k}{p_j} r_{kj} \dots\dots\dots \text{Eq. 3.26}$$

Carle (Carle, 1996) addresses this issue but simply indicates that lateral rates can be inferred from the vertical rates without explicitly suggesting how this should be

accomplished. As part of this research, we developed a strategy for averaging the vertical transition rates to come up with the lateral rates. This averaging technique involves a three-step procedure:

1. Compute the lower-half rate (R_{LS}) that will satisfy symmetry with the upper-half rate (R_u) for the vertical data using Eq. 3.26.
2. Set the lower-half rate for the lateral direction (R'_L) equal to $\frac{(R_L + R_{LS})}{2}$
3. Compute the upper-half rate for the lateral direction R'_U that will satisfy symmetry with the new lower-half rate (R'_L) using Eq. 3.26.

In addition to the averaging technique applied to the off-diagonal terms of the lateral transition rates, an adjustment was made to the diagonal terms of the rate matrix. It will be remembered that the diagonal terms correlate to the average mean lengths by Eq. 3.25. Generally in depositional patterns, lateral mean lengths are larger than their counterparts. Therefore, the lateral mean length generally need to be increased by a factor F prescribed by the user. This factor F is equal to the ratio of the lateral mean lengths/vertical mean lengths. If the mean length for category K increases by a factor F , the corresponding transition rate will decrease by that same factor F according to Eq. 3.25. And the row sum for category K must therefore decrease by the factor F according to Eq. 3.13. For example, given the vertical transition rate matrix R_z , if the lateral mean lengths are all increased by a factor $F = 10.0$, the resulting lateral rate matrix R_x with the ratio affect only (without the average technique applied) would be

$$R_z = \begin{bmatrix} -0.296 & 0.092 & 0.021 & 0.183 \\ 0.026 & -0.103 & 0.012 & 0.065 \\ 0.019 & 0.133 & -0.262 & 0.109 \\ 0.037 & 0.103 & 0.025 & -0.165 \end{bmatrix} \rightarrow R_x = \begin{bmatrix} -0.0296 & 0.0092 & 0.0021 & 0.0183 \\ 0.0026 & -0.0103 & 0.0012 & 0.0065 \\ 0.0019 & 0.0133 & -0.0262 & 0.0109 \\ 0.0037 & 0.0103 & 0.0025 & -0.0165 \end{bmatrix}$$

Eq. 3.27 illustrates the extrapolation of a vertical transition rate matrix to a lateral transition rate matrix with application of both the averaging technique as well as the adjustment for the ratio of vertical to lateral average mean lengths.

$$R_z = \begin{bmatrix} \dots & \dots & R_U & \dots \\ \dots & \dots & \dots & \dots \\ R_L & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \rightarrow R_x = \begin{bmatrix} \dots & \dots & \frac{R_U}{F} & \dots \\ \dots & \dots & \dots & \dots \\ \frac{R_L}{F} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \dots \text{Eq. 3.27}$$

Once the transition rate matrices have been defined for each of the three primary directions, the three 1-D Markov chains can then be interpolated to any direction with

$$|r_{jk,\phi}| = \sqrt{\left(\frac{h_x}{h_\phi} r_{jk,x}\right)^2 + \left(\frac{h_y}{h_\phi} r_{jk,y}\right)^2 + \left(\frac{h_z}{h_\phi} r_{jk,z}\right)^2} \quad \forall j, k \neq \beta \dots \text{Eq. 3.28}$$

where β denotes the background category, h_x , h_y and h_z are the x , y and z direction components of $h_\phi = \sqrt{h_x^2 + h_y^2 + h_z^2}$. For the negative lag vector components, say h_{-x} , entries from the rate matrix R_{-x} corresponding to the opposite direction $-x$ are defined by

$$r_{jk,-x} = \left(\frac{p_k}{p_j}\right) r_{kj,x} \dots \text{Eq. 3.29}$$

3.5 Eigensystem Analysis

MCMOD performs an eigensystem analysis because of the mathematical requirements of developing a continuous-lag Markov chains as a geostatistical model of spatial variability. This process requires the following calculations:

- Evaluate the matrix exponential form of Markov chain give by Eq. 3.4
- Evaluate the matrix logarithm of a transition probability matrix given by Eq. 3.30
- Convert a discrete-lag Markov chain to a continuous-lag Markov chain by combing Eq. 3.4 and Eq. 3.30

$$\mathbf{R}_\phi = \frac{\ln[\mathbf{T}(\Delta h_\phi)]}{\Delta h_\phi} \dots\dots\dots \text{Eq. 3.30}$$

where \mathbf{R}_ϕ is a transition rate matrix, \mathbf{T} is a transition probability matrix, and Δh_ϕ is the change in the lag in the direction ϕ . Equations 3.4 and 3.30 cannot be computed directly from the matrix entries. Eigenvalues for both \mathbf{R}_ϕ and \mathbf{T} must be computed in the analysis.

3.6 Developing a Simulation

The generation of an indicator Kriging simulation via the transition probability geostatistics approach involves three general steps:

- 1) Generate Markov chains in the primary directions

- 2) Extrapolate 1-D Markov chains to a 3-D system
- 3) Establish an initial simulation of material identifications based on the Markov chains and then condition the simulation to borehole data.

In this research, we utilize the T-PROGS software package developed by Steven F. Carle. The T-PROGS software consists of three separate algorithms to perform each of these steps, GAMEAS, MCMOD, and TSIM respectively.

GAMEAS is an algorithm that processes borehole data and determines geologic characteristics such as material proportions and transition probability curves in a given direction. Typically, GAMEAS is only applicable to vertical data because lateral data is too sparse for accurate results. Input for GAMEAS includes a parameter file, (Figure 3.7), that has site characteristics such as material names, lag identification, and statistical information.

START OF PARAMETERS	
data.eas	/input file
1 2 3	/x,y,z columns
4 4 5 6 7	/nvar, var 1,2,3,...columns
-1. 2.	/vmin, vmax
datatpz.eas	/output file
41	/# lags
0.3	/lag spacing
0.15	/lag tolerance
1	/ndir
0.0 90. 0.25 -90. 22.5 0.25	/az, daz, azbw; dip,.....
16	/# of bivariate statistics
1 1 11	/j, k, 11=tp
1 2 11	
1 3 11	
1 4 11	
2 1 11	
2 2 11	
2 3 11	
2 4 11	

```

3 1 11
3 2 11
3 3 11
3 4 11
4 1 11
4 2 11
4 3 11
4 4 11

```

Figure 3.7 Sample parameter file for GAMEAS. (Carle, 1999)

Furthermore, GAMEAS requires a site data file that contains a material identification at each x, y, and z location, as shown in Figure 3.8.

```

Data
7
x = easting
y = northing
z = elevation above mean sea level
1 = debris flow
2 = floodplain
3 = levee
4 = channel
2132.8      2487.4      137.07      0  1  0  0
2132.8      2487.4      136.77      0  1  0  0
2132.8      2487.4      136.47      0  1  0  0
2132.8      2487.4      136.17      0  1  0  0
2132.8      2487.4      135.87      1  0  0  0
2132.8      2487.4      135.57      1  0  0  0
2132.8      2487.4      132.27      0  1  0  0
2132.8      2487.4      131.97      0  1  0  0
2576.2      2695.5      186.48      0  1  0  0
2576.2      2695.5      182.28      0  0  0  1
2576.2      2695.5      181.98      0  0  0  1
2576.2      2695.5      181.68      0  0  0  1
2576.2      2695.5      181.38      0  0  0  1
2576.2      2695.5      181.08      0  1  0  0
2576.2      2695.5      175.98      1  0  0  0
.
.
.

```

Figure 3.8 Sample data file. (Carle, 1999)

Hence, a typical application of GAMEAS would involve the creation of the parameter and data file from vertical borehole data.

GAMEAS parses through the borehole data and determines the proportions of each material as well as the transition relationships between each material. The transition probability curves generated by GAMEAS provide a model from which geologic data such as proportions, vertical mean lengths, and transition probability rates can be extrapolated. The proportions, mean lengths, and transition rates assist in the generation of vertical Markov chains required in MCMOD and TSIM. Other geologic information like embedded transition probabilities and frequencies can be inferred from the borehole data. The vertical transition probability parameters generated with the aid of GAMEAS can then be used to develop Markov chains in the strike/dip directions. This can be accomplished with application of Walther's Law or utilization of more interpretive methods of generating Markov chains like the maximum entropy approach or the embedded transition approach. The computation of Markov chains in the three primary directions enables the generation of a 3-D Markov chain with MCMOD.

MCMOD is a utility that converts three individual 1-D Markov chains into a single 3-D Markov chain system. MCMOD also requires a parameter file that contains some geologic and grid data and the definition of each 1-D Markov chain in the primary directions (Figure 3.9). The definition of each Markov chain includes the method used to generate the Markov chain and the corresponding matrix of rates.

4	/# of categories
0.565 0.19 0.179	/proportions
2	/background category
../lnl/tp/mcmod1_21097.dbg	/name of debugging file
../lnl/tp/tpxyz1.bgr	/output file for 3-D model

../lnl/tp/det.bgr	/output file for determinant
0.05 0.05 0.05	/determinant extent for 3-D model
3.0 10. 0.3	/dx, dy, dz for 3-D model
../lnl/tp/lnltpxm.eas	/X-direction output file
200 1.	/X-direction: # lags, spacing
1	/1=r, 2=d, 3=etp, 4=etf, 5=maxe
-0.125 0. -1. -1.	/row 1 transition rates
0. 0. 0. 0.	/row 2 transition rates
0.0042 0. -0.167 -1.	/row 3 transition rates
0.004 0. 0.084 -0.1	/row 4 transition rates
../lnl/tp/lnltpym.eas	/Y-direction output file
200 2.5	/Y-direction; # lags, spacing
1	/1=r, 2=d, 3=etp, 4=etf, 5=maxe
-0.042 0. 0.0036 0.0022	/row 1 transition rates
0. 0. 0. 0.	/row 2 transition rates
0.0013 0. -0.05 0.016	/row 3 transition rates
0.0008 0. 0.017 -0.02	/row 4 transition rates
../lnl/tp/lnltpzm.eas	/Z-direction output file
200 0.1	/Z-direction; # lags, spacing
2	/1=r, 2=d, 3=etp, 4=etf, 5=maxe
../lnl/tp/data.eas	/data file
2	/lag #

Figure 3.9 Sample parameter file for MCMOD (Carle, 1999)

The output from MCMOD is a 3-D transition probability model file and a determinant file that are used in TSIM.

TSIM is a utility that builds alternative, equally probable spatial distributions of a random variable that honor hard data at specified locations. TSIM has two major functions: First, it establishes an initial configuration with the sequential indicator simulation (SIS) algorithm using a transition probability-based cokriging estimate. SIS includes four basic steps at each “unsimulated” nodal location chosen sequentially along a random path.

1. A search is conducted for nearby data locations.
2. A local conditional probability distribution is estimated by cokriging values of the nearby conditioning data and already “simulated” data.

3. A random number is chosen which determines the category from the probability distribution.
4. The simulation is updated. (Carle, 1996)

These four steps are repeated at each “unsimulated” node visited along the random path, with each cokriging estimate conditioned to the sequentially updated version of the simulation.

After completing the SIS, TSIM iteratively improves the initial configuration by the simulated quenching (zero-temperature annealing) algorithm. The quenching process iteratively minimizes an objective function, O that computes the quality of the outcome as compared to the measured data by

$$O = \sum_{l=1}^K \sum_{j=1}^K \sum_{k=1}^K (t_{jk}(h_l)_{MEAS} - t_{jk}(h_l)_{MOD})^2 \dots\dots\dots \text{Eq. 3.31}$$

Figure 3.10 illustrates the improvement realized from the quenching process.

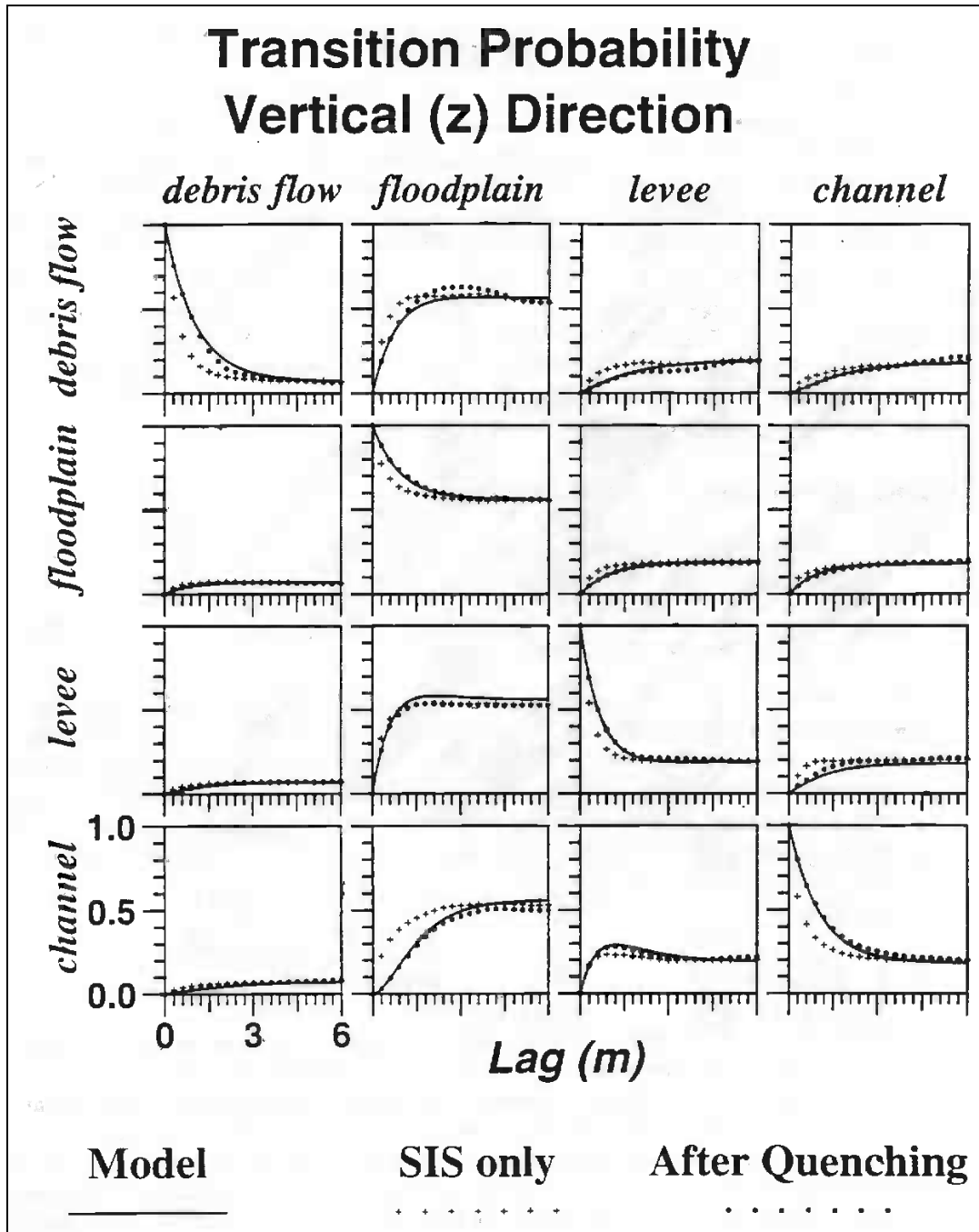


Figure 3.10 Vertical (z)-direction transition probabilities for models of spatial variability with SIS only and with SIS and quenching. (Carle, 1996)

The two processes are mutually exclusive because the SIS step does not accurately honor known data and the simulated quenching cannot function without an initial configuration.

The product of TSIM is a 3-D array of material identification numbers corresponding to

the grid information entered in the parameter file illustrated in Figure 3.11. The input to TSIM includes the data file used in GAMEAS, a parameter file, and the two output files of MCMOD: a 3-D transition probability model file and a determinant file.

4	/number of categories
0.066 0.565 0.19 0.179	/proportions
../lnl/sim/simxyz.bgr	/output file
1	/output format: 1 = binary, 2 = ascii
1	/debugging level
tpsim.dbg	/debugging file
4175	/seed
1	/number of realizations
1966.3 -20 3.	/xcenter, nx+, xsiz
3023.5 -20 10.0	/ycenter, ny+, ysiz
142.07 -20 0.3	/zcenter, nz+, zsiz
1 4	/ndmin, ndmax
1	/ibasis: 0 = cov, 1 = tp
0.01	/wratio
../lnl/tp/tpxyz.bgr	/trans. Prob. Model file
../lnl/tp/det.bgr	/determinant file
../lnl/data/data.eas	/input data file
0. 0.	/azimuths: coord, true
0. 0.	/dip: coord, true
junkaz. bgr	/azimuth int*1 file
junkdip.bgr	/dip int*1 file
4 0.0001 -1	/maxit; tol; -1 = no dcl, 1 = lag1
0.4	/limit by determinant

Figure 3.11 Sample TSIM parameter file. (Carle, 1999)

Thus the final product of these three algorithms (GAMEAS, MCMOD, and TSIM) is a set of N 3-D arrays of stochastically-generated geologic units that honor known data. Sample applications of the T-PROGS software will be illustrated in Chapter 6—Case Study.

4 INDICATOR ARRAYS TO HUF ARRAYS

Using transition probability geostatistics with MODFLOW models results in two basic limitations. First, the underlying indicator Kriging equations used by the T-PROGS software are formulated such that the MODFLOW grid must have uniform row, column, and layer widths. The row width can be different from the column width, but each row must have the same width. This results in a uniform orthogonal grid. While MODFLOW grids are orthogonal in x and y, the layer thickness is allowed to vary on a cell by cell basis. This makes it possible for the layer boundaries to accurately model the ground surface and the tops and bottoms of aquifer units. If a purely orthogonal grid is used, irregular internal and external layer boundaries must be simulated in a stair-step fashion either by varying material properties or by activating/inactivating cells via the IBOUND array. A second limitation is that in order to get a high level of detail in the simulated heterogeneity, the grid cell dimensions are generally kept quite small. This can result in difficulties in the vertical dimension. The large number of layers with small layer thicknesses near the top of the model generally ensures that many of the cells in this region will be at or above the computed water table elevation (for simulations involving unconfined aquifers). As a result, these cells will undergo many of the numerical instabilities and increased computational effort issues associated with cell wetting and drying.

4.1 Hydrogeologic Unit Flow (HUF) Package

The new Hydrogeologic Unit Flow (HUF) package is an alternative to the Block-Centered Flow (BCF Package) and the Layer Property Flow (LPF Package) in MODFLOW 2000. Each of these packages is used to compute cell-to-cell conductances from the layer geometry and aquifer properties. These packages manage the flow parameters required to solve the governing flow equation in MODFLOW. The HUF algorithm was developed by Evan Anderman (Anderman, 2000). This package includes a set of arrays defining the model stratigraphy in a grid independent fashion. The stratigraphy data are defined using a set of elevation and thickness arrays. These arrays have the same number of rows and columns as the MODFLOW grid but are independent of the MODFLOW grid in the vertical direction. The first array defines the top elevation of the model. The remaining arrays define the thicknesses of a series of hydrogeologic units, starting at the top and progressing to the bottom of the model. For each array of thicknesses, many of the entries in the array may be zero. This makes it possible to simulate complex heterogeneity, including pinchouts and embedded lenses that would be difficult to simulate with the LPF and BCF packages. When MODFLOW runs, it parses through each MODFLOW grid cell and determines the percentage of each HUF layer with its corresponding hydraulic conductivity and computes an equivalent hydraulic conductivity and applies it to the entire MODFLOW grid cell. The hydraulic governing equation computes the heads based on these equivalent hydraulic conductivities.

The HUF package has some unique advantages over the BCF and LPF packages in MODFLOW. First, property arrays can be defined independently of the grid.

Therefore, soil heterogeneity and detail can be maintained in the model without increasing the cell density in the grid. This reduces computational time and memory requirements. Furthermore, representing detailed soil heterogeneity with a grid approach would require thin layers. Thin layers near the top of the grid poses a problem in MODFLOW because of the potential for the cells in those layers to go dry. If the cells go dry at any time while MODFLOW is calculating a solution, the cells become inactive and remain dry. Therefore, cells that go dry might not accurately represent the hydraulic head in that cell. MODFLOW has a “cell rewetting” option that rewets cells after going dry, but this presents more complexity in the solution.

4.2 T-PROGS to HUF Conversion Algorithm

The primary goal of my research was to develop an algorithm for merging output from the T-PROGS software to input data for the HUF package. Using this algorithm, it is possible to generate stochastic indicator simulations of a study area with or without borehole data and read them into HUF layers, thus combining the benefits of the HUF package, transition probability geostatistics, and stochastic simulations. The basic approach used by the algorithm is to overlay a dense background grid on the MODFLOW grid and run T-PROGS on the background grid. A set of HUF arrays is then extracted from the background grid for use with the MODFLOW model. The main steps of the algorithm are as follows:

1. Create a MODFLOW grid.
2. Create a Background grid with a higher grid density in the vertical direction relative to the MODFLOW grid.

3. Run T-PROGS to get a set of indicator arrays on the background grid.
4. Convert the T-PROGS output from the background grid to the HUF input arrays.

The end result of this conversion process is N sets of HUF input arrays, each array corresponding to one 3D indicator array from the T-PROGS simulation. These sets can then be used as input to a stochastic flow simulation. Each of these steps will now be described in more detail.

4.2.1 Create a MODFLOW Grid

The first step is to create a MODFLOW grid with the desired number of layers. The user performs this step, but the remainder of the steps are performed internally. The layer elevations are interpolated to match the aquifer boundaries. The row and column widths are uniform but the layer thicknesses may vary from cell to cell. The MODFLOW grid can have curvilinear layer boundaries.

4.2.2 Create a Background Grid

The next step is to create a background grid that encompasses the MODFLOW grid. The rows and columns of the background grid match the MODFLOW grid but the layer thicknesses are uniform and relatively thin, resulting in a much greater number of layers than the MODFLOW grid. The user specifies the number of layers incorporated into the background grid. The two grids are identical in plan view, but in side view the background grid encompasses an area that bounds the MODFLOW grid. If the MODFLOW grid is curvilinear, then the background grid will potentially have cells that

exist above and below the MODFLOW grid in a given column. This becomes important when the indicator arrays generated with T-PROGS are interpolated to the MODFLOW grid.

4.2.3 Run T-PROGS

Once the background grid is generated, a T-PROGS simulation is performed to get a set of indicator arrays on the background grid. A brief review of this process is given below. A detailed description of this process is given in Appendix B—GMS/T-PROGS Interface.

1. Initialize a T-PROGS simulation. Select materials included in the simulation (if boreholes exist, materials are inherited from the boreholes) and the background material.
2. Define the vertical Markov chain with geologic intuition. If borehole data exist, the T-PROGS utility, GAMEAS can aid in this process.
3. Define the lateral (strike/dip) Markov chains from the vertical Markov chain.
4. Specify the number of layers in the background grid and run T-PROGS with the HUF option selected.

4.2.4 Convert the T-PROGS Output

Once the indicator data have been generated on the background grid, each of the indicator arrays is transferred from the background grid to a set of HUF elevation/thickness arrays. As mentioned above, the row/column configuration is identical for both grids. Each column of indicator data corresponds to a column in the MODFLOW grid. Therefore, each column of indicator data is applied to its corresponding column in the MODFLOW grid with the following process illustrated with

Figure 4.1. The colors correlate to HUF arrays defined with hydrogeologic units in a background grid, and the MODFLOW grid is outlined with black cells.

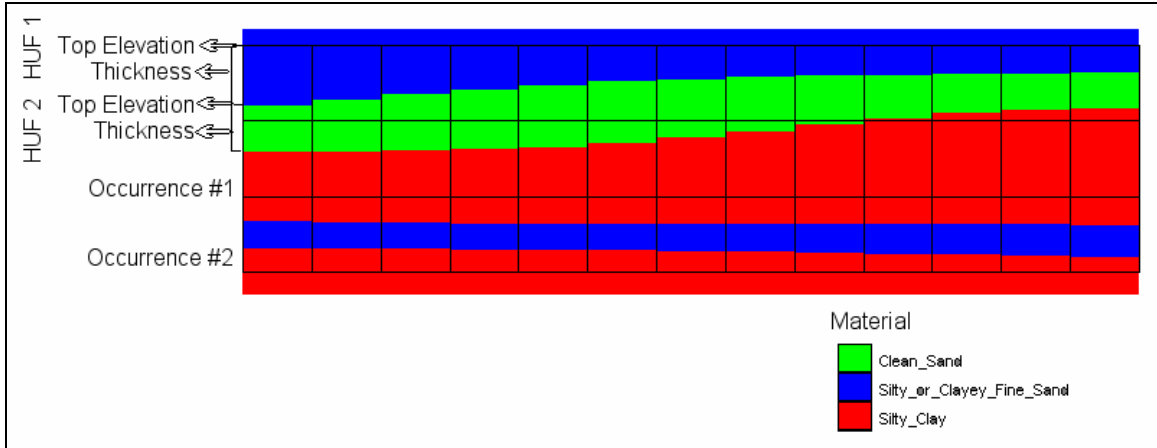


Figure 4.1 Illustration of T-PROGS --> HUF.

Consider the first grid column from the left in Figure 4.1. Since the top of the background grid in a particular column may not be the same as the MODFLOW grid, the elevation of the top of the MODFLOW grid is identified in the background grid, and the material identification is extracted at that location (Silty_or_Clayey_Fine_Sand in this case). It can be seen from Figure 4.1 that multiple instances of a material can be encountered in a single column (Silty_Clay and Silty_or_Clayey_Fine_Sand). The index of occurrence of each material is stored because the index of occurrence and the material identifies the HUF layer at each location in the MODFLOW grid. The HUF layer is retrieved or created based on the material identification and index of occurrence. The consecutive sequence of cells with a common material is then converted to an entry into the appropriate HUF thickness array, and the top elevation of those consecutive cells is entered in the HUF top elevation array. When a change in the material assignment is found (Clean_Sand in this case), the process is repeated. This process is repeated until

the bottom elevation of the MODFLOW grid is encountered. Once again, this bottom elevation might not correlate with the bottom elevation of the background grid. This entire process is then repeated for the remaining columns.

Figure 4.2 illustrates the side view of a MODFLOW grid with four curvilinear layers. The background grid used for the T-PROGS algorithm had the same plan view as the MODFLOW grid, but it contained twenty vertical layers. The different colors correlate to material zones that were mapped to different HUF layers generated by the process described above. Figure 4.2 demonstrates the ability to define soil heterogeneity independent of the MODFLOW grid. It also demonstrates the ability to use a curvilinear MODFLOW grid. The vertical scale has been exaggerated by a factor of 5.0 in this figure.

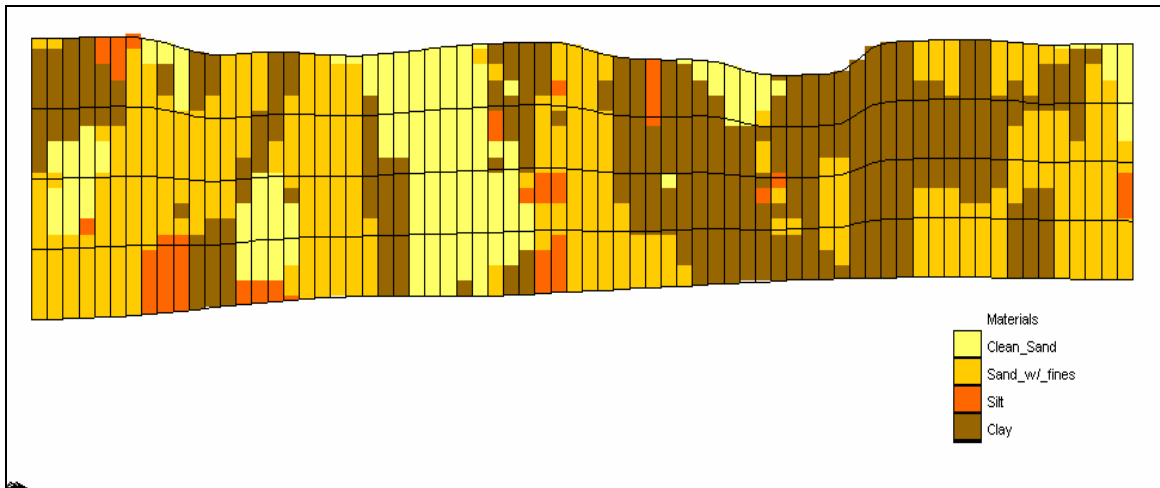


Figure 4.2 Sample HUF arrays for a four-layer grid.

After populating the HUF package in MODFLOW with the output from T-PROGS, hydraulic parameters like vertical and horizontal conductivity are assigned to each material. Head solutions are then generated with MODFLOW. The combination of T-

PROGS and the new HUF package in MODFLOW 2000 allows users to maintain detail in soil heterogeneity while reducing complexity, memory requirements, and computational time by minimizing the number of grid cells.

5 SINGLE LAYER MODFLOW MODELS

Although MODFLOW is a three-dimensional model, a majority of the MODFLOW models constructed by typical users are 2D models consisting of one model layer. There are several reasons why 2D models are so common. One reason is that many of these models are regional models where the aquifer thickness is very small compared to the lateral extent of the model. As a result, the flow directions are primarily horizontal and little improvement is gained by adding multiple layers to the model. Even with local scale models, the aquifer thickness is often small enough that one-layer models are considered adequate. 2D models are also attractive due to the simplicity of the model increased computational efficiency. One of the problems associated with using multiple layers for MODFLOW models with unconfined aquifers is that as the water table fluctuates, the upper cells may go dry. These cells will not rewet even if the water table subsequently rises, unless the rewetting option has been selected in the flow package (BCF, LPF, or HUF). The rewetting issues can often be avoided with a one-layer model.

When developing a one-layer model, the modeler must determine how to distribute the hydraulic conductivity values within the layer. One option is to assume a homogenous aquifer; this is typically a gross over-simplification since aquifers are usually highly heterogeneous. Therefore, a common approach is to delineate zones of hydraulic conductivity by examining the subsurface stratigraphic data. In many cases,

these data are in the form of borehole logs. These borehole logs often exhibit substantial heterogeneity and don't always exhibit definitive trends between adjacent boreholes. Furthermore, the boreholes are often clustered with large regions of the model lacking any borehole data. The modeler then faces a difficult task of trying to determine a rational approach to delineating two-dimensional zones of hydraulic conductivity based on complex 3D borehole data.

As part of this research, we developed a technique for developing 2D zones of hydraulic conductivity from borehole logs using transition probability geostatistics. The technique is simple, fast, and preserves proportions and trends exhibited by the borehole data.

5.1 Description of Algorithm

The 2D T-PROGS approach provides a simple, rational approach to representing borehole data in a one-layer, 3-D ground water model. The T-PROGS algorithm accurately represents the spatial variability trends in the borehole with a 3-D Markov chain system developed by GAMEAS and MCMOD. GAMEAS and MCMOD are used to develop 1-D Markov chains in the primary directions: vertical, strike, and dip. The Markov chains mathematically represent the heterogeneity in the boreholes. The process up to this point is identical for a single or multiple layer model. Then, the algorithm parses through each borehole and computes a predominant material as illustrated in Figure 5.1. In the input file to TSIM, a single record is written out for each borehole. This record includes the x and y coordinates and a z coordinate corresponding to the

average grid elevation. The z coordinate is not critical in this case since the interpolation is performed entirely in the xy plane. For the category data, a 1 is assigned to the column that correlates to its predominant material.

Data						
7						
x	=	strike				
y		dip				
z	=	elevation				
66	=	Clean_Sand				
67	=	Sand_w/_fines				
68	=	Silt				
69	=	Clay				
805.9328947746		238.49310916799	0.5	0	1	0
1417.9793180903		576.48151815065	0.5	0	1	0
1389.6281820497		482.22296477063	0.5	0	0	1
1354.9372347225		389.27654432545	0.5	0	0	1
1338.0178639188		293.67577502163	0.5	0	0	1
1406.4565280048		579.67546690643	0.5	0	1	0
1378.5744959855		482.16444655222	0.5	0	0	1
1362.9725108618		382.37749820571	0.5	0	0	1
1342.112347496		284.19605961268	0.5	0	0	1
1413.4748425483		552.67796589309	0.5	1	0	0
1404.3222176013		554.41832694301	0.5	0	1	0
1406.0826987637		528.37334266456	0.5	0	1	0
1398.0210373393		530.1381698742	0.5	0	1	0
1398.6116578279		503.83467832929	0.5	0	0	1
1390.7362323798		506.14815493626	0.5	0	0	1
1439.4370915416		573.07076185611	0.5	0	1	0
1435.4624256808		565.03580502396	0.5	0	1	0
1411.055821404		469.17748004541	0.5	0	0	1
.						
.						
.						
.						
.						
.						
.						

Figure 5.1 Data file for a one-layer model.

When TSIM runs, the predominant material for each borehole is assigned to its corresponding location in the one-layer grid, and during the quenching process, simulations are conditioned to those data points. Although the simulation can honor the

spatial variability evident in the entire borehole, it can only condition to one material at each borehole location. With our algorithm, this material always corresponds to the predominant material in the borehole. The materials assigned to the other cells honor the proportion, transition, and juxta-positioning tendencies inherent in the borehole data. Once again, the output from this algorithm can be utilized for stochastic simulations.

Figure 5.2 illustrates an example of the 2D approach. The boreholes are labeled with their corresponding predominant material. The figure illustrates the ability to accurately condition to known borehole data.

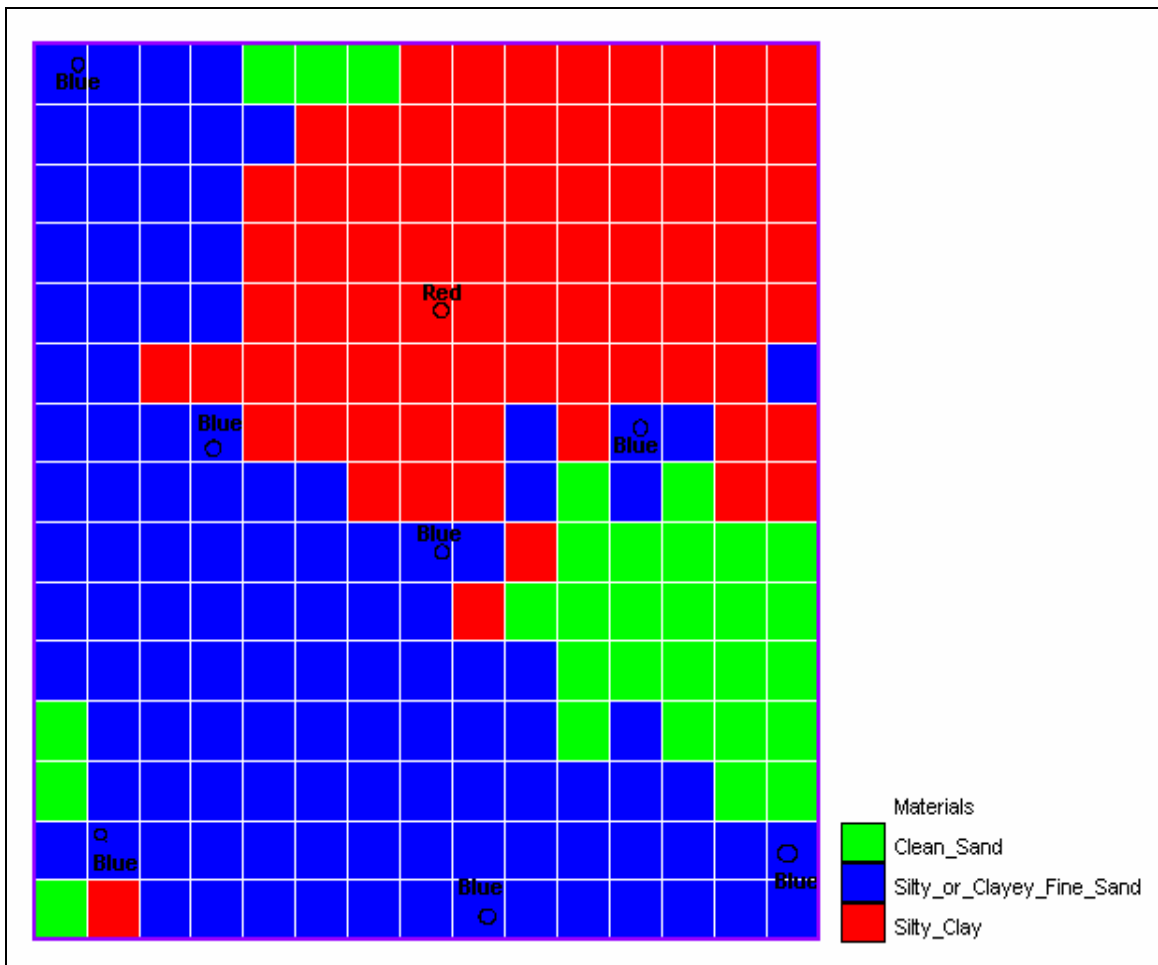


Figure 5.2 3D borehole condensed into a one-layer model using the 2D approach.

6 CASE STUDIES

This chapter describes a case study used to illustrate the new algorithms described in the previous two chapters. The case study includes indicator simulations generated with T-PROGS, MODFLOW head solutions, and particle tracking analyses. The case study demonstrates the application of the new algorithms in the context of a stochastic simulation. Four different methods will illustrate this application: 3-D orthogonal grid, HUF arrays, single layer grid, and indicator simulation in the absence of borehole data.

6.1 Site Description

The site selected to illustrate the transition probability approach of MODFLOW modeling is the Longhorn Army Ammunition Plant (LHAAP) in Eastern Texas. The plant is inactive now, but during its active status, it produced, distributed, and decommissioned various types of munitions. These processes contaminated the surface and groundwater hydrologic systems with a mixture of chemicals from solvents and oxidizers to explosives.

Since the purpose of this case study is to demonstrate the application of stochastic generation of subsurface stratigraphy and not to perform a comprehensive analysis of the groundwater transport at this site, some general assumptions were made about the site conditions and model construction. A detailed, rigorous groundwater and surface water

modeling study was performed by the U.S. Army Engineer Waterways Experiment Station (WES). The borehole data and general hydrologic properties of the site were extracted from the WES model provided by Cary Talbot ([REFERENCE](#)).

6.2 Description of Model

The entire LHAAP site is illustrated in Figure 6.1. The site covers approximately 8,420 acres.

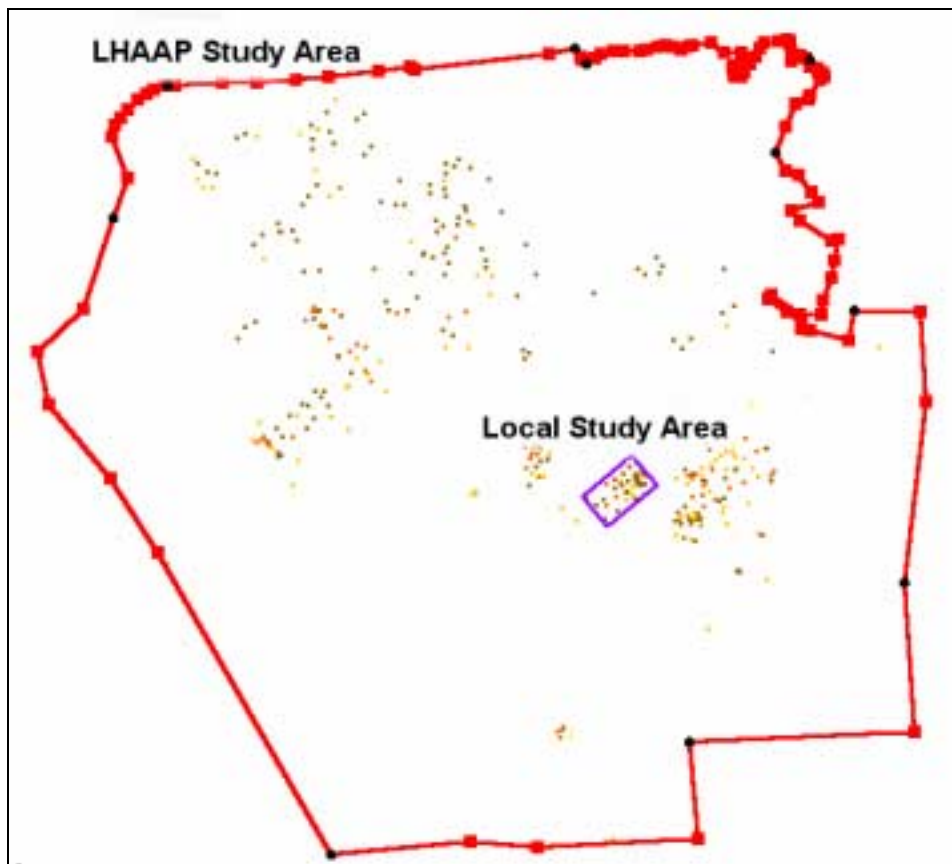


Figure 6.1 LHAAP and "Local study area". The dots represent boreholes.

The area modeled in this case study is a local scale model covering approximately 40 acres near the center of the LHAAP site indicated by the small rectangular area on Figure

6.1. The technique described in this thesis is more applicable to local scale models than to large-scale, regional models. In order for the T-PROGS software to accurately model the transition trends, the grid cells used in MODFLOW must be smaller than the average lens thickness of the materials being modeled. If the grid cells are larger than the average material lens thicknesses, the transition trends are not developed, and while the material proportions are correct, the material assigned to each cell is random. With this in mind, the subset selected for the local study area was chosen because it contained adequate borehole data to develop transition trends, and it was small enough to create cells that were smaller than the average lens lengths but not computationally unrealistic.

There are 77 boreholes inside the local study area. The borehole logs originally contained 60 different materials defined by the Unified Soil Classification System included in Table 6.1.

Table 6.1 Original materials in site boreholes.

Materials								
CH	CLCL	FI	LI	MLSP	SCCL	SMCL	SP	SPSP
CHCH	CLLI	GC	MH	OH	SCML	SMGM	SPCL	SW
CHCL	CLML	GCGC	MHMH	OL	SCSC	SMML	SPFI	SWSM
CHSC	CLOL	GM	ML	OLCL	SCSM	SMSC	SPGP	TS
CHSM	CLPT	GMGM	MLCL	PT	SCSP	SMSM	SPML	

CL	CLSC	GP	MLMH	PV	SH	SMSP	SPSC	
CLCH	CLSM	GW	MLSM	SC	SM	SMSW	SPSM	

In order to make the problem manageable with the transition probability approach, the materials were condensed down to four primary materials: clean sand, sand with fines, clay, and silt. The proportions for each material extracted from the boreholes inside the local study area are listed in Table 6.2.

Table 6.2 Materials in the local study area with corresponding proportions and hydraulic conductivity.

Material	Proportion	Horizontal Conductivity (ft/d)	Vertical Conductivity (ft/d)
Clean Sand	9%	230	115
Sand with Fines	51%	12	6
Silt	6%	2	1
Clay	34%	0.03	0.015

An extraction well was introduced into the site that pumped an average of 300 GPD. The well was screened in different layers depending on the number of layers in the model. However, the plan-view location of the well illustrated in Figure 6.2 by the “X” was constant in all the sample models. Specified head boundaries were assigned to the east-west boundaries, and no-flow boundaries marked the north-south boundaries. Heads were assigned to produce a water flow pattern from west to east. Figure 6.2 illustrates the specified head for the single layer model.

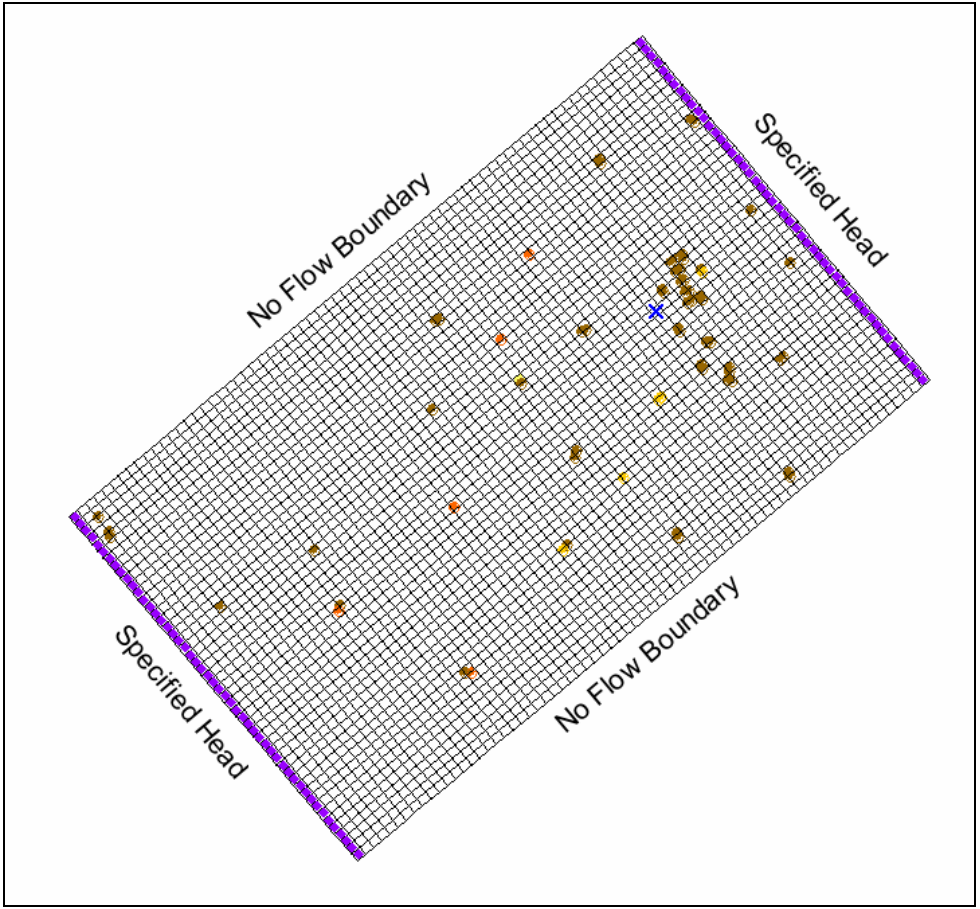


Figure 6.2 Plan view layout of the grid. Dots represent borehole locations.

The plan-view grid dimensions are the same in all the sample models. The dimensions are 1,674 feet in the strike direction and 1,031 feet in the dip. The grid is partitioned into 70 divisions in the strike direction and 50 in the dip direction. These dimensions correlate to cell lengths and widths of 24 and 20 feet respectively. This is important because the average material lens lengths must be less than the cell dimensions in the corresponding direction to develop any meaningful transition trends. The average lens lengths for the strike/dip and vertical directions are illustrated in Table 6.3.

Table 6.3 Strike/dip and vertical lens lengths for each material.

Material	Strike/Dip Lens Length (ft)	Vertical Lens Length (ft)
Clean Sand	59	5.9
Sand with Fines	132	13.2
Silt	39	3.9
Clay	73	7.3

6.2.1 Transition Trends

The same vertical, strike, and dip transition trends were used for all three of the models: 3-D orthogonal grid, single layer grid, and HUF arrays. Markov chains were developed in the vertical, strike, and dip directions with the assistance of the GAMEAS utility inside T-PROGS. Figure 6.3 illustrates the transition data in the vertical direction used to develop the Markov chains. A relatively accurate fit exists between the measured curve and the Markov chain curve.

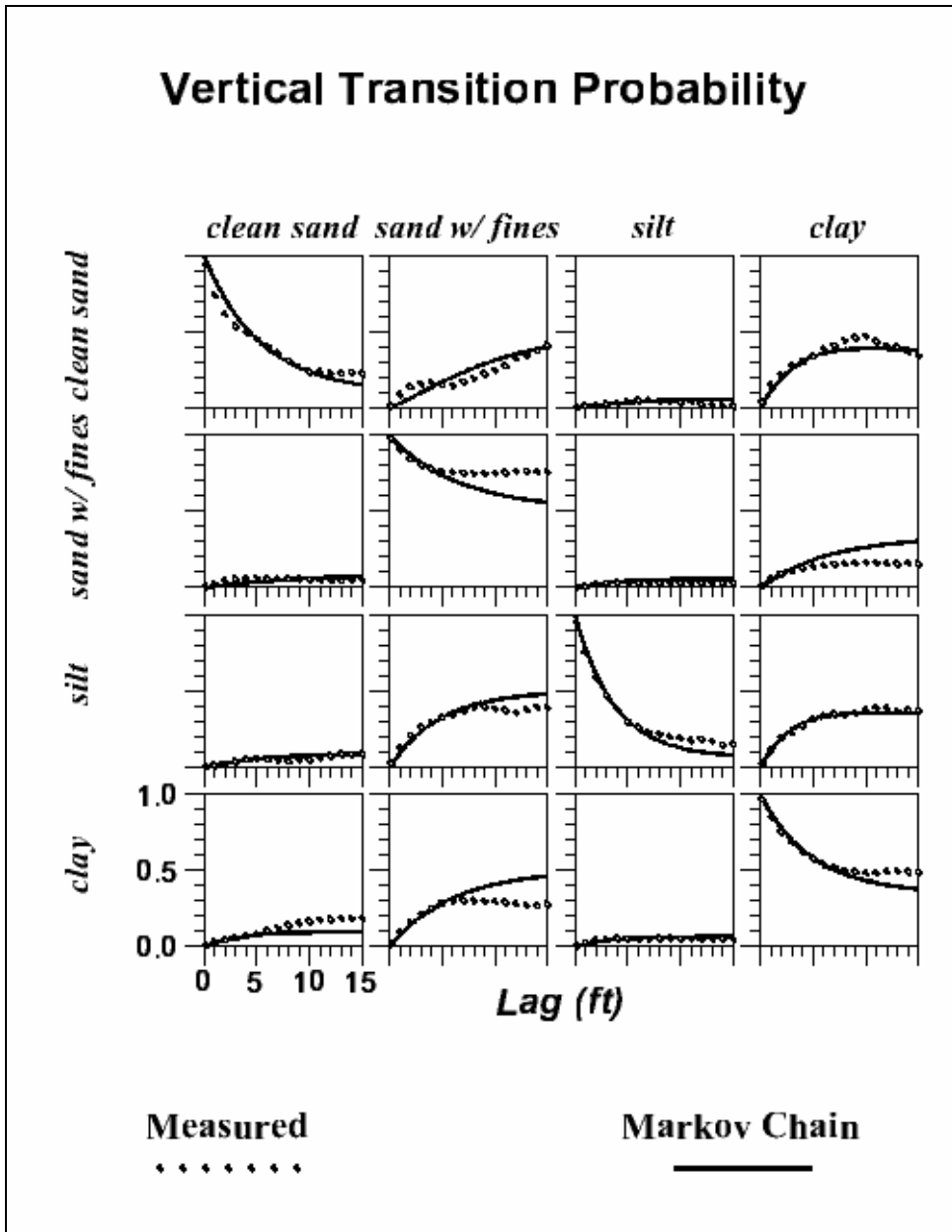


Figure 6.3 Vertical transition probability curves.

The average vertical lens lengths extracted from these curves for each material are listed in Table 6.3. The lateral (strike and dip) transition probability curves have the same shape, but the average lens lengths are ten times the lengths of their vertical counterparts.

6.3 3-D Orthogonal Grid

The first sample application involves a twenty-layer orthogonal grid of the subset area illustrated above. The purpose of this example is to illustrate the application of the transition probability approach in the general 3D case, as opposed to the HUF method developed through this research. For this case, the extraction well is screened in layer 15. T-PROGS requires uniform cell widths, lengths, and heights for a multi-layer model. The grid dimensions in the vertical direction are 86 feet partitioned into twenty layers. These dimensions correlate to cell dimensions of 2.9 feet in the vertical direction. This is important because the average material lens lengths must be less than the cell dimensions in the corresponding direction to develop any meaningful transition trends. The average vertical lens lengths for each material are illustrated in Table 6.3.

A 3D Markov chain was developed with MCMOD. Seventy-five realizations were generated with TSIM. Figure 6.4 and Figure 6.5 display an oblique view of two of the realizations. The vertical scale has been exaggerated by a factor 5.0 in each of the figures.

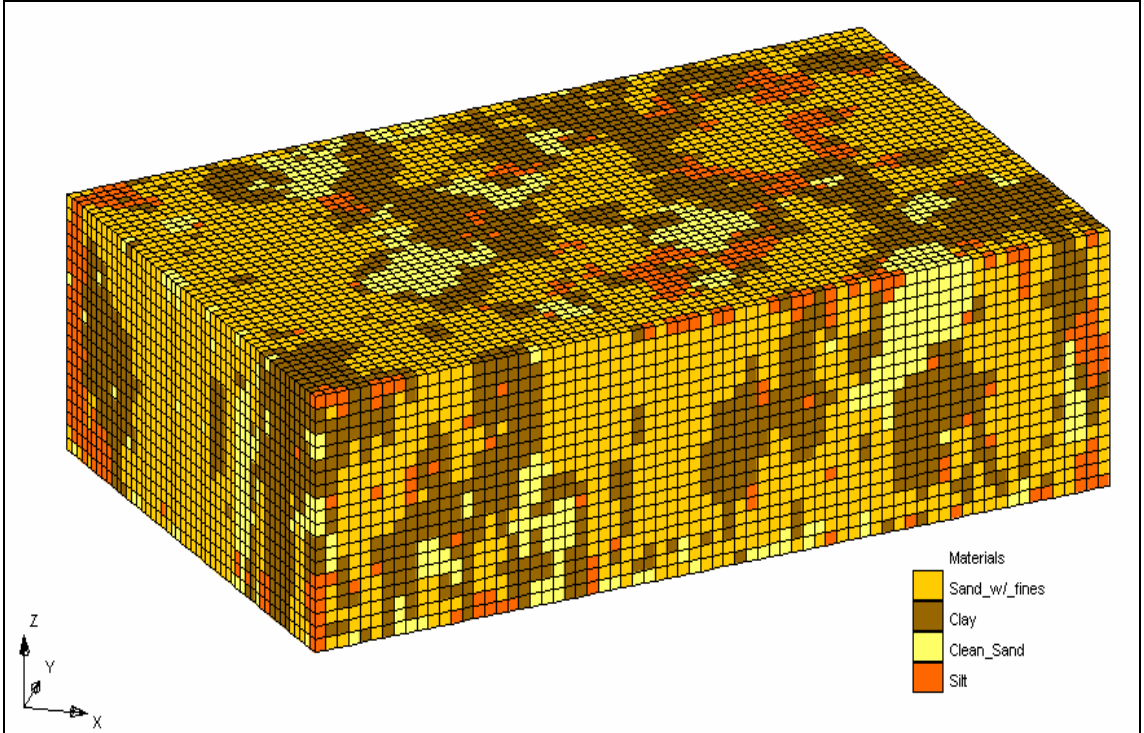


Figure 6.4 Oblique view of the realization #1 with cells filled with material colors.

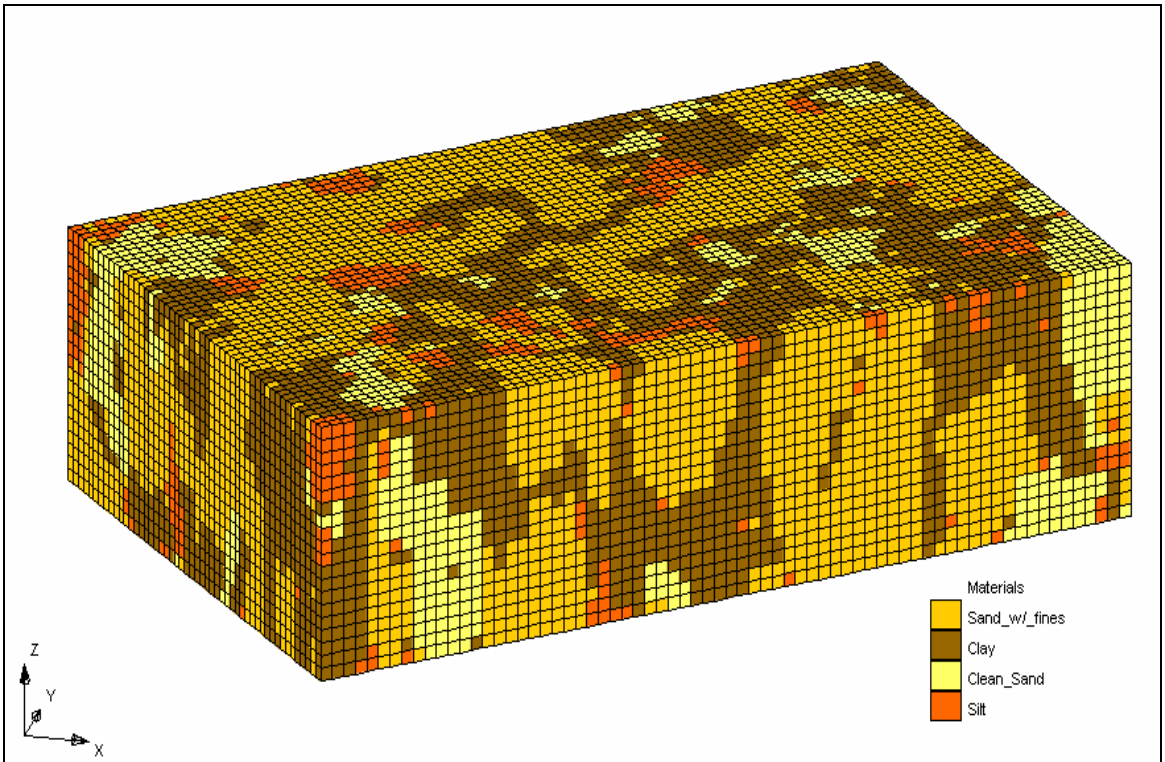


Figure 6.5 Oblique view of the realization #2 with cells filled with material colors.

MODFLOW head solutions for the top layer of each grid are illustrated in Figure 6.6 and Figure 6.7 (black contours). In addition, a particle tracking analysis was performed on each solution with MODPATH (blue lines).

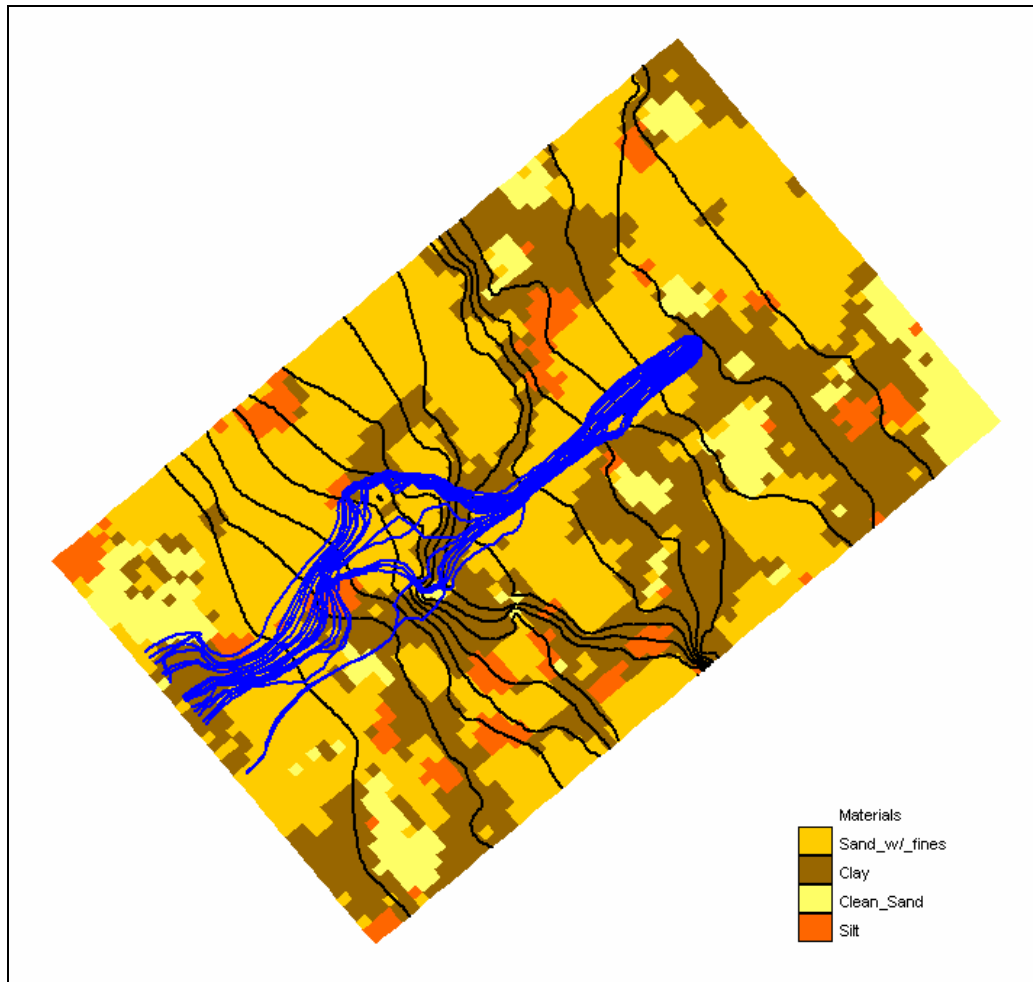


Figure 6.6 Head (black) and particle flow line (blue) solution for realization #1.

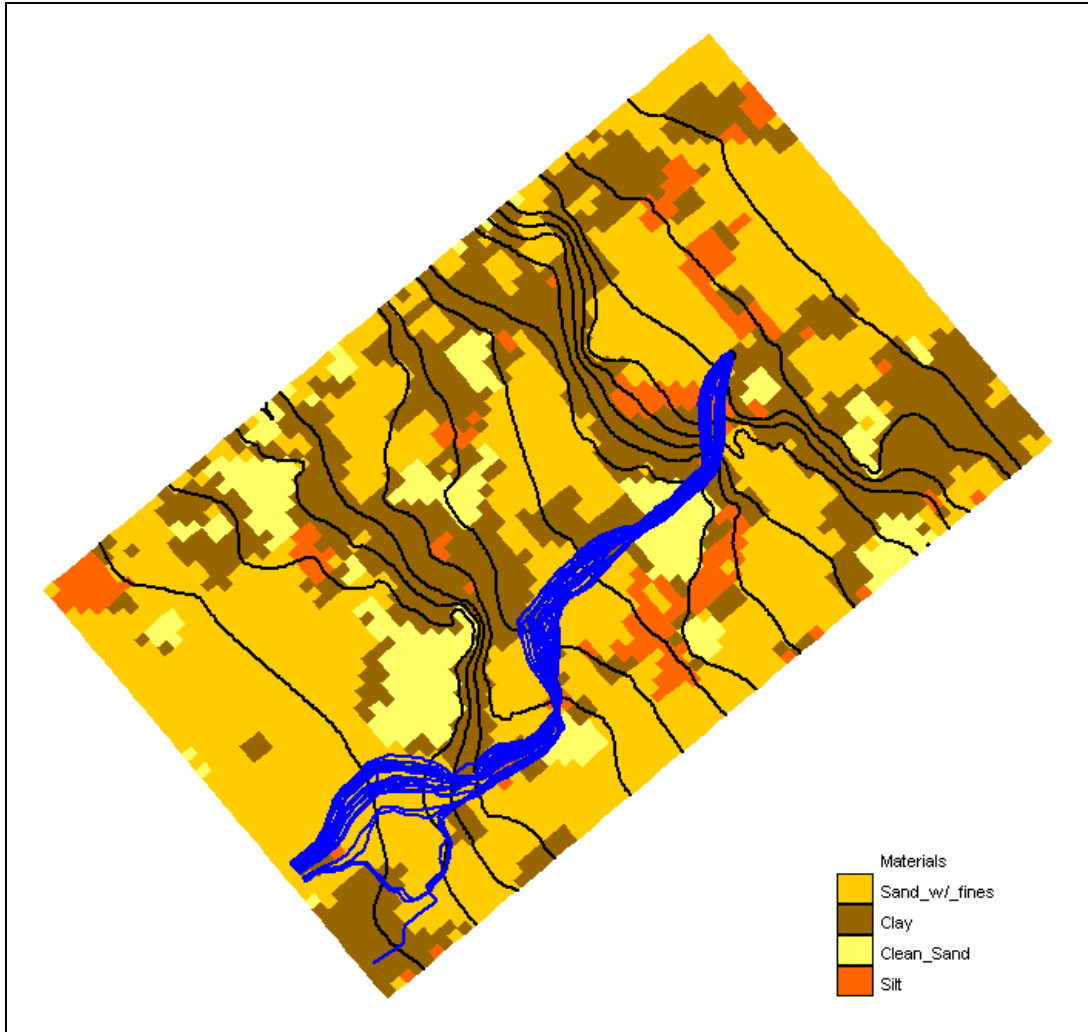


Figure 6.7 Head (black) and particle flow line (blue) solution for realization #2.

Note the differences in the pathlines for the two samples shown in Figure 6.6 and Figure 6.7. One of the benefits of this approach is that the model realistically portrays complex capture zones resulting from aquifer heterogeneity. This is further illustrated by the model cross-sections shown in Figure 6.8 and Figure 6.9.

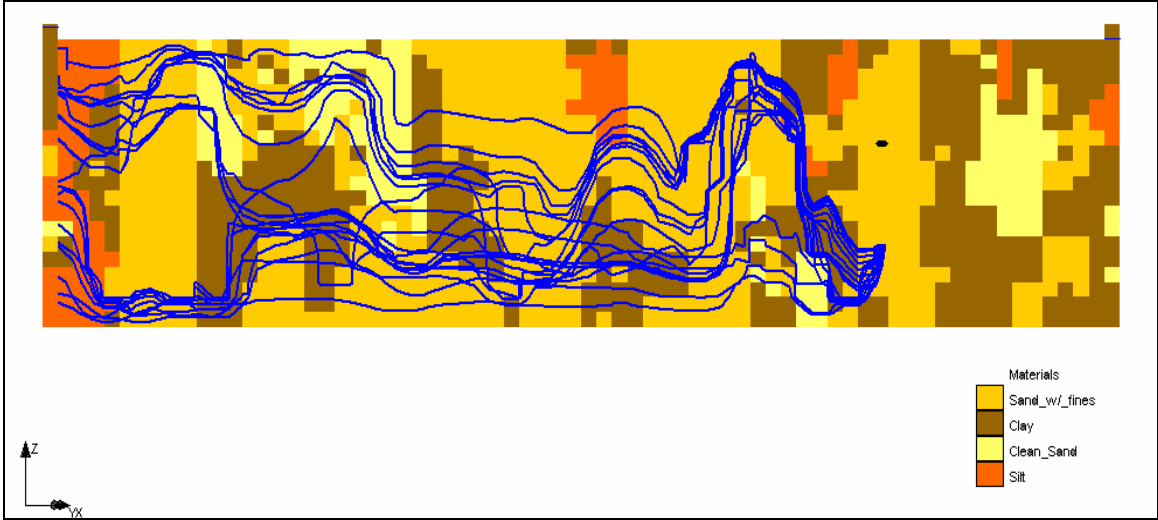


Figure 6.8 Side view of flow lines and water table.



Figure 6.9 Side view of flow lines and water table.

As part of the particle tracking analysis, a 3D probabilistic capture zone was developed from the MODFLOW solutions in which the cell at the well was not dry. Some of the realizations developed by T-PROGS assigned clay to the cells including and adjacent to the well. For some of the realizations, the well could not extract the quantity of water specified by the pumping rate with the given conductivity of clay. Therefore, the cell

went dry. In fifty-one of the seventy-five solutions, the cell including the well did not go dry, and those solutions were used in the probabilistic capture zone analysis. A 3D probabilistic capture zone is generated by placing a particle at each cell center of the grid for each simulation and tracking the cell backward in time to determine if it terminates at the cell. The particle tracking analysis was conducted with MODPATH (Pollock, 1994). The probability that the particle at the cell reaches the well is computed by dividing the number of simulations in which the particle from that cell reached the well by the total number of simulations. The probability at each cell is then contoured with iso-surfaces. A set of iso-surfaces developed for different probabilities of capture is illustrated in Figure 6.10-Figure 6.13.

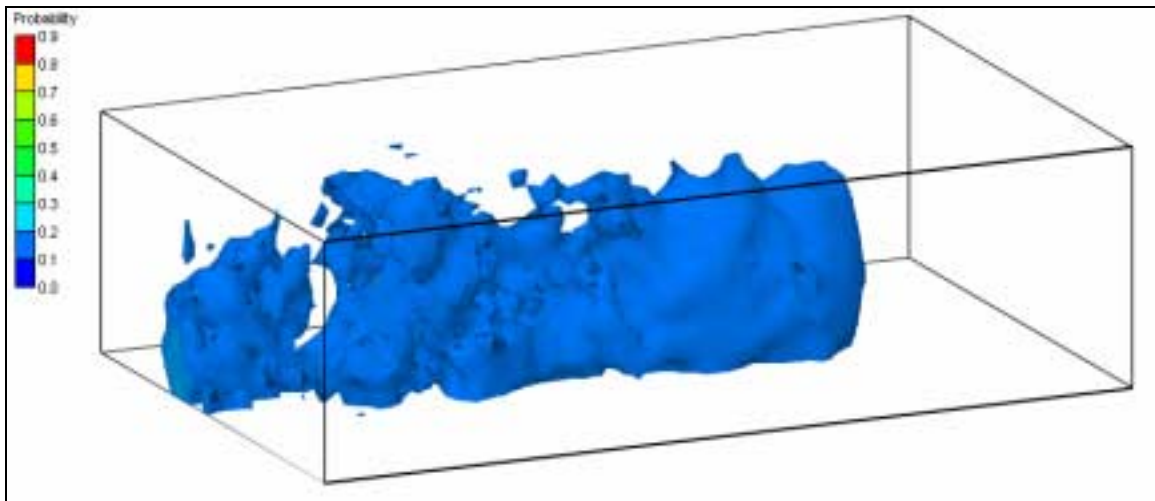


Figure 6.10 3D capture zone for a probability shell equal to 10%.

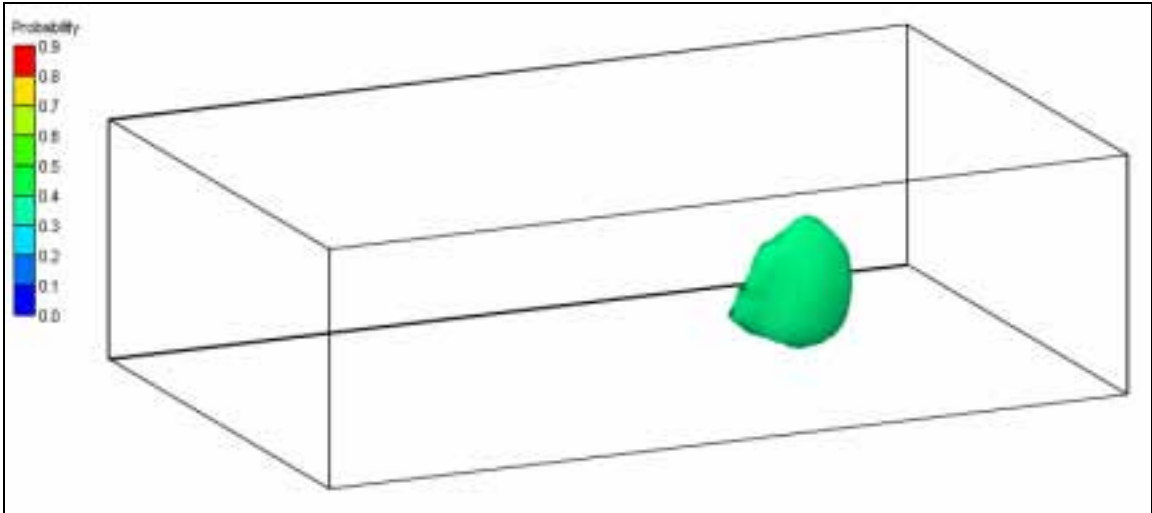


Figure 6.11 3D capture zone for a probability shell equal to 30%.

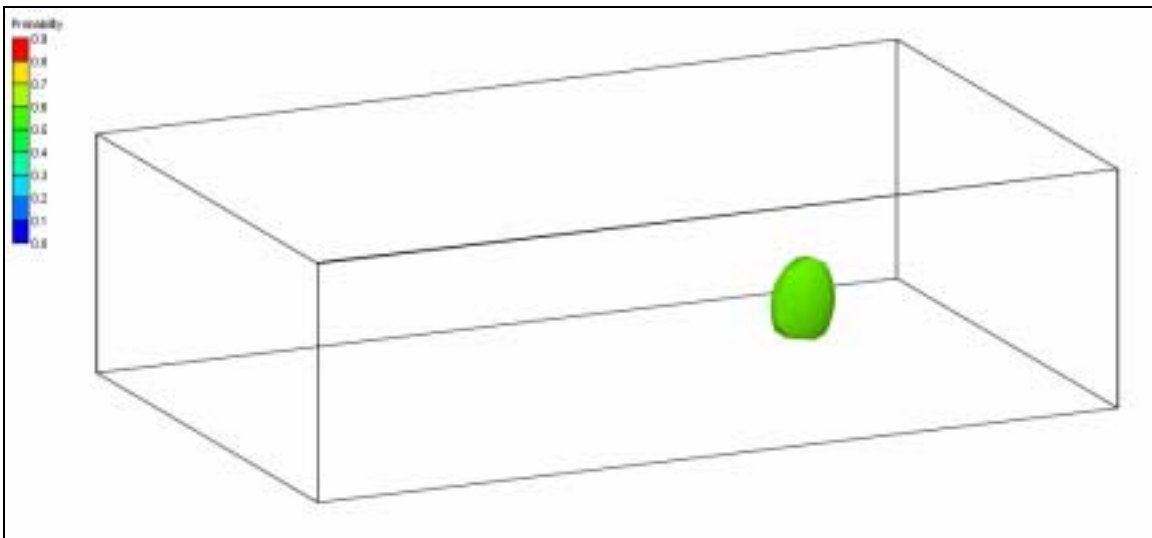


Figure 6.12 3D capture zone for a probability shell equal to 50%.

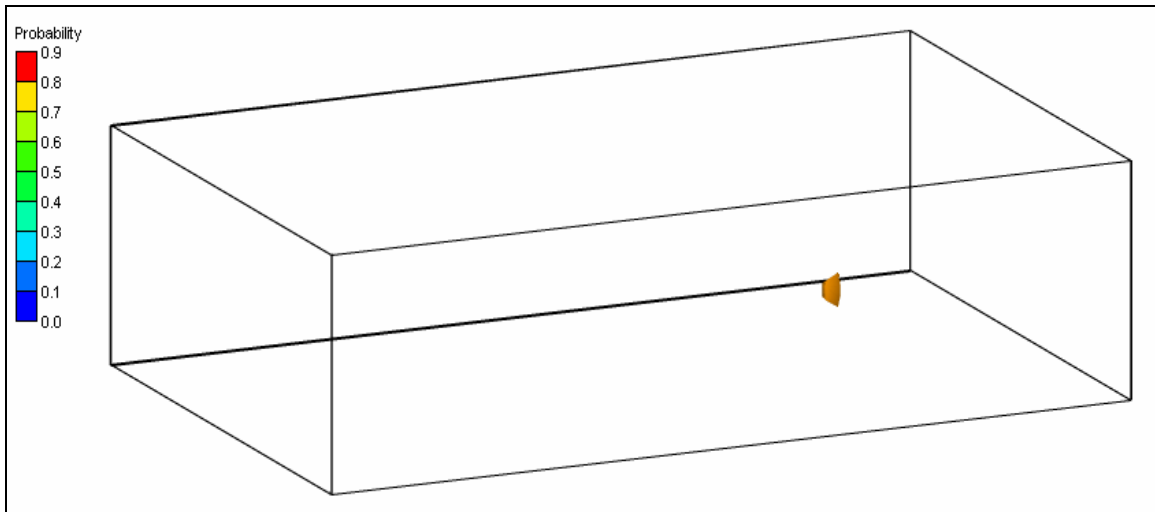


Figure 6.13 3D capture zone for a probability shell equal to 70%.

A cutaway of the iso-surfaces was also generated to illustrate the interior configuration of the capture zone (Figure 6.14 and Figure 6.15).

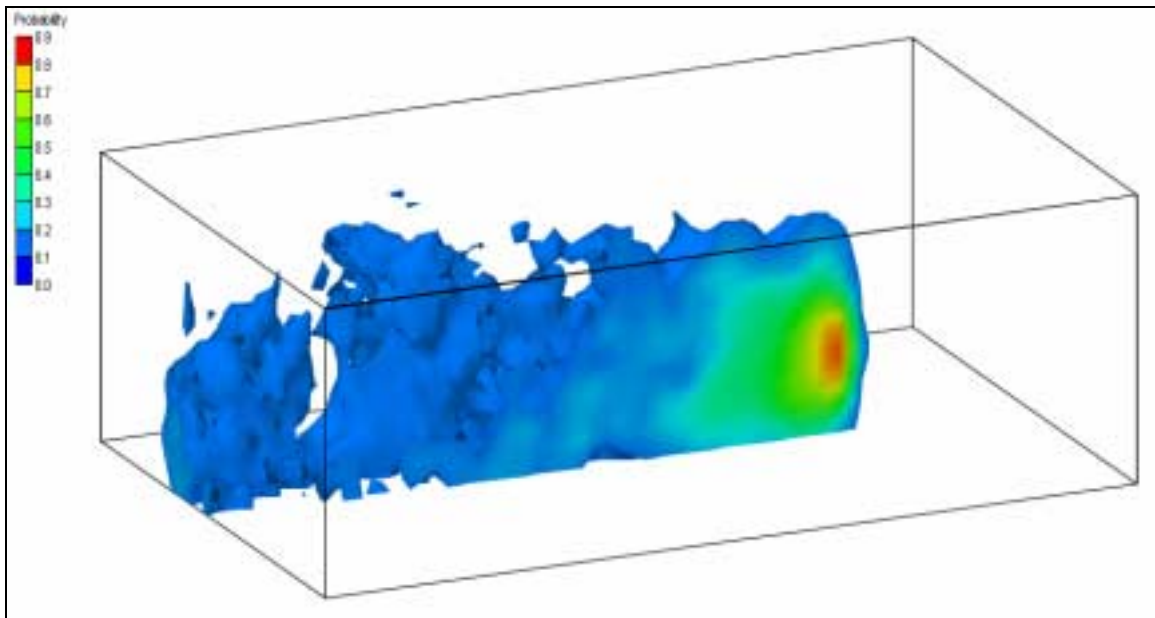


Figure 6.14 3D cutaway of the interior of the capture zone with iso-surfaces capped.

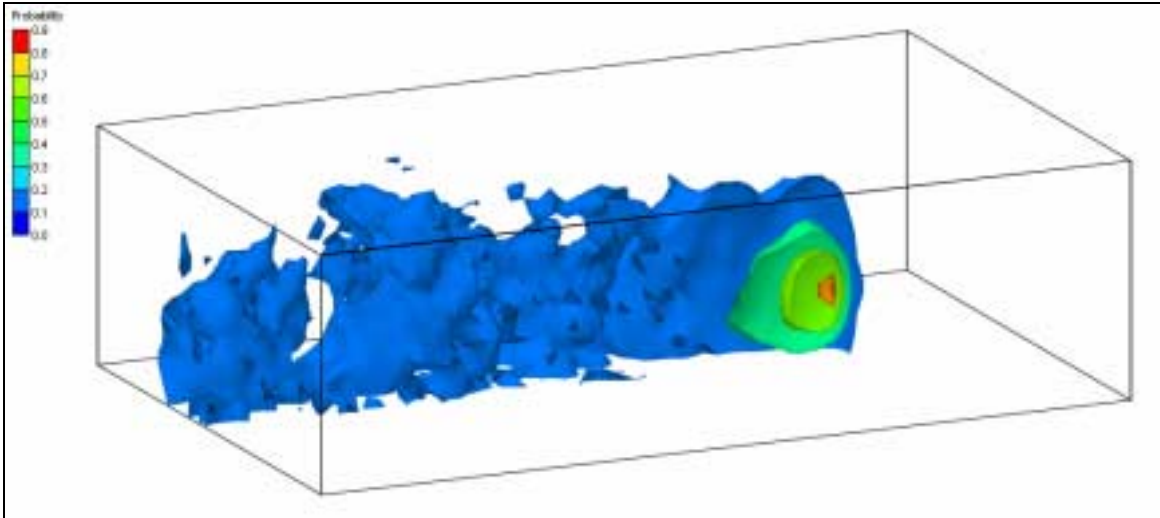


Figure 6.15 3D cutaway of the interior of the capture zone with iso-surfaces not capped.

6.4 Single Layer Model

The grid for the single layer model has similar plan-view dimensions as the 3D case, but the grid has only one layer in the vertical direction. The 2D Approach outlined in Chapter 6 was used when conditioning the realizations to the borehole data. Ninety realizations were developed for the local study area. Two of the realizations are illustrated in Figure 6.16 and Figure 6.17. The cells are filled with their corresponding material identification color. The heads computed by MODFLOW are also contoured (black), and the flow lines computed by MODPATH are displayed in blue.

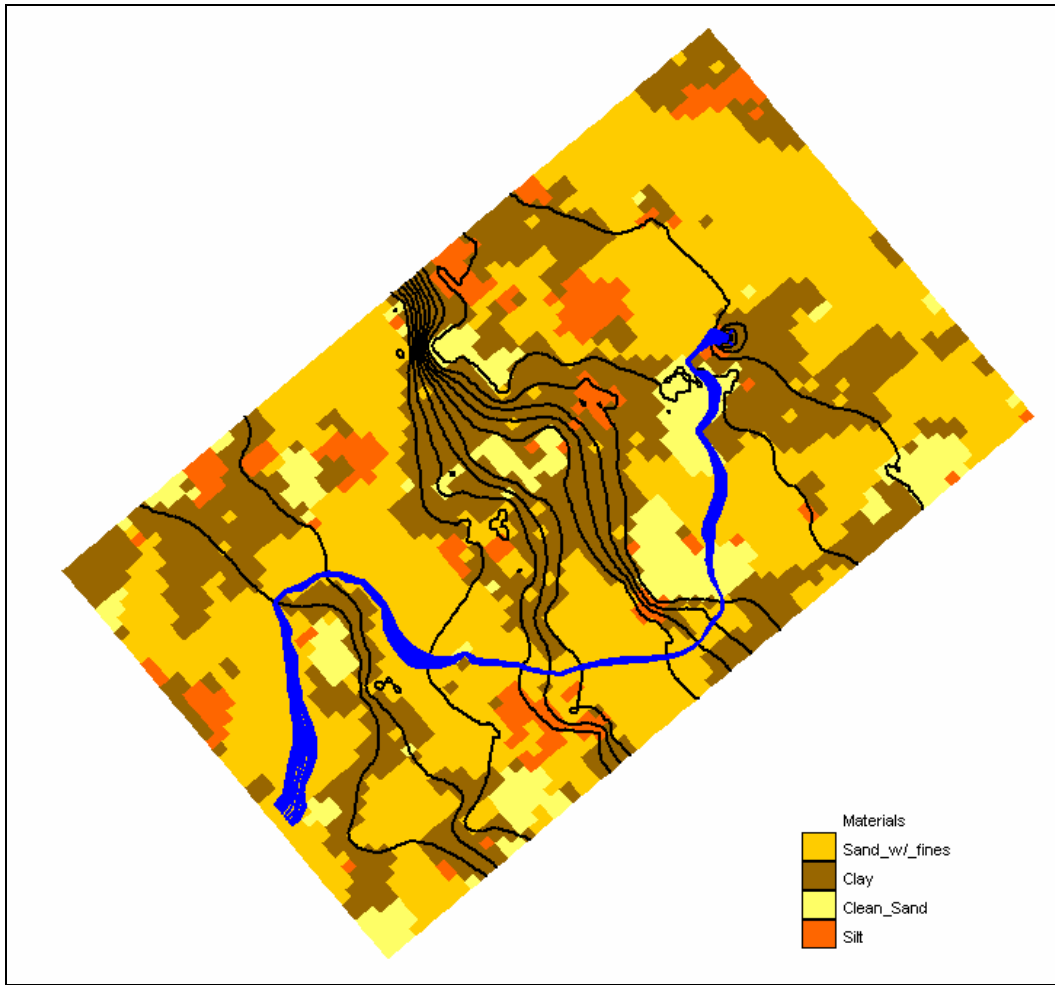


Figure 6.16 Realization #1 for 2D approach with MODFLOW and MODPATH solutions.

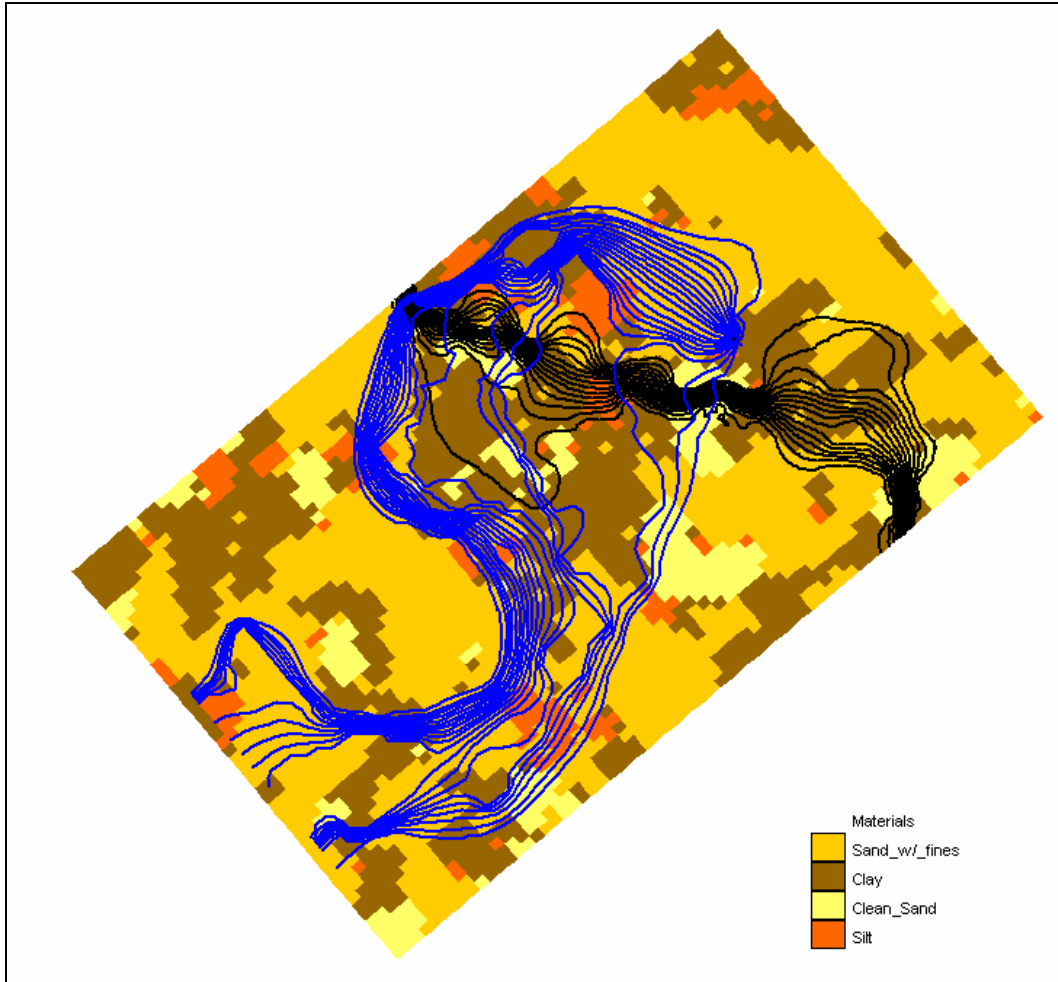


Figure 6.17 Realization #2 for 2D approach with MODFLOW and MODPATH solutions.

A visual analysis of the material distribution in each of these realizations confirms the material proportions listed in Table 6.2 and the impact of variability in soil stratigraphy on head and transport behavior. The pathlines clearly show channeling effects and there is a large difference in the well capture zone, even though both cases are conditioned to the same borehole data.

In addition to the head solutions, a 2D probabilistic capture zone illustrated in Figure 6.18 was computed for those solutions in which the cell with the well assigned to

it did not go dry. Eighty-three of the ninety solutions were used in the capture zone analysis. The 2D probabilistic capture zone differs from the 3D capture zone described above. For the 2D case, particles are distributed at the water table surface for every cell instead of the cell center. A backward particle tracking analysis is then performed and the probability of capture is computed the same way as in the 3D case. The contours correlate to the probability of a particle being captured by the well marked with an “X”.

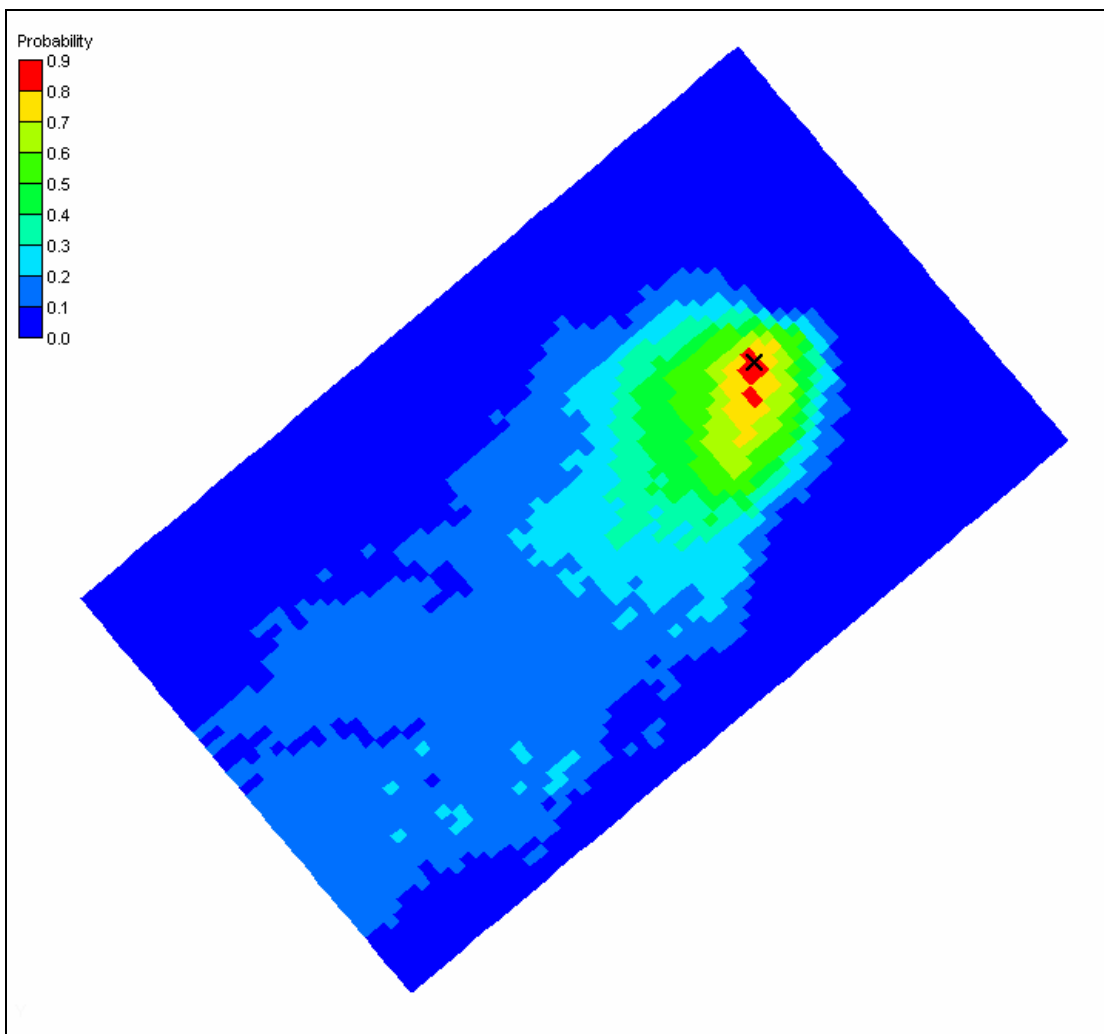


Figure 6.18 2D probabilistic capture zone for the well.

6.5 HUF Arrays

The grid for the HUF case has similar plan-view dimensions as the 2D and 3D case. However, the grid used to model the HUF array application has four layers. Furthermore, the grid has uniform row and column dimensions, but the layer dimensions vary throughout the grid and are curvilinear as evident in Figure 6.19. The extraction well is screened in layer three. The approach outlined in Chapter 4 was used to develop realizations of HUF arrays with the T-PROGS software. Ninety realizations were developed for the local study area. Two of those realizations are illustrated in Figure 6.19 and Figure 6.20. The effect of the soil variability is demonstrated by the flow lines and head contours. The cells are filled with the material identification color corresponding to the hydrogeologic unit defined by the HUF arrays.

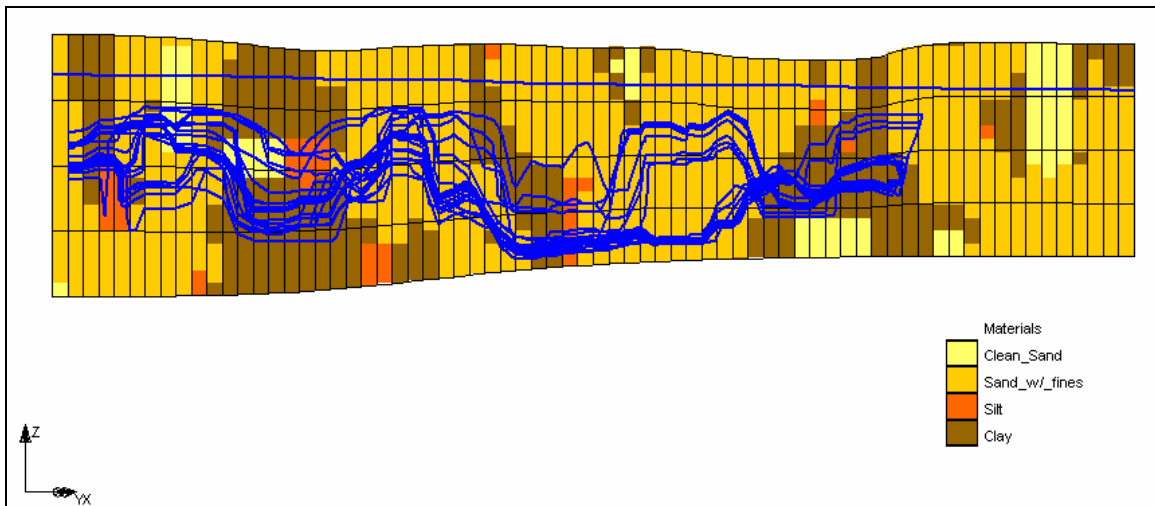


Figure 6.19 Realization #1 for HUF approach with MODFLOW solutions (side view).

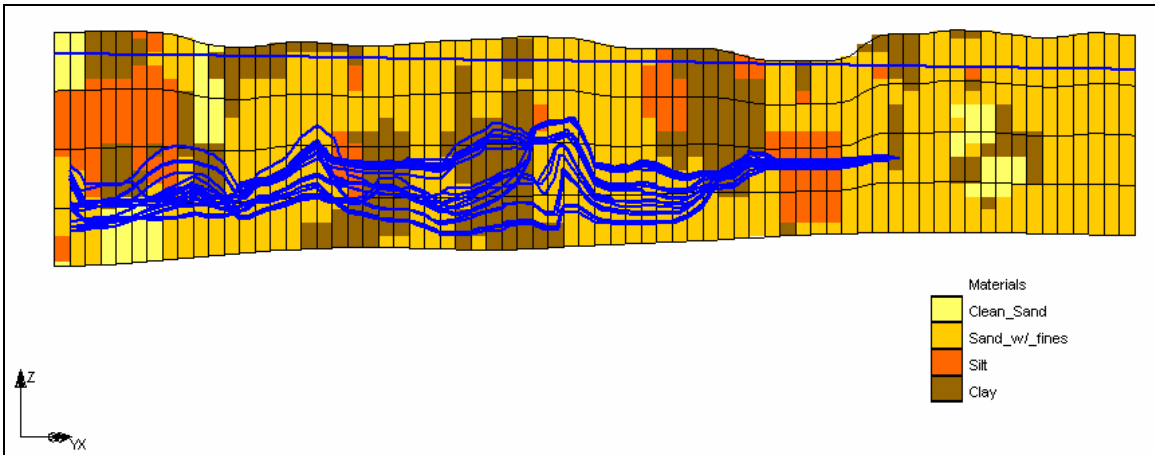


Figure 6.20 Realization #2 for HUF approach with MODFLOW solutions (side view).

The plan view figures (Figure 6.21 and Figure 6.22), illustrate the heads computed by MODFLOW contoured in black and the flow lines computed by MODPATH in blue.

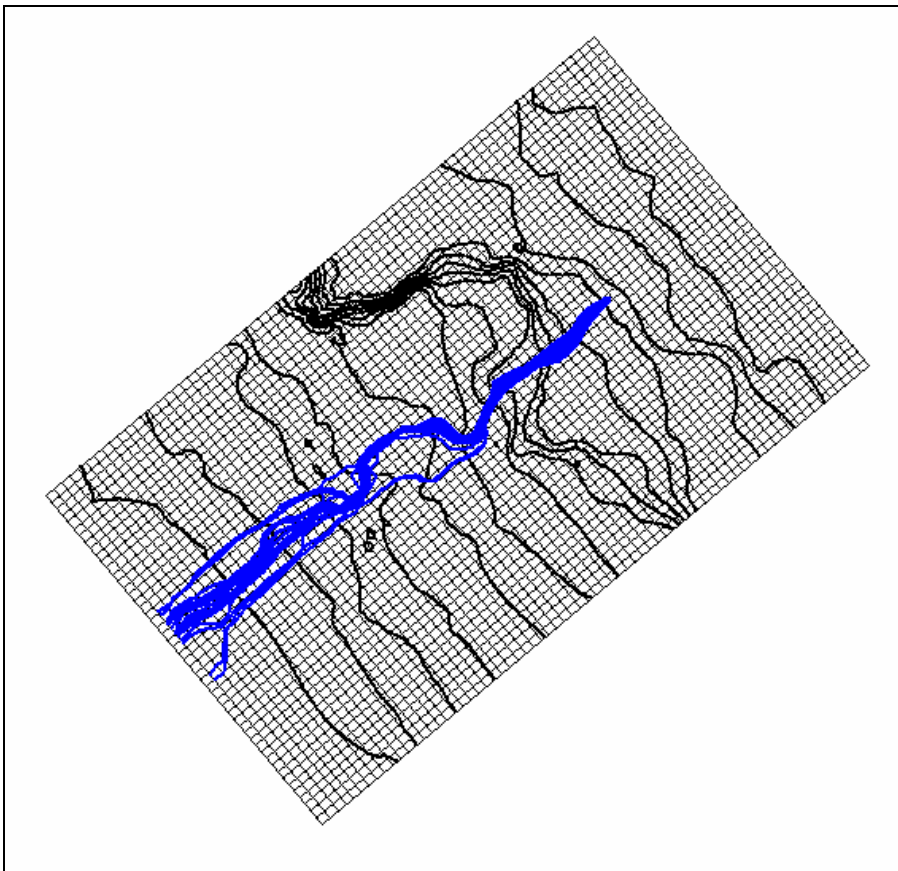


Figure 6.21 Realization #1 for HUF approach with MODFLOW and MODPATH solutions (plan view).

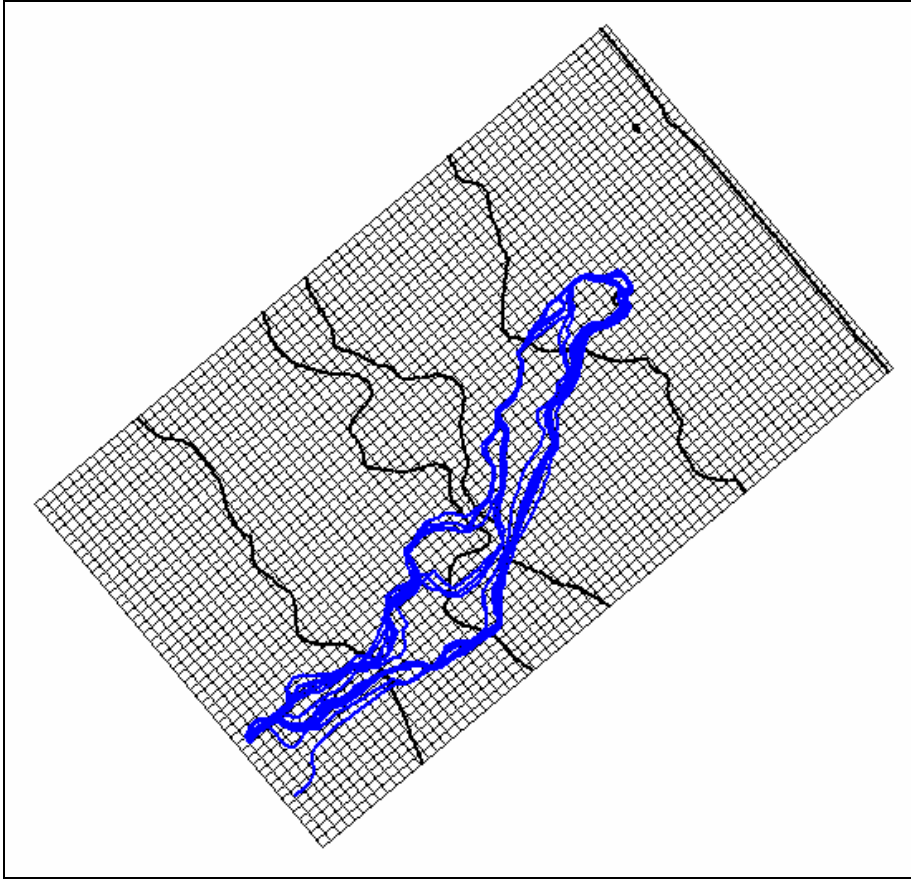


Figure 6.22 Realization #2 for HUF approach with MODFLOW and MODPATH solutions (plan view).

In addition to the head solutions, a probabilistic capture zone was computed for all the simulations. A set of iso-surfaces corresponding to differing values of probability of capture was developed for the zone is illustrated in Figure 6.23-Figure 6.26.

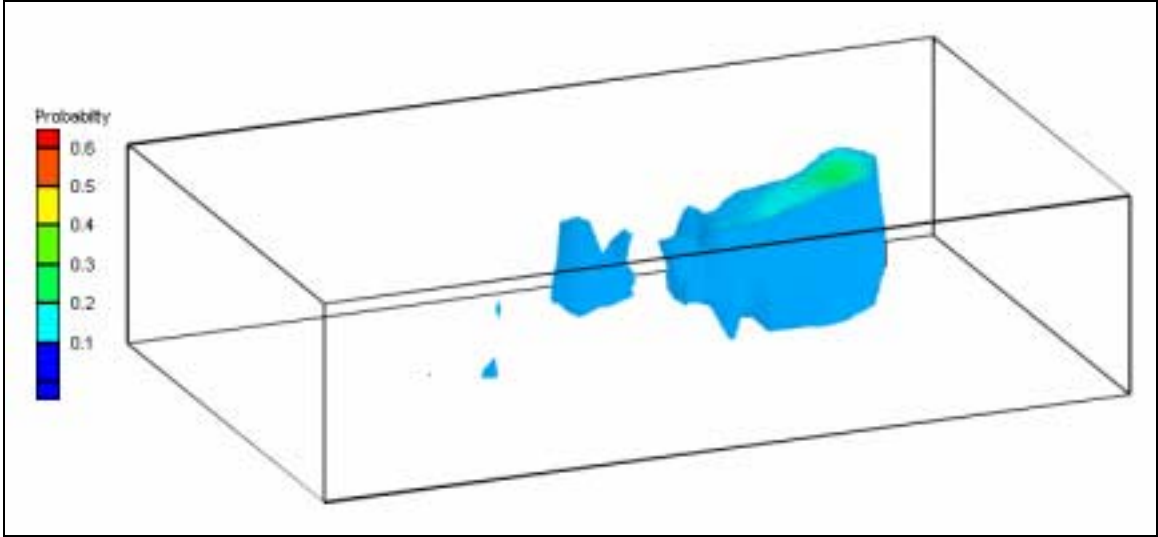


Figure 6.23 HUF capture zone for a probability shell equal to 10%.

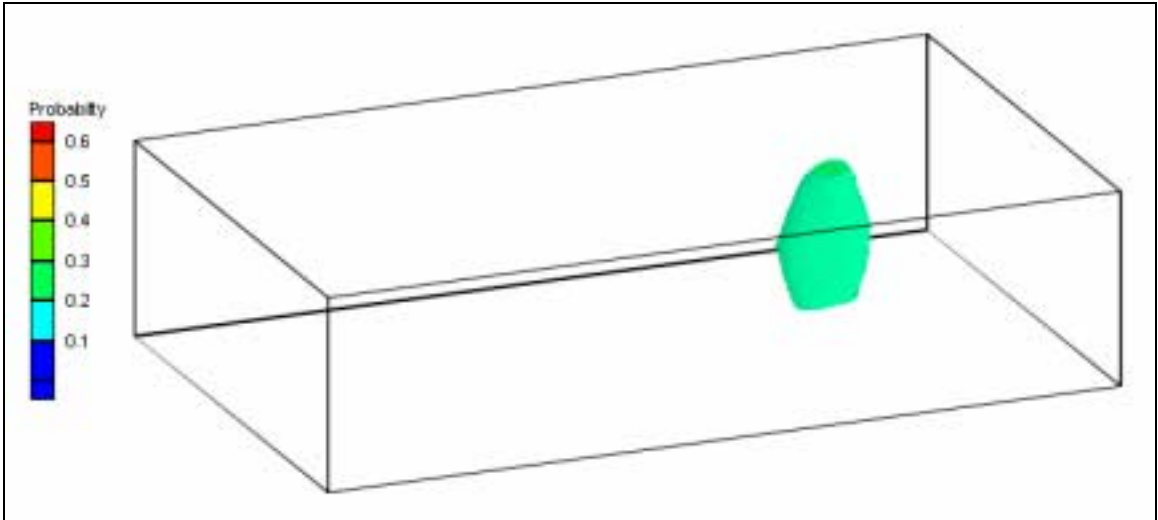


Figure 6.24 HUF capture zone for a probability shell equal to 20%.

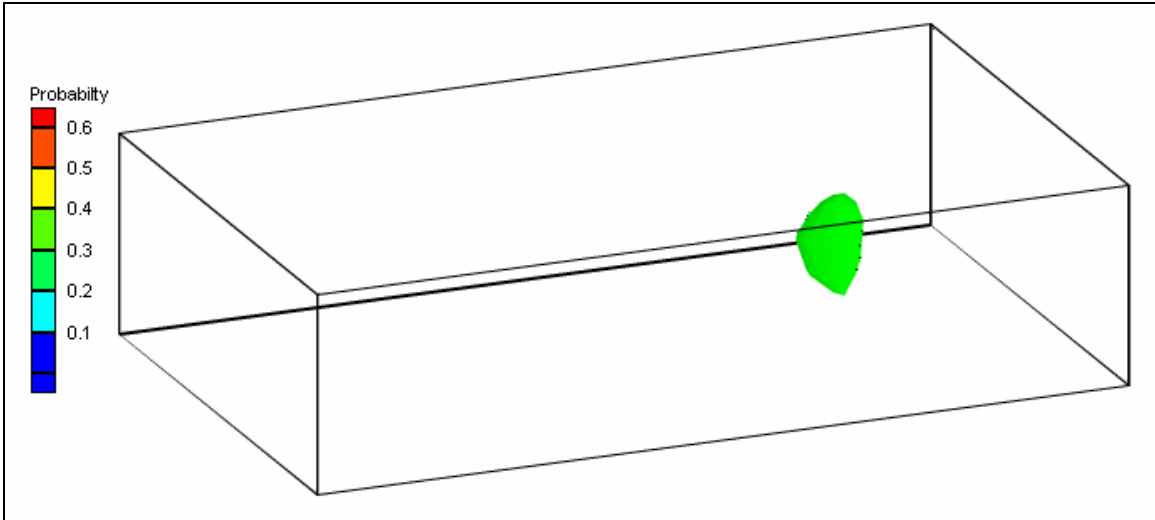


Figure 6.25 HUF capture zone for a probability shell equal to 30%.

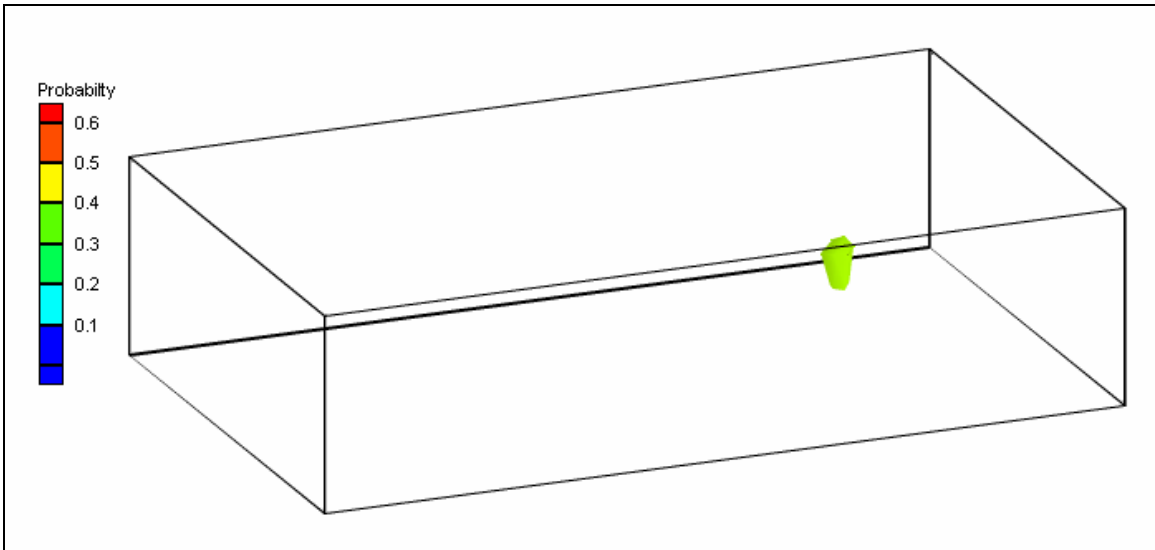


Figure 6.26 HUF capture zone for a probability shell equal to 40%.

A cutaway of the iso-surfaces was also generated to illustrate the interior configuration of the capture zone (Figure 6.27).

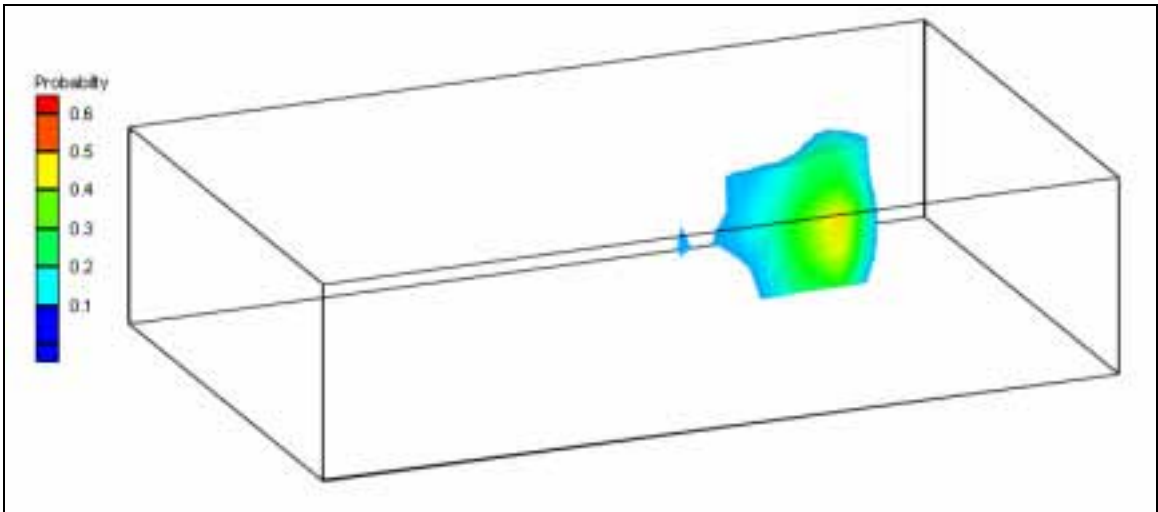


Figure 6.27 HUF cutaway of the interior of the capture zone with iso-surfaces capped.

7 CONCLUSION

Soil stratigraphy in ground water modeling is typically defined by large, homogeneous polygonal zones. However, actual soil deposition is dominated by small lenses, and soil parameters such as hydraulic conductivity can vary by two orders of magnitude or more inside a ten foot radius. Various stochastic simulators exist to generate N equally probable realizations of variability in soil stratigraphy. Transition probability geostatistics for indicator simulation presents an attractive method of generating soil stratigraphy from typical site characterization data such as boreholes. Some unique qualities of the transition probability approach make it the candidate of choice. First, since it is a type of indicator simulation, this algorithm operates on discrete parameters, which correlates well with geologic facies. It is conducive to stochastic inverse modeling like PEST and other models. Stochastic realizations can be conditioned to site data introduced by boreholes. The transition probability approach also has some unique advantages over the traditional Kriging methods like cross-variograms. The transition probability approach provides a conceptual framework for incorporating geologic interpretations such as material proportions, average lens lengths, background material, and depositional patterns into the model. It also enhances the development of asymmetric depositional patterns like fining-upwards or radial morphology. In addition, the transition probability approach enhances stratigraphic development in lateral directions where the absence of adequate borehole data makes curve fitting unrealistic. A

transition probability software package, T-PROGS, developed by Steven F. Carle was the tool used to incorporate stochastic stratigraphy simulation into MODFLOW with the Department of Defense Groundwater Modeling System (GMS) interface.

The indicator arrays generated by the transition probability simulation can be converted into material identification arrays for use in the Layer Property Flow (LPF) package in MODFLOW 2000. The material set approach presents a rational way of generating soil stratigraphy laterally between borehole data. Furthermore, material sets can be used to assign three dimensional borehole data to a one-layer model. The indicator arrays generated by the transition probability simulation can also be converted into layer elevation and thickness arrays for use with the new Hydrogeologic Unit Flow (HUF) package in MODFLOW 2000. This makes it possible to preserve complex heterogeneity while using reasonably sized grids. This approach can also be used to generate multiple realizations of aquifer zonation for MODFLOW simulations involving one-layer models.

The stochastic tools described above enable MODFLOW users to generate soil stratigraphy conceptually with geologic knowledge in the presence of sparse or even no borehole data. These tools also enable more realistic modeling of local transport dominated by low and high conductivity lenses. Furthermore, the compatibility of the transition probability approach with stochastic inverse modeling enables the use of an automated parameter estimation engine such as PEST to optimize the parameter values associated with the material zones generated for each model instance. The end result of this process is that each of the N model runs is calibrated to field observed heads and flows. This provides a greater confidence in the probabilistic model predictions.

Additional topics of research that would extend and improve the application of the transition probability approach in MODFLOW exist. One topic of research being pursued currently by Jonathan Green and Dr. Norman Jones in the Environmental Modeling Research Lab (EMRL) at Brigham Young University is probabilistic capture zones generated from multiple MODFLOW head solutions based on T-PROGS output. Although a description of the application of the transition probability approach has been given in this these, it has not been demonstrated that this approach accurately models aquifer behavior. Therefore, further research is recommended to ascertain the validity of stochastic simulations. Furthermore, additional research could demonstrate if a higher degree of detail correlates to a more accurate model of aquifer behavior. Another issue that arises with stochastic simulations of soil stratigraphy involves modeling wells. If a well is included in a model and a stochastic simulator is used to generate realizations of soil stratigraphy, it is likely that a low-conductivity material will populate the cells surrounding the well. This situation creates a potential for the cells to go dry around the well. In field conditions, wells are developed to flush out fines surrounding the well, which increases the hydraulic conductivity around the well. Further research could address this dilemma and potentially develop a method of simulating a developed well in MODFLOW.

APPENDIX A REFERENCES

- Anderman, E.R. & Hill, M.C., MODFLOW-2000, THE U.S. GEOLOGICAL SURVEY MODULAR GROUND-WATER MODEL-DOCUMENTATION OF THE HYDROGEOLOGIC-UNIT FLOW (HUF) PACKAGE, USGS, Denver, CO, 2000.
- Carle, Steven F., A Transition Probability-Based Approach to Geostatistical Characterization of Hydrostratigraphic Architecture, reprint of Ph. D. dissertation, Hydrologic Sciences Graduate Group University of California, Davis, 1996.
- Carle, Steven F. et al., Conditional Simulation of Hydrofacies Architecture: A Transition Probability/Markov Approach, Hydrologic Sciences Graduate Group University of California, Davis, 1998.
- Carle, Steven F. & Graham, Fogg E., Transition Probability-Based Indicator Geostatistics, Hydrologic Sciences Graduate Group University of California, Davis, 1995.
- Carle, Steven F. & Graham, Fogg E., Modeling Spatial Variability with One and Multidimensional Continuous-Lag Markov Chains, Hydrologic Sciences Graduate Group University of California, Davis, 1997.
- Carle, Steven F., T-PROGS: Transition Probability Geostatistical Software Version 2.1, Hydrologic Sciences Graduate Group University of California, Davis, 1999.
- Davis, John C., Statistics and Data Analysis in Geology, 2nd Edition, New York: John Wiley & Sons, Inc., 1986.
- Deutsch, Clayton V. and Journel, Andre G., GSLIB Geostatistical Software Library and User's Guide, 2nd Edition, New York: Oxford University Press, 1998.
- Doherty, 2000, PEST: Model Independent Parameter Estimation, Watermark Numerical Computing.
- Harbaugh, A., Banta, E.R., Hill, M.C., & McDonald, M.G., MODFLOW-2000, THE U.S. GEOLOGICAL SURVEY MODULAR GROUND-WATER MODEL-USER GUIDE TO MODULARIZATION CONCEPTS AND THE GROUND-WATER FLOW PROCESS, USGS, Reston, VA, 2000.

Parzen, E., Stochastic Processes, Department of Statistics, Stanford University, San Francisco, Calif., Holden-Day, Inc., 1962.

Pollock, D. W., 1994, User's Guide for MODPATH/MODPATH-PLOT, Version 3: A particle tracking post-processing package for MODFLOW, the U.S. Geological Survey finite difference ground-water model, U.S. Geological Survey, Open-File Report 94-464, Reston, Virginia, Sept., 1994.

Tompson, Andrew F. B., Ababou, Rachid, and Gelhar, Lynn W., Implementation of the Three-Dimensional Turning Bands Random Field Generator, The American Geophysical Union, 1989.

APPENDIX B GMS/T-PROGS INTERFACE

As part of this research, a graphical user interface to the T-PROGS software was developed in the Department of Defense Groundwater Modeling System (GMS). The interface is contained in the T-PROGS Menu in the Borehole Module. The GMS interface takes a five-step approach to generating spatial realizations of geologic units. Algorithms from the T-PROGS software are used intermittently throughout those five steps. If the user has not initialized a T-PROGS simulation, the five steps are accomplished in a wizard with each step automatically following the previous step. Otherwise, each step can be completed individually. These five steps are illustrated in Figure BB and include:

- 1) General options
- 2) Vertical Markov chain definition
- 3) Strike Markov chain definition
- 4) Dip Markov chain definition
- 5) Simulation generation

These options become initialized when the *New Simulation* command is selected. They are dimmed when the current simulation is deleted.

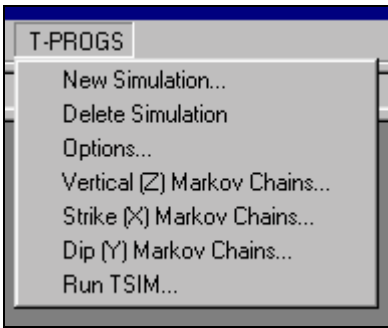


Figure BB T-PROGS menu commands.

B.1 Building a Grid

A prerequisite for building T-PROGS data is a three-dimensional grid. The indicator arrays generated by the T-PROGS algorithm are interpolated to a grid, and therefore, a grid must exist. The T-PROGS algorithm is compatible with two types of grid. If the option to generate material sets for a multi-layer grid is selected, the grid must be orthogonal with uniform row, column, and layer dimensions. The row dimensions can be different from the column dimensions, but the row dimensions must be the same throughout the grid. The same conditions exist for the columns and layers. If the option to generate HUF arrays or material sets for a one-layer grid is selected, the grid must have uniform row and column widths, but cells can have varying Z dimensions and a curvilinear geometry.

B.2 General Options

The first step in the GMS interface includes the definition of general T-PROGS options. When a user selects the command to initialize a T-PROGS simulation, *New Simulation*, the dialog illustrated in Figure CC or Figure DD appears. If boreholes do not exist in the model, an unconditioned simulation will be generated. In this case, the user

selects the materials to be used and a corresponding background material. The “Materials Editor” button enables users to quickly create, delete, or rename materials and the material list is automatically regenerated.

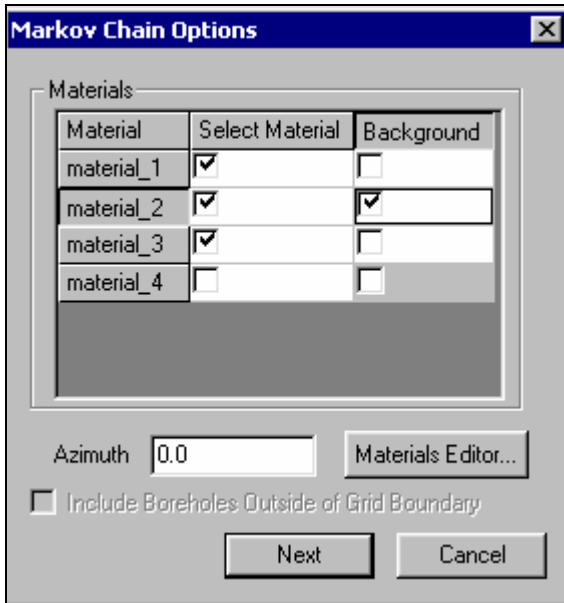


Figure CC Markov chain general options dialog without borehole data.

If boreholes do exist in the model, the list of materials is automatically generated as in Figure DD. Since this list is inherited from the boreholes, the user cannot alter the list of materials and so the *Materials Editor* button is dimmed. In addition to generating the material list, the code parses through the boreholes to compute the most dominant material and automatically selects the default background material. However, the user can change the background material. It is possible that the grid defined by the user does not circumscribe all of the borehole data. In this case, the user can opt to use all the boreholes in the model to generate the transition trends. Or the user can only use the

boreholes inside the grid boundary for generating the transition trends. The user specifies this option with the button, “Include Boreholes Outside of Grid Boundary.” The use of the boreholes inside the grid boundary is the default.

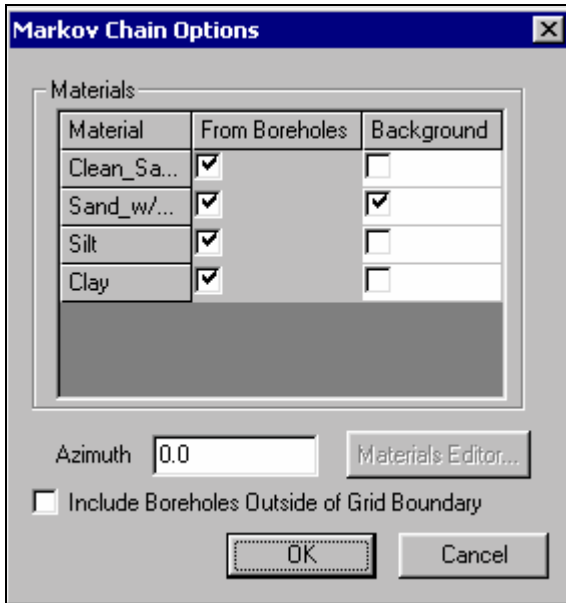


Figure DD Markov chain general options dialog with borehole data.

The user also enters an azimuth in this dialog. The azimuth is illustrated in Figure EE and determines the orientation of the primary directions of the depositional trends in the strike/dip directions. These trends generally are aligned with the primary directions of horizontal flow in the aquifer. Theoretically, the azimuth can be oriented independently from the grid orientation. However, in practice, if the grid and azimuth orientations are offset by more than about 40 degrees, checkerboard patterns appear in the indicator array results. Hence, the azimuth orientation is set equal to the grid orientation by default. However, the grid angle is defined counterclockwise as illustrated in Figure

FF, and the azimuth angle is clockwise as illustrated in Figure EE. Therefore, if the grid angle is 40 degrees, then the azimuth angle will be -40 degrees by default.

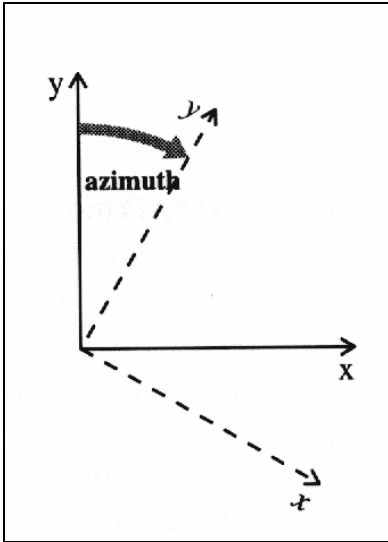


Figure EE Definition of azimuth in T-PROGS algorithm

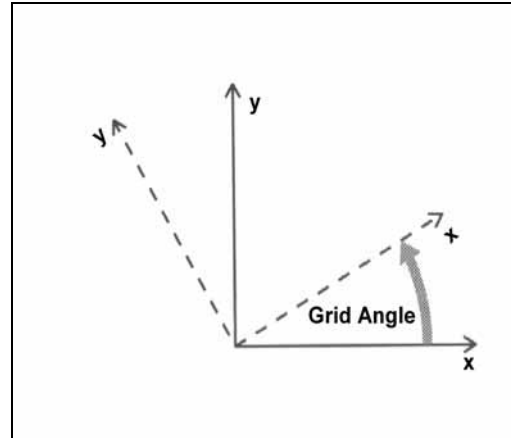


Figure FF Grid angle orientation.

One limitation for both the cases with and without boreholes is that a maximum of five materials can be used in the T-PROGS algorithm. This limitation was imposed to keep the data processing and user-interface reasonably simple. Although five materials present a limitation, borehole data can generally be easily condensed down to five or fewer materials. Furthermore since this is a stochastic approach, which is based on probability, the detail generated with numerous materials is rarely justifiable anyway. In addition, as the number of materials increase, the ratio of process time to detail becomes inefficient.

B.3 Vertical (Z) Markov Chains

Once the general options are input, the Vertical (Z) Markov Chains dialog illustrated in Figure GG appears. This dialog is composed of three main sections: a plot section, a Markov Chains section, and a spreadsheet section.

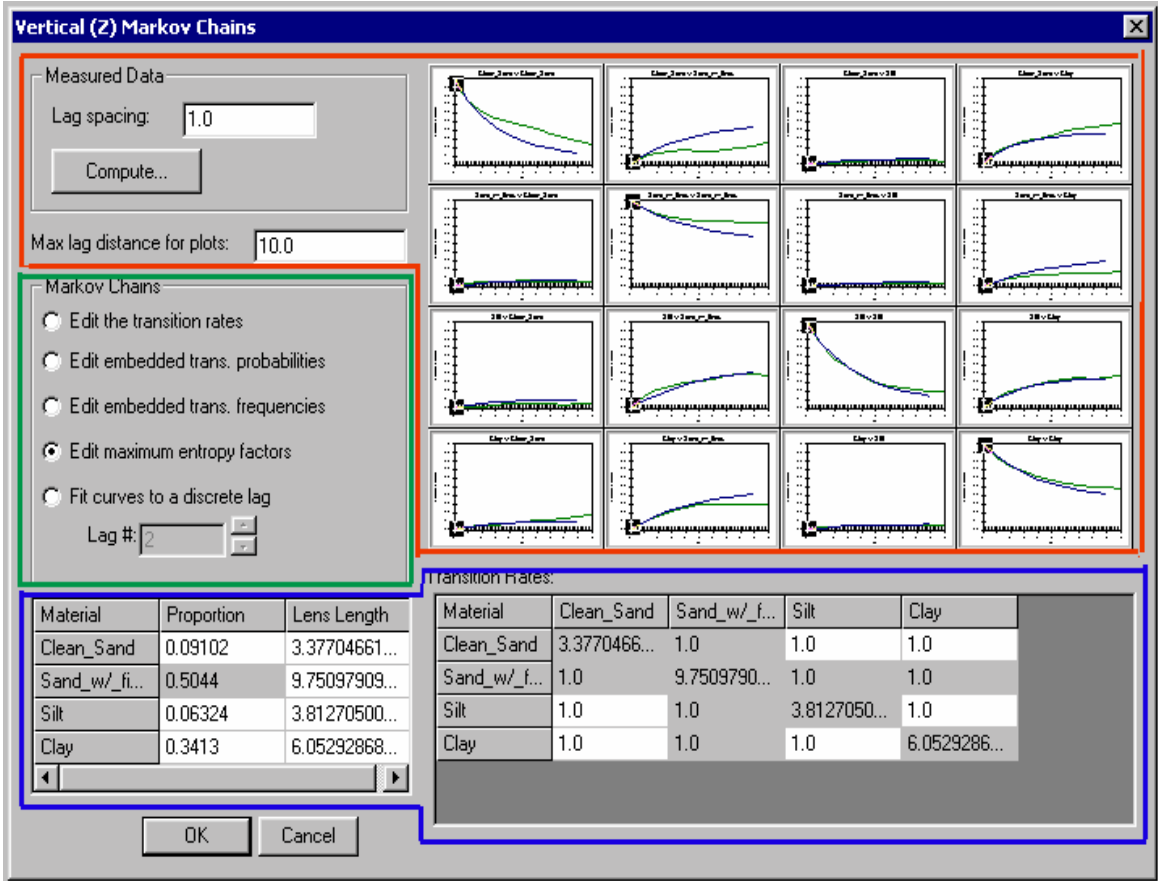


Figure GG Vertical Markov chain dialog.

All three sections enable the user to develop a 1-D Markov chain in the vertical direction.

B.3.1 Plot Section

The plot section (area in red) includes the array of curves, the *Lag spacing* edit

field, the *Compute* button, and the *Max lag distance for plots:* edit field. The number of plots in the array produced correlates to the number of materials used in the simulation. If N materials are used, an N by N array of plots will illustrate the transition probabilities for each material with respect to every other material. Every plot is labeled with a name and units and can be maximized with a command in the menu produced by right-clicking on the curve in question (Figure HH).

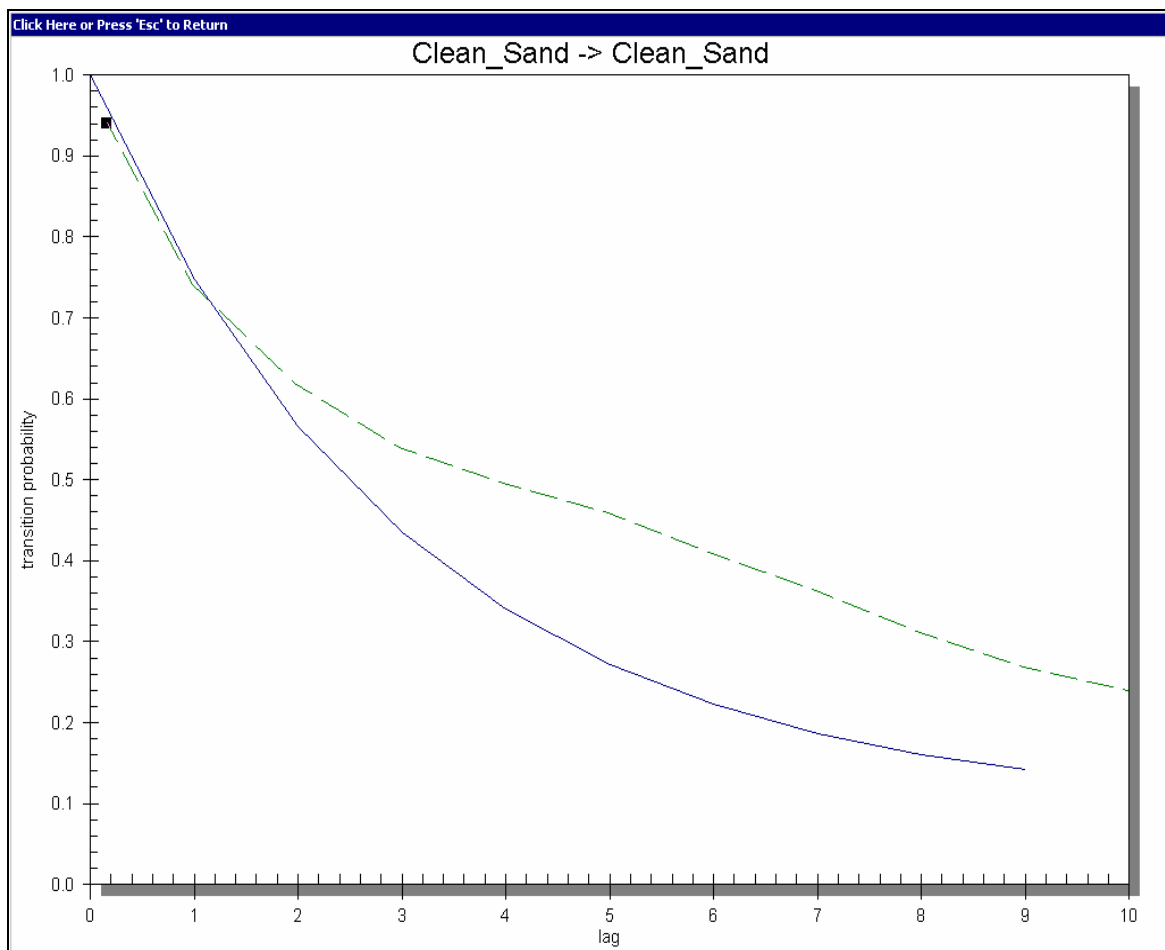


Figure HH Maximized plot.

The curves are automatically regenerated anytime a change is made in the other sections of the dialog.

By default, the curves plot transition probabilities that correspond to an equal probability of each material occurring as illustrated in Figure II.

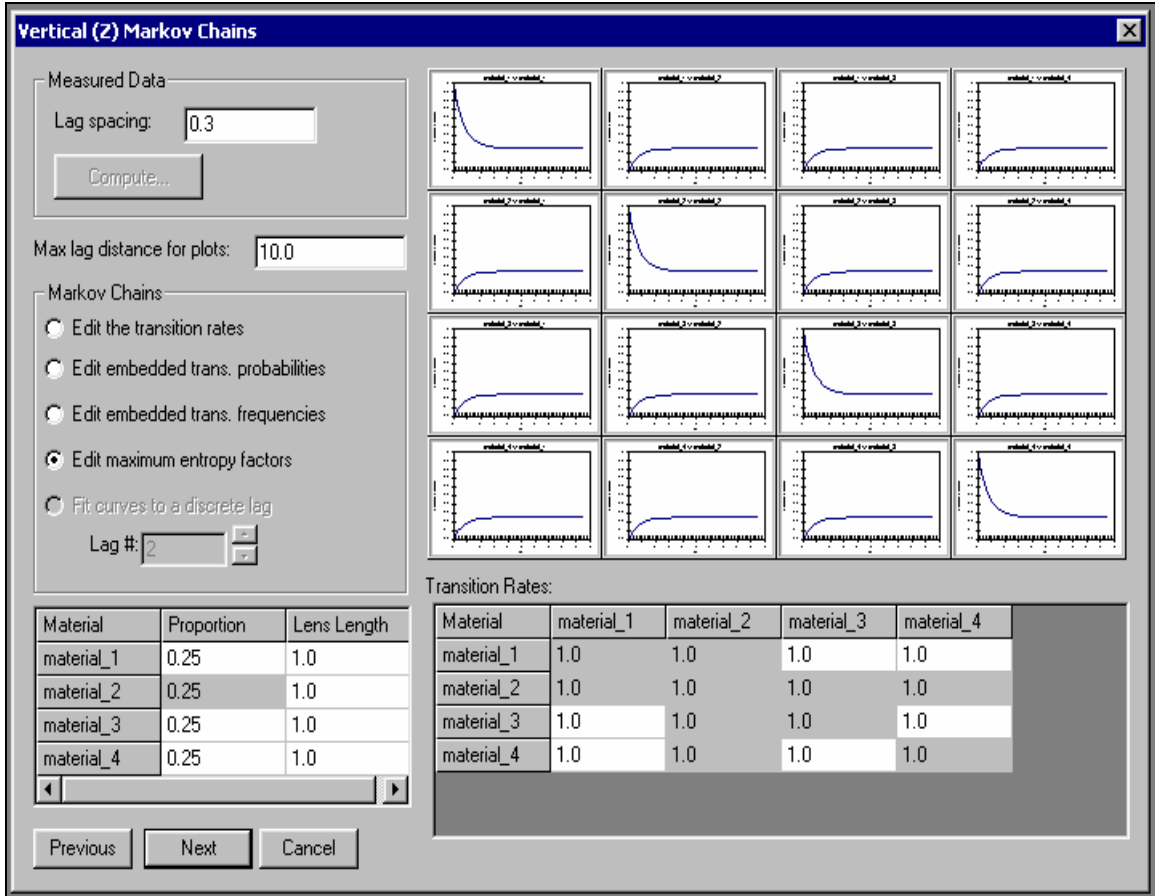


Figure II Curves with equal probability.

If boreholes exist in the model, the *Compute* button becomes undimmed. When the user clicks the *Compute* button, the parameter files required for running GAMEAS are generated and GAMEAS is executed. A separate window displays the details of the GAMEAS run, which allows the user to monitor the output from the GAMEAS simulation as shown in Figure JJ.

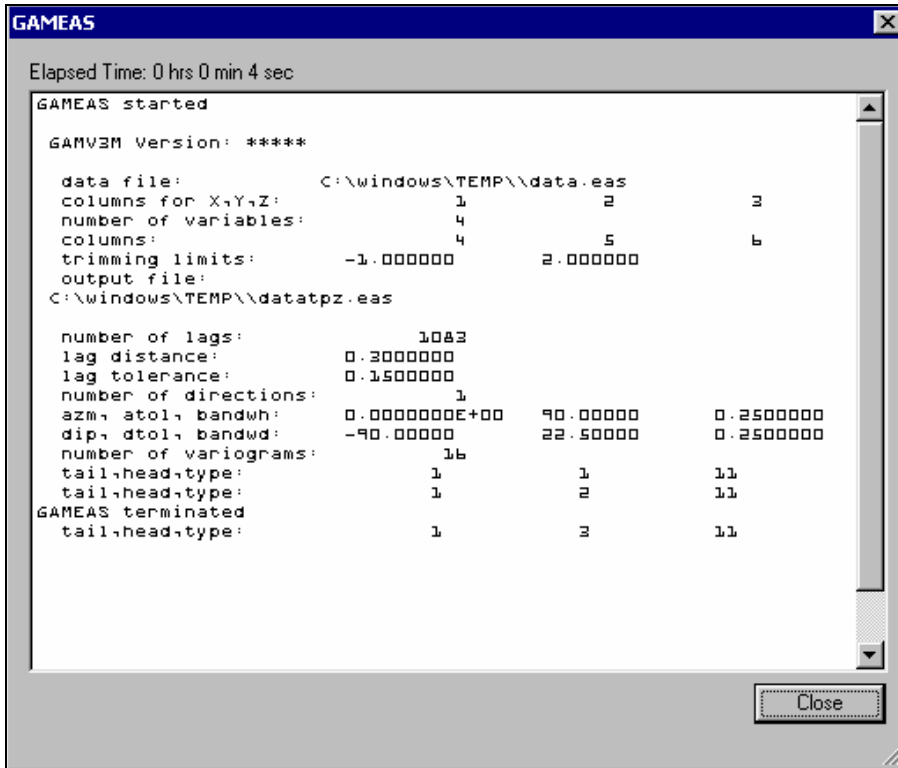


Figure JJ GAMEAS executable inside GMS interface.

When GAMEAS completes a successful run, the results, including the material proportions and transition probability curves from the measured data, are read into the corresponding data fields in the dialog as illustrated in Figure GG. Furthermore, the transition rates which correspond to the slope of the transition probability curves when $h_\phi = 0$ (see Chapter 3) are interpolated from the measured data curves. In addition to running GAMEAS, the code parses through the borehole data and calculates the embedded transition probabilities and frequencies as defined in Chapter 3. These values are stored in arrays that correspond to options #2 and #3 in the *Markov Chains* section of the dialog in Figure GG.

In addition to the curves generated from the measured data, a set of curves is

generated from the Markov chains as defined by the option selected in the *Markov Chains* section. In Figure GG, the measured data curve is green, and the Markov chain curve is blue. Hence, each plot contains two curves: a curve from the measured data and a curve from the Markov chain definition. Only the Markov chain curve is used in later calculations. The curve from the measured data plays a curve-fitting role. Generally, the parameters in the dialog are changed until the Markov chain curve mimics the measured curve as illustrated in Figure GG.

The *Lag spacing* determines how dense the curves are, and the *Max lag distance for plots* determines the range of the curves. The curves always range from 0 to the *Max lag distance for plots* horizontally, and they range from 0 to 1 vertically to honor probability constraints.

B.3.2 Markov Chains Section

There are five alternate methods of generating Markov chains in this section (area in green in Figure GG). Four of these five methods correspond to the theory described in the Markov chain section of Ch. 3. Furthermore, these methods are in a radio-button configuration to allow the user to conveniently change from one method to another. Each of these five methods will be described. The first four options have previously been described in the Markov Chain section of Chapter 3. The curves used to illustrate the different methods in the following section were derived from borehole data extracted from the Longhorn Army Ammunition Plant (LHAAP) in Eastern Texas.

The first option, *Edit the transition rates*, is exactly that. One can directly edit the

array of transition rates that are listed in the *Transition Rates* section. The theory that supports this option is described in the Markov chain section of Ch. 3. This option is useful after selecting the *Compute* button and running GAMEAS because slopes can be inferred from the measured data curves. GAMEAS outputs transition probability curves.

Transition rates used in this option

correspond to the slope of the transition probability curve at a lag, $h(\phi) = 0$. When reading the output from GAMEAS, the transition probability rates, $r_{jk,\phi}$, are interpolated as

$$r_{jk,\phi} = 0.57 * r1_{jk,\phi} + 0.29 * r2_{jk,\phi} + 0.14 * r3_{jk,\phi} \dots\dots\dots \text{Eq. B.1}$$

where r1, r2, and r3 are the slopes defined by a straight line from the origin out to lag1, lag2, and lag3 respectively. It will be remembered that the transition rate obeys Eq.

3.11. Therefore, as the lag approaches zero, more weight should be given to the corresponding slope. Hence, a weight of 0.57, 0.29, and 0.14 were assigned to r1, r2, and r3 respectively. Once the slopes are computed for each entry in the matrix, the mean lengths for each category are computed by

$$\overline{L}_{j,\phi} = \frac{-1}{r_{jj,\phi}} \dots\dots\dots \text{Eq. B.2}$$

Figure KK illustrates the Vertical (z) Markov Chain dialog when option #1 is selected.

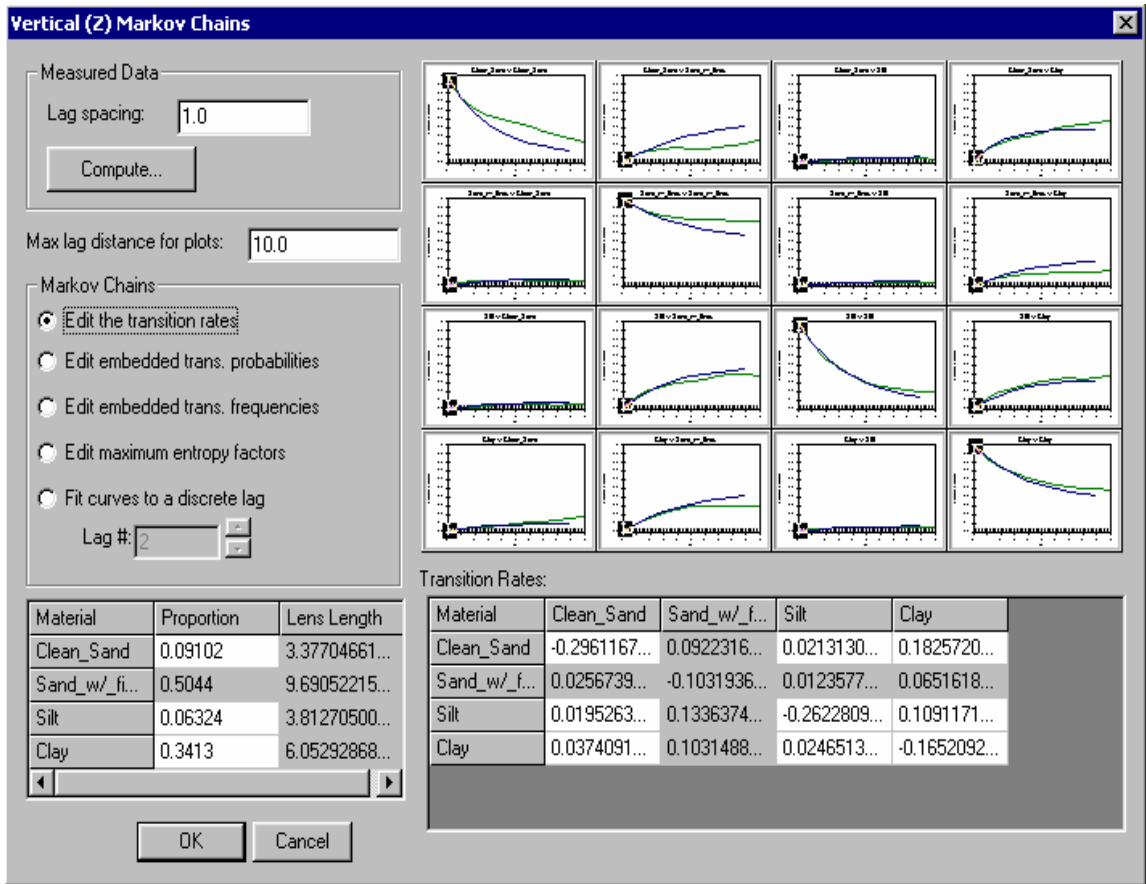


Figure KK Vertical (Z) Markov Chain dialog with option #1 selected.

Regardless of which Markov Chain option is selected, the background row and column, *Sand_w/_fines*, is dimmed because the values in this row and column are automatically computed from the remaining entries by probability constraints described in the Markov chain section of Ch. 3. In addition, with this option selected, the *Lens Length* column is also dimmed because the lens lengths are automatically computed and updated from the diagonal terms in the *Transition Rates* spreadsheet. The diagonal terms of the *Transition Rates* spreadsheet must be negative to obey probability rules. With this data, this method produces an accurate fit between the measured (green) and the Markov chain (blue) curves at small lag spaces.

The second option is *Edit embedded trans. probabilities*. This is a more intuitive method of generating Markov chains and is conducive to sites with and without data. It is conducive to sites with data because the embedded transition probabilities can be determined from the borehole data. When a simulation is initialized, if borehole data exist default embedded transition probabilities are computed from the borehole data. If borehole data do not exist, the embedded transition probabilities can be estimated with some basic geologic knowledge including the average mean lengths of each material for each direction and depositional trends. Figure LL illustrates the curves generated with this method.

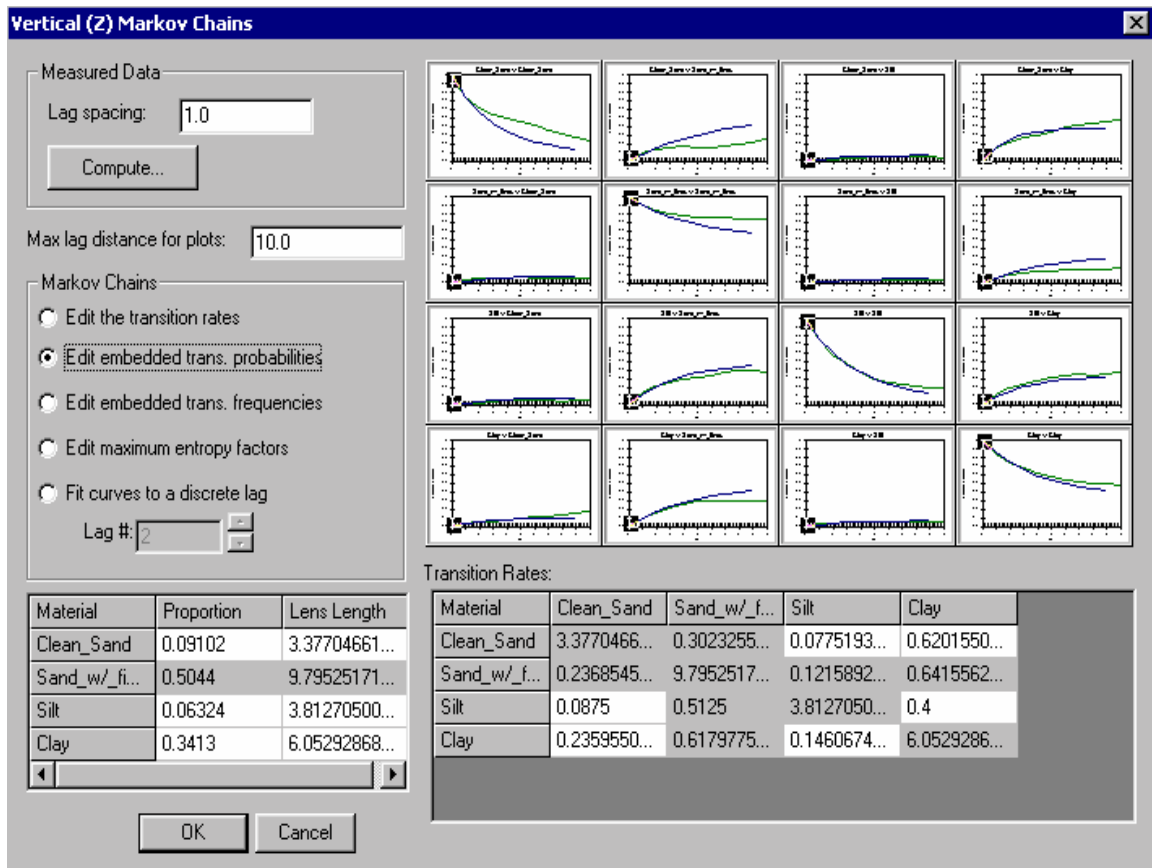


Figure LL Vertical (Z) Markov Chain dialog with option #2 selected.

With this option selected, the diagonal terms are dimmed because these values are derived from the values entered in the *Lens Length* column using Eq. B.2. Adjusting the proportions, lens lengths, or the off-diagonal terms in the *Transition Rates* spreadsheet alters the curves.

The third option in the Markov Chains section is *Edit embedded trans.*

Frequencies (Figure MM).

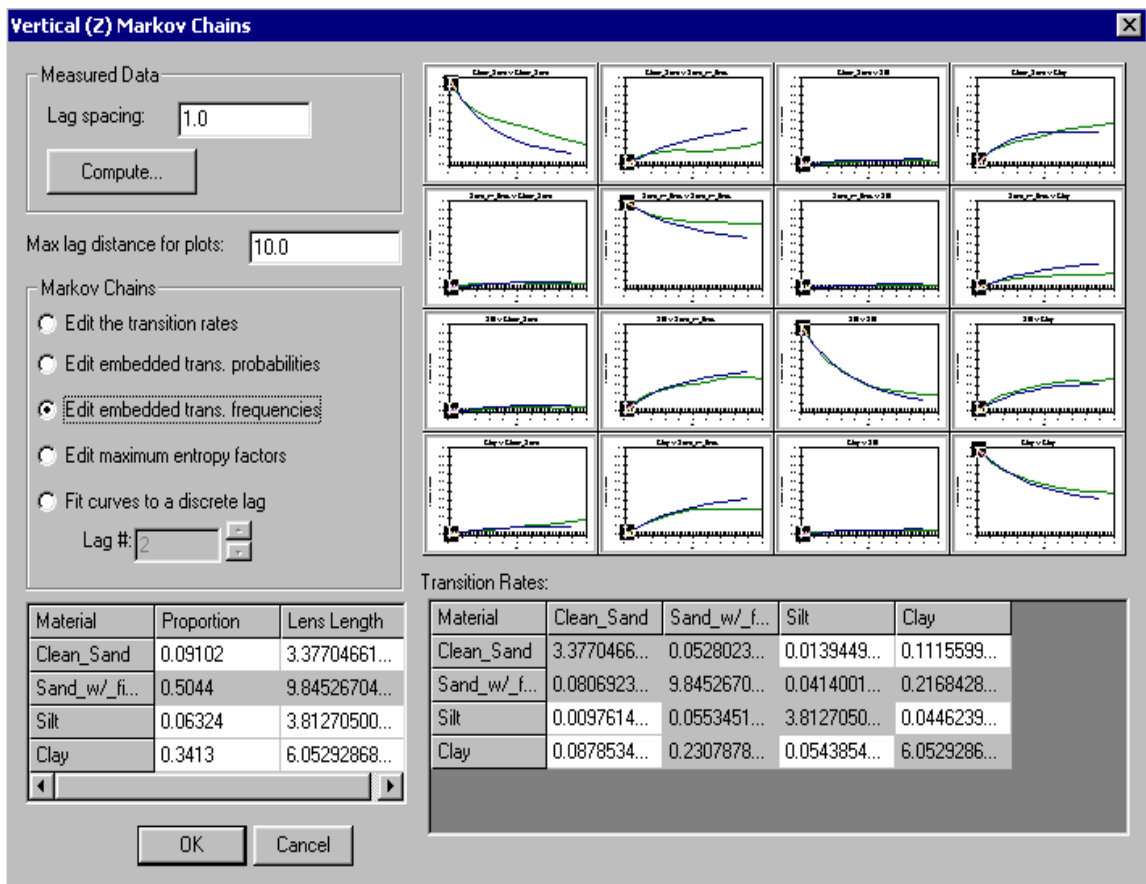


Figure MM Vertical (Z) Markov Chain dialog with option #3 selected.

This option is similar to option #2, except embedded transition frequencies populate the off-diagonal terms of the *Transition Rates* spreadsheet. Embedded transition frequencies

are also computed from borehole data when a simulation is initialized if borehole data exist. Once again, this is an intuitive framework with geologic knowledge including average mean lengths and depositional trends.

The fourth option is *Edit maximum entropy factors* (Figure NN).

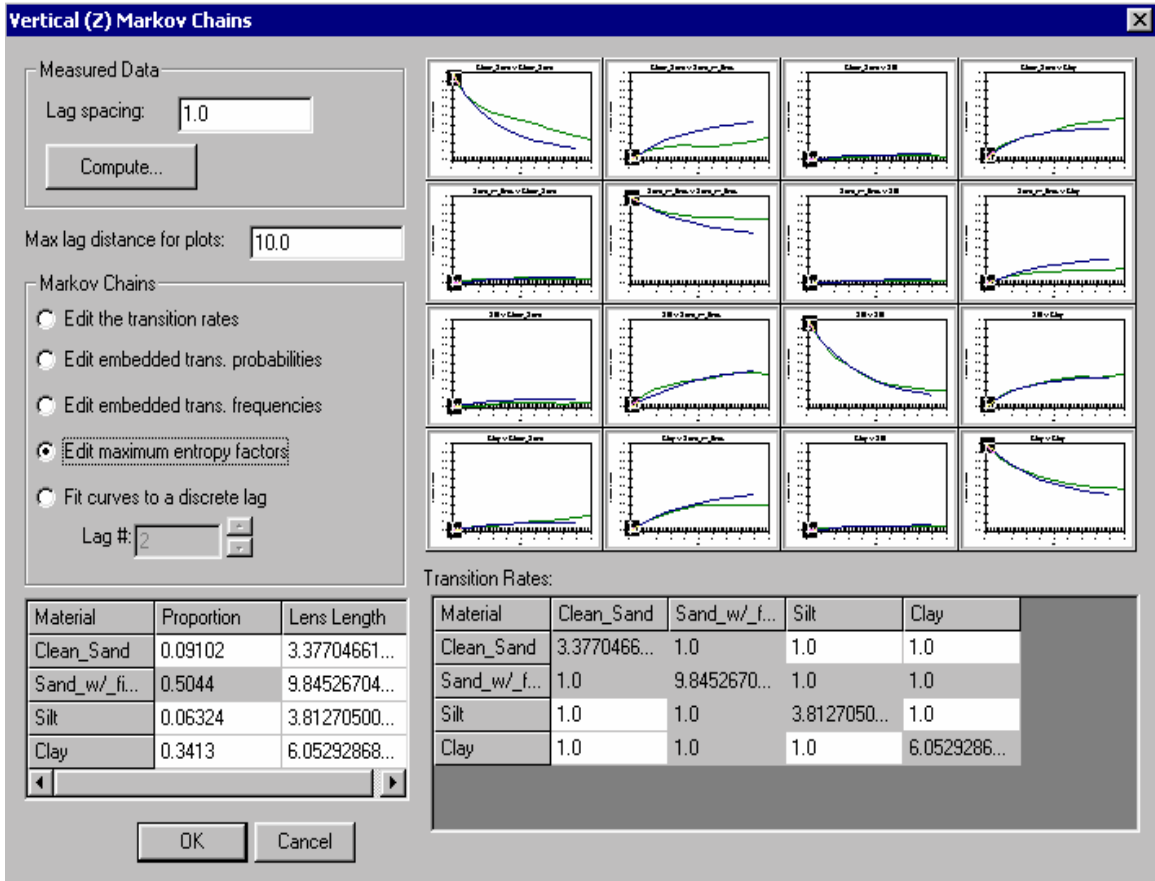


Figure NN Vertical (Z) Markov Chain dialog with option #4 selected.

While this option can be used with borehole data, it is ideally it is ideally suited for cases without borehole data. With this option, the user edits the proportions for all but the background material and the means lens lengths for all materials. The lens lengths are used to populate the diagonal terms of the *Transition Rates* spreadsheet using Eq. B.2,

and the maximum entropy factors fill the off-diagonal terms of the spreadsheet. As described in Chapter 3, the maximum entropy factors represent the ratio of the transition rate to the maximum entropy transition rate. A maximum entropy factor of 1.0 represents maximum disorder in depositional tendencies. A rate greater than 1.0 indicates that the two categories tend to occur next to each other. A factor less than unity would infer the opposite. This is an intuitive method of generating Markov chains and is conducive to all types of sites. This method enables logical incorporation of anisotropy into the model with the maximum entropy factors.

The fifth option, *Fit curves to a discrete lag*, is only undimmed if a transition probability curve from measured data exists in memory. When this option is selected, the *Lag #* edit field is undimmed and the user enters the discrete lag the curves will be fit to (Figure OO).

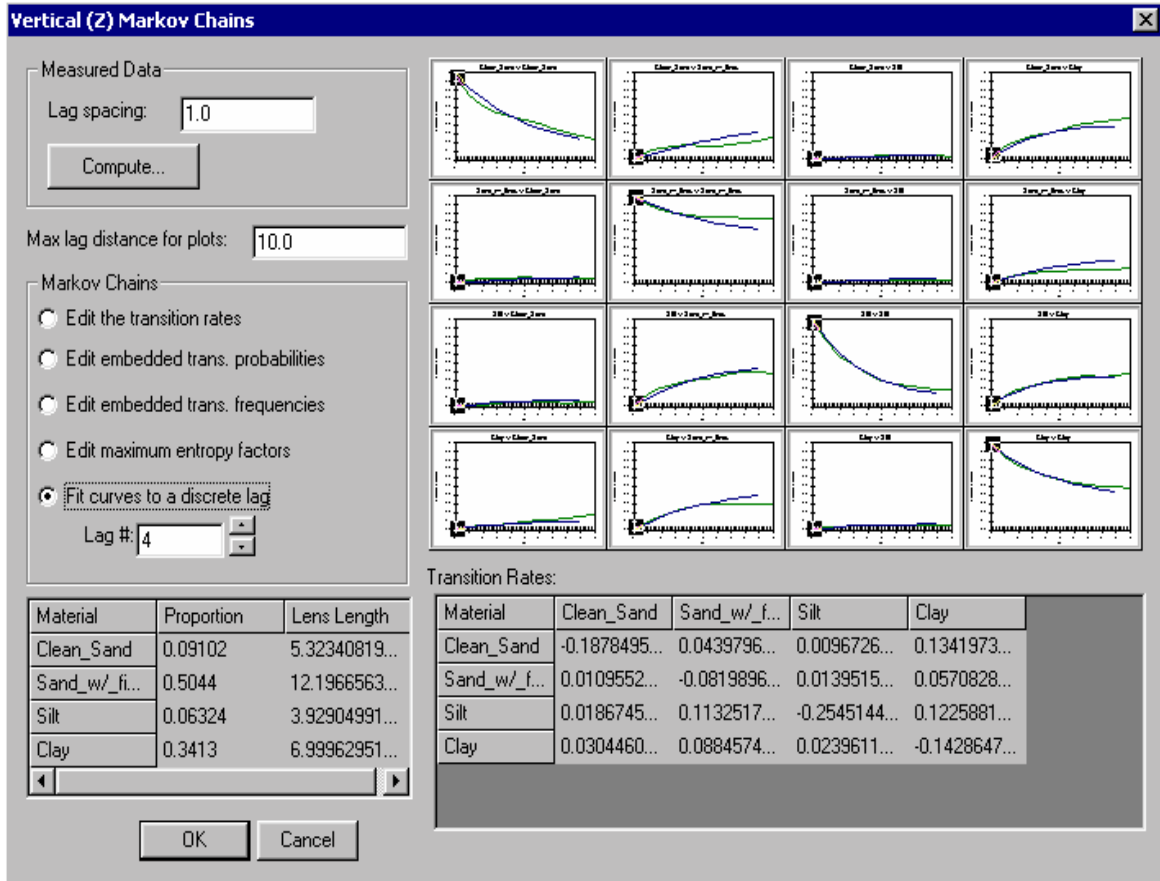


Figure OO Vertical (Z) Markov Chain dialog with option #5 selected.

This option produces Markov chains that are computed from the measured transition rates produced by GAMEAS. GAMEAS computes a set of transition probabilities at each lag specified by the user. This option computes an array of transition rates from the slope of the curves generated by GAMEAS (green). The transition rates correlate to the shape of the curve from the origin to the lag # specified by the user in the *Lag #* edit field in Figure OO. Therefore, the cells of all the spreadsheets in the dialog are dimmed because the values are inherited from the measured curves (green). The values in the spreadsheets change depending on the lag # entered.

B.3.3 Spreadsheet Section

This section, area in blue in Figure GG, includes two separate spreadsheets: *Transition Rates* and *Proportions & Mean Lengths* spreadsheets. The *Transition Rates* spreadsheet contains the rate entries that correspond to the selected option in the *Markov Chains* section. The *Proportions* column holds the proportions for each material. The *Mean Lengths* column contains the average mean length in the vertical direction for each material.

B.4 Strike (X) Markov Chains

Once the vertical Markov chains have been defined, the user is then presented with the *Strike (X) Markov Chains* dialog. This dialog has the same setup as the vertical dialog (Figure PP).

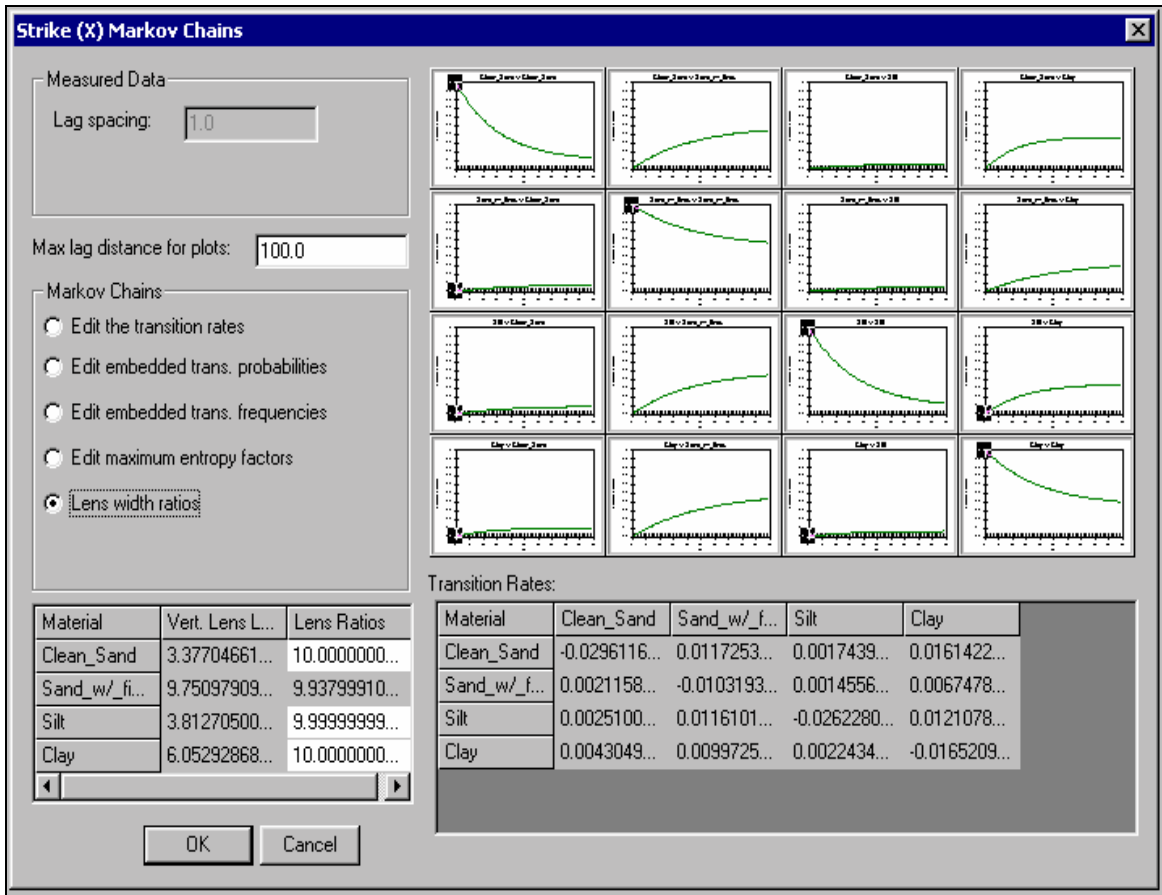


Figure PP Strike/Dip Markov chain dialog.

There are three minor differences between this dialog and the vertical dialog. The first difference is in the *Markov Chains* section. The option, *Lens width ratios*, replaces the *Fit curves to a discrete lag* option. The *Fit curves to a discrete lag* option is not applicable because horizontal measured transition probability curves do not exist due to lack of data in the horizontal direction. The *Lens width ratios* option is the default option and should be used in most cases. This option allows users to apply the transition data entered in the vertical direction to the horizontal direction. The transition rate matrix is automatically populated based on the Walther's Law assumption and by using the three-step process described in the *Multidimensional Markov Chains* Section of Chapter 3. The

proportion data are directly inherited from the vertical data. In addition, rather than entering the mean lengths, the user enters a ratio corresponding to the ratio of the lens length in the x direction to the lens length in the z direction. For example, if the lens length for material A is 5 feet in the z direction and the user enters 10.0 for the ratio, then the lens length in the x direction would be 50 feet. Due to the lack of measured data in the horizontal direction, another minor change is that each plot in the *Plot Section* contains only one curve: the Markov chain curve. The *Lens Width Ratios* option is particularly useful because the only required input is lens length ratios for the non-background materials. The remaining data are all inherited from the vertical data

B.5 Dip (Y) Markov Chains

The dip dialog has an identical appearance and functionality as the strike dialog.

B.6 Running TSIM

Once the general options and the 1-D Markov chains in all three primary directions have been defined, a simulation can be generated. Selecting the *Run TSIM* command in the T-PROGS menu in Figure BB initializes this process.

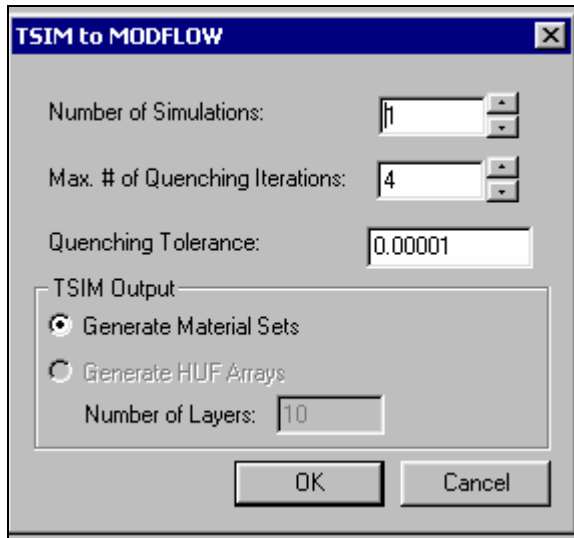


Figure QQ General options for simulation generation.

The command produces the dialog shown in Figure QQ. This dialog contains some general options including the number of simulations created, maximum number of quenching iterations, and the type of output desired. Either material sets or HUF arrays can be generated. A material set is an array of material ids that correspond to each cell in the grid, and each material has hydraulic parameters associated with it. HUF arrays are arrays of material ids that can be independent of the number of grid layers in the grid. This option is only undimmed if MODFLOW is initialized. In Figure QQ, MODFLOW has not been initialized so this option as well as the *Number of layers*: edit field is dimmed.

When the *OK* button is selected on the *TSIM Options* dialog, the parameter files for MCMOD are generated and the executable is launched in the same way as the GAMEAS executable (Figure RR).

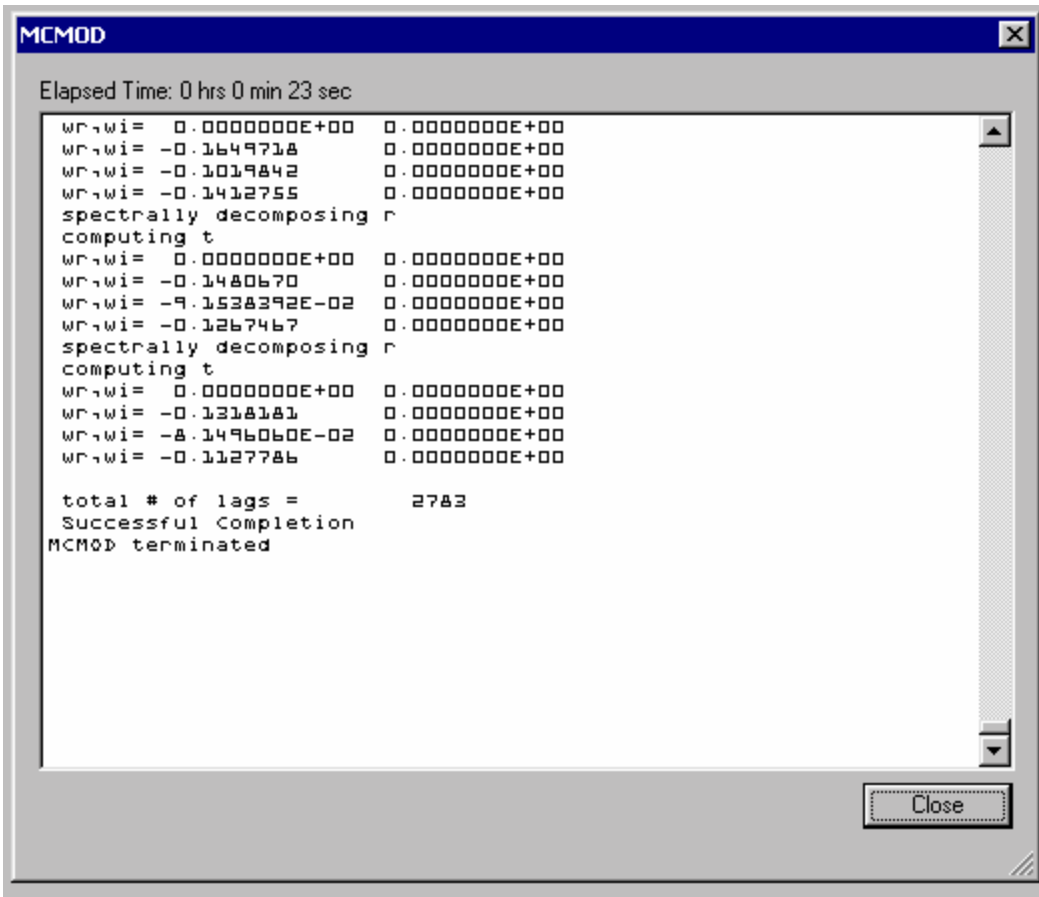


Figure RR MCMOD executable inside GMS interface.

After a successful completion, TSIM is automatically executed. After its successful completion the output of spatial variability is read into the appropriate material set arrays or HUF arrays as prescribed in the *TSIM Options* dialog. Both for the MCMOD and TSIM executable, the time remaining for completion is estimated and displayed in the top right-hand dialog as illustrated in Figure RR. Furthermore, prior to completion, one can exit the program by selecting the *Abort* button in the bottom of the dialog displayed in Figure SS.

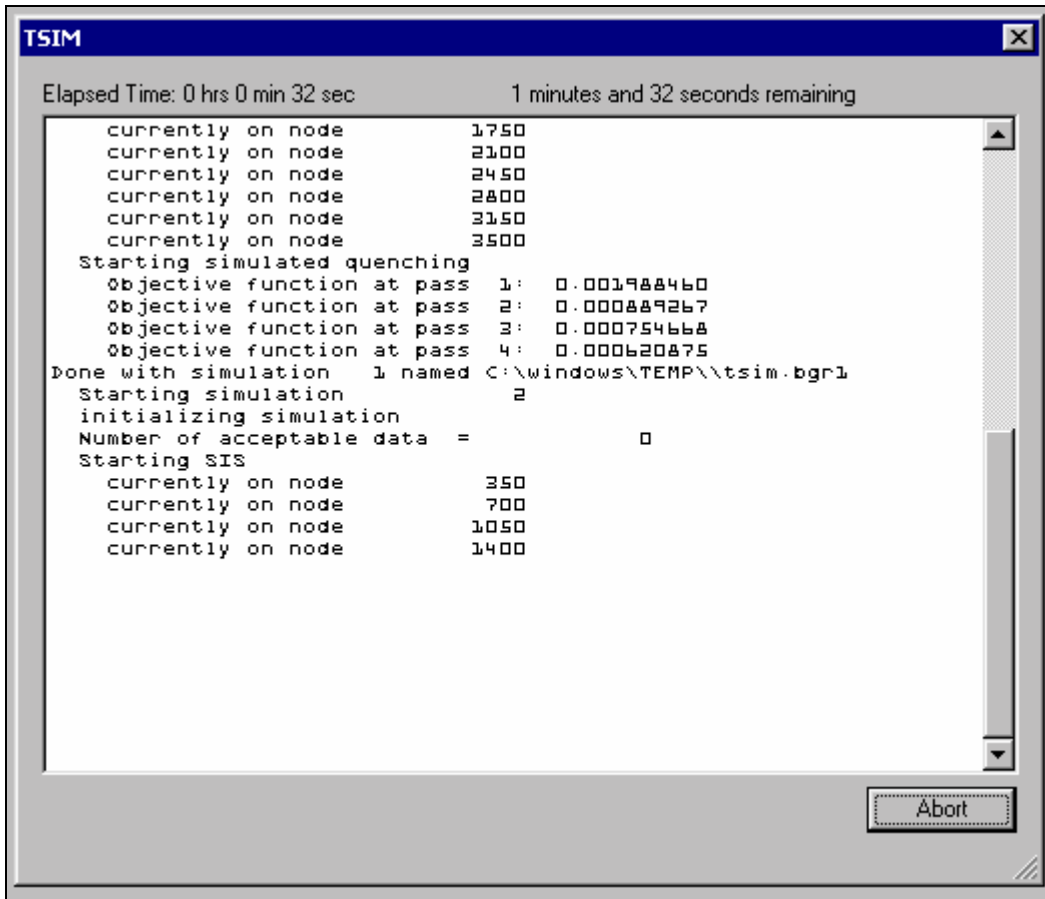


Figure SS TSIM executable inside GMS interface.

Only one T-PROGS simulation can exist at once. The data associated with a T-PROGS simulation is written out and read in with the GMS project data. The interface is user-friendly with sufficient warning and HELP messages throughout the interface.

B.7 Post-Processing

The indicator arrays generated by T-PROGS are read into material sets or HUF arrays as outlined in Chapters 4 and 5. If a simulation of N material sets is generated, all N material sets are organized into a *Material Set* folder stored in the 3-D grid module inside a data tree window as illustrated in Figure TT.

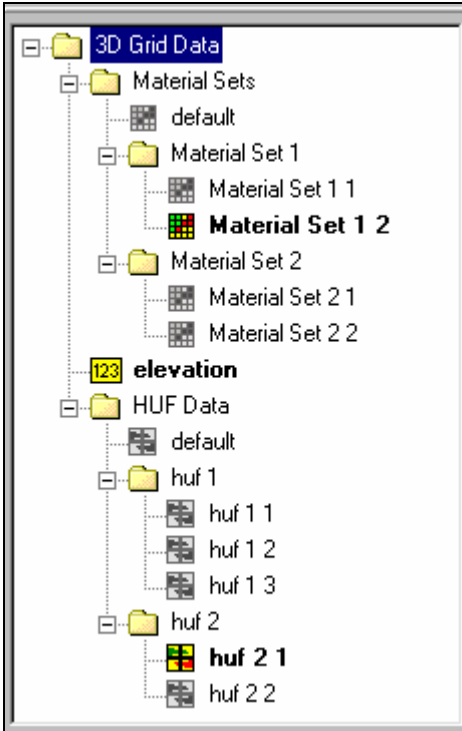


Figure TT Data tree used to organize HUF and Material sets.

The simulation folders can be renamed, moved, or deleted by the user by right-clicking on the item or by simply dragging the folder. Furthermore, not only can multiple realizations be generated, but also multiple sets of realizations can be generated, saved, organized, and manipulated efficiently and simply. The different indicator arrays or material sets are applied to the grid by simply selecting the corresponding item in the data tree and the grid display is automatically updated. The active material set or data set that is currently incorporated into the grid is marked with a bolded name and icon. In

Figure UU, the grid material identification array is populated with data from the data from the active material set, Material Set 2 2. Another material set simulation exists as well as HUF data set simulations.

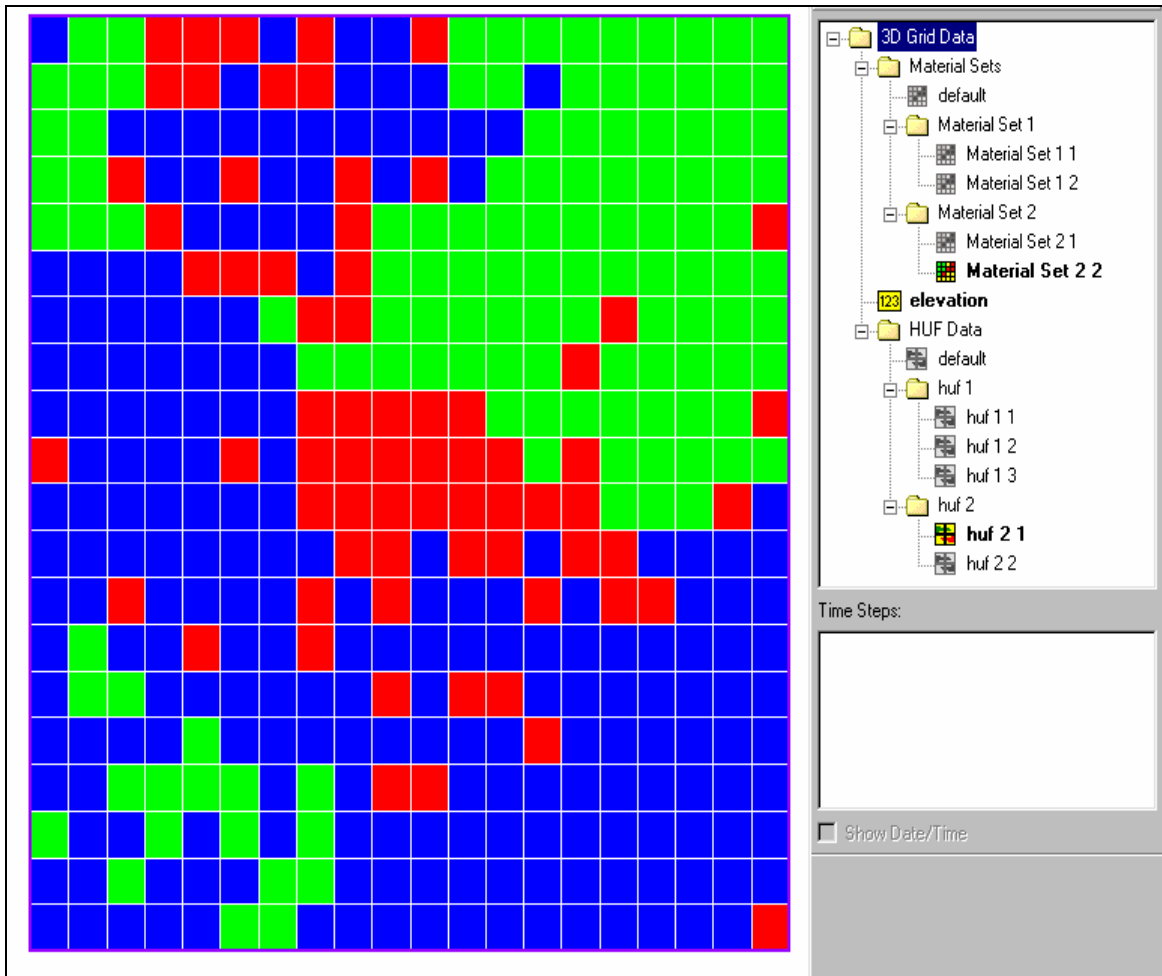


Figure UU Grid populated with data from Material Set 2 2.

HUF arrays are stored and organized in a similar way. Except the HUF data are all stored in a *HUF Data* folder.