Characterization of subsurface heterogeneity: Integration of soft and hard information using multi-dimensional Coupled Markov chain approach

Eungyu Park[1+], Amro Elfeki[2*], and Michel Dekking[3]

1. Environmental Sciences Division, Oak Ridge National Laboratory, PO Box 2008, MS6036, Oak Ridge, TN 37831-6036, USA.
+ Contact: parke@ornl.gov

2. Section Hydrology and Ecology, Faculty of Civil Engineering and Geosciences, Delft University of Technology P.O. Box 5048, 2600 GA, Delft, The Netherlands

3. Section of Applied Probability, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

Abstract

Subsurface heterogeneity is a complex mixture of discrete structures characterized by more or less discontinuous boundaries (e.g., lithologies, faults) and random features that may be characterized by statistical models. The development of models capable of mathematically mimicking such complex variability has proven difficult. Conventional semivariogram-based geostatistical methods have difficulty incorporating geologic interpretations into models and are unable to describe non-stationary distributions and asymmetric juxtapositional tendencies. Transition probability-based indicator models have been proposed to address these problems. Carle and Fogg (1997) proposed a Markov chain model in conjunction with a conventional sequential indicator simulation algorithm. Elfeki and Dekking (2001) developed a two-dimensional coupled Markov chain model that uses a conditional simulation algorithm with explicit transitional conditioning probability equations. More recently, Park et al. (2002) extended this model to three-dimensional space and improved the algorithm to more readily utilize sparsely located hard data. Advantages of the coupled Markov chain approach include: (1) the procedure is simpler than conventional sequential indicator simulation because the computation of the transition probability matrix does not require parametric fitting of a semi-variogram model, (2) asymmetric heterogeneity structures can be modeled because there is no intrinsic symmetry assumption in the coupled Markov chain model, (3) conditioning to measured values is straight-forward, and (4) geological observations and principles (e.g., fining up/down sequences) can be directly implemented in the transition probability matrix. In this extended abstract, we will demonstrate a three-dimensional coupled Markov chain model to characterize subsurface heterogeneity using hypothetical borehole lithology data.

1. Introduction

Characterization of the subsurface is an important step for subsurface flow and transport simulations that may reveal useful information about the existing heterogeneity of the fields. The subsurface is inherently not fully accessible with existing techniques, hence there exist huge amounts of uncertainties that need to be resolved to produce a plausible picture of the ground. Geostatistical simulations are currently the most promising solutions to these uncertainty problems and their effectiveness has been proven in many case studies. Indicator-based simulation methods are well suited for many geological characterization cases (Carle and Fogg, 1996). They have been intensively used in many fields such as soil, stratigraphy, hydrogeology, sedimentology, etc. In most of these simulations, indicator variograms have been widely applied.

---

[*] on leave from Faculty of Engineering, Mansoura University, Mansoura, Egypt.

Recently, Carle and Fogg (1997) and Carle et al. (1998) developed a new type of sequential indicator simulation (SIS) algorithm that uses Markovian transition probabilities instead of the indicator variograms. The Markovian property states that the conditional distribution of any future state given the past and the present states, is independent of the past history and depends only on the present state (Ross, 2000). The approaches of Carle et al. (1998) brought many improvements in terms of asymmetry, which could not be modeled by conventional geostatistical approaches. By using the transition probability approach, many soft and hard data in an early stage of the model can be utilized. In the approach presented by Carle and Fogg (1998), the transition probabilities are built in a Markovian framework in the estimation phase, however in the simulation phase they use the conventional simulation methods (i.e. sequential indicator simulation (SIS) followed by simulated quenching technique). Therefore, their approach loses some merits of the Markovian transition probabilities. Elfeki and Dekking (2001) developed a two-dimensional conditional indicator simulation algorithm using Markovian transition probabilities under a Markovian framework. This model is directly branched from the one-dimensional Markov chain model developed by Krumbein (1967). The coupled Markov chain (CMC) model is very convenient to implement for stochastic simulation because the CMC model does not require parametric fitting of a semi-variogram model nor cumbersome indicator co-kriging techniques. Also conditioning is simple by using explicit formulae. The main objective of this study is to extend the CMC to 3D in end to make the model better suited for many practical problems.

2. Theoretical Background

In the Markovian framework, the conditional distribution of any future state is independent of the past history if the present state is given. In formulas, let the discrete stochastic process $\{Z_i, i = 0,1,2,\ldots\}$ be a sequence of random variables taking values in the state space $\{S_1, S_2,\ldots, S_M\}$. The sequence is a Markov chain or a Markov process if,

$$\Pr\left(Z_i = S_k \,|\, Z_{i-1} = S_l, Z_{i-2} = S_n, Z_{i-3} = S_r, \cdots, Z_0 = S_p\right) =$$
$$\Pr\left(Z_i = S_k \,|\, Z_{i-1} = S_l\right) = p_{lk}. \tag{1}$$

In one-dimensional problems a Markov chain is described by a single transition probability matrix. Transition probabilities correspond to relative frequencies of transitions from certain state to other states. These transition probabilities can be arranged in a square matrix form,

$$\boldsymbol{p} = \begin{bmatrix} p_{11} & \cdot & \cdot & \cdot & p_{1M} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & p_{lk} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{M1} & \cdot & \cdot & \cdot & p_{MM} \end{bmatrix}, \tag{2}$$

where p_{lk} denotes the probability of transition from state S_l to state S_k, and M is the number of states in the system. Thus the probabilities of a transition from S_1 to $S_1, S_2,\ldots, S_M$ are given by $p_{1l}, p_{12},\ldots, p_{1M}$ where $l = 1,2,\ldots, M$ and so on. The matrix $\boldsymbol{p}$ has to fulfill specific properties: (1) its elements are non-negative, $p_{lk} \geq 0$; (2) the elements of each row sum up to one. The transition probabilities considered in Eq. 2 are called one-step transitions. One considers also N-step transitions, which means that transitions from a state to another takes place in N steps. The N-step transition probabilities can be obtained by multiplying the single-step transition probability matrix by itself N times. Under some mild conditions on the transition matrix (aperiodicity and

irreducibility), the successive multiplications lead to identical rows (w_1, w_2..., w_M). So the w_k (k=1, 2,...,M) are given by

$$\lim_{N \to \infty} p_{lk}^N = w_k, \tag{3}$$

and are called marginal probabilities, where w_k is no longer dependent on the initial state S_l.

The aforementioned method is a discrete time Markov chain model, and the methods for building continuous-time Markov chains are also available and thoroughly reviewed by Carle et al. (1998). It is worthy of note that using the continuous-lag Markov chains, we can build Markovian transition probabilities even out of sparsely distributed data, because the continuous-lag formulation can handle irregularly spaced data using the transition rate and the exponential decaying lag-transition model. This method is potentially helpful when we build the horizontal transition matrix using sparsely distributed lithological data (Carle et al. 1998).

3. 3D Coupled Markov Chain Model

The coupled transition probability $p_{lma,bfh}$ on the state space $\{S_1, S_2,..., S_M\} \times \{S_1, S_2,..., S_M\} \times \{S_1, S_2,..., S_M\}$ is given by, (Figure 1)

$$p_{lma,bfh} = p_{lb}^{h_x} \cdot p_{mf}^{h_y} \cdot p_{ah}^{v}. \tag{4}$$

These transition probabilities form a stochastic $M^3 \times M^3$ matrix.

For the three-dimensional generalization of the 2D CMC model, a triple coupling of one-dimensional Markov chains (X_i), (Y_j), and (Z_k) is used for constructing a three-dimensional spatial stochastic process (Γ_{ijk}) on a cubic lattice as shown in Figure 1. Each cell has a layer number k, a row number j, and a column number i. Using the analogy applied for developing the 2D CMC model (Elfeki and Dekking, 2001), one can write the three coupled chains on the cubic lattice by forcing these three independent chains to have the same outcome as,

$$p_{lmn,o} = \Pr\left(\Gamma_{i,j,k} = S_o \,\middle|\, \Gamma_{i-1,j,k} = S_l, \Gamma_{i,j-1,k} = S_m, \Gamma_{i,j,k-1} = S_n\right)$$
$$= \frac{p_{lo}^{h_x} \cdot p_{mo}^{h_y} \cdot p_{no}^{v}}{\sum_f p_{lf}^{h_x} \cdot p_{mf}^{h_y} \cdot p_{nf}^{v}}, \quad o = 1,...,M. \tag{5}$$

One can also write the conditional probability of cell (i,j,k) to be in state S_o, given the past (cell $(i-1,j,k)$ is in state S_l, cell $(i,j-1,k)$ is in state S_m, and cell $(i,j,k-1)$ is in state S_n), and the future (cell (N_x,j,k) is state S_p, and cell (i,N_y,k) is in state S_q) as

$$p_{lmn,o|p,q} = \Pr\left(\Gamma_{i,j,k} = S_o \,\middle|\, \Gamma_{i-1,j,k} = S_l, \Gamma_{i,j-1,k} = S_m, \Gamma_{i,j,k-1} = S_n, \Gamma_{N_x,j,k} = S_p, \Gamma_{i,N_y,k} = S_q\right)$$
$$= C \frac{p_{lo}^{h_x} p_{op}^{h_x(N_x-i)}}{p_{lp}^{h_y(N_x-i+1)}} \frac{p_{mo}^{h_y} p_{oq}^{h_y(N_y-j)}}{p_{mq}^{h_y(N_y-j+1)}} p_{no}^{v}, \tag{6}$$

where C is a normalizing constant which is given by

$$C = \left(\sum_r \frac{p_{lr}^{h_x} p_{mr}^{h_y} p_{nr}^{v} p_{rp}^{h_x(N_x-i)} p_{rq}^{h_y(N_y-j)}}{p_{lp}^{h_x(N_x-i+1)} p_{mq}^{h_y(N_y-j+1)}} \right)^{-1}. \tag{7}$$

Also in the derivation of Eq. (6), we assume that the conditioning data can only be found along the horizontal plane and the conditioning scheme goes from top to bottom layer by layer.
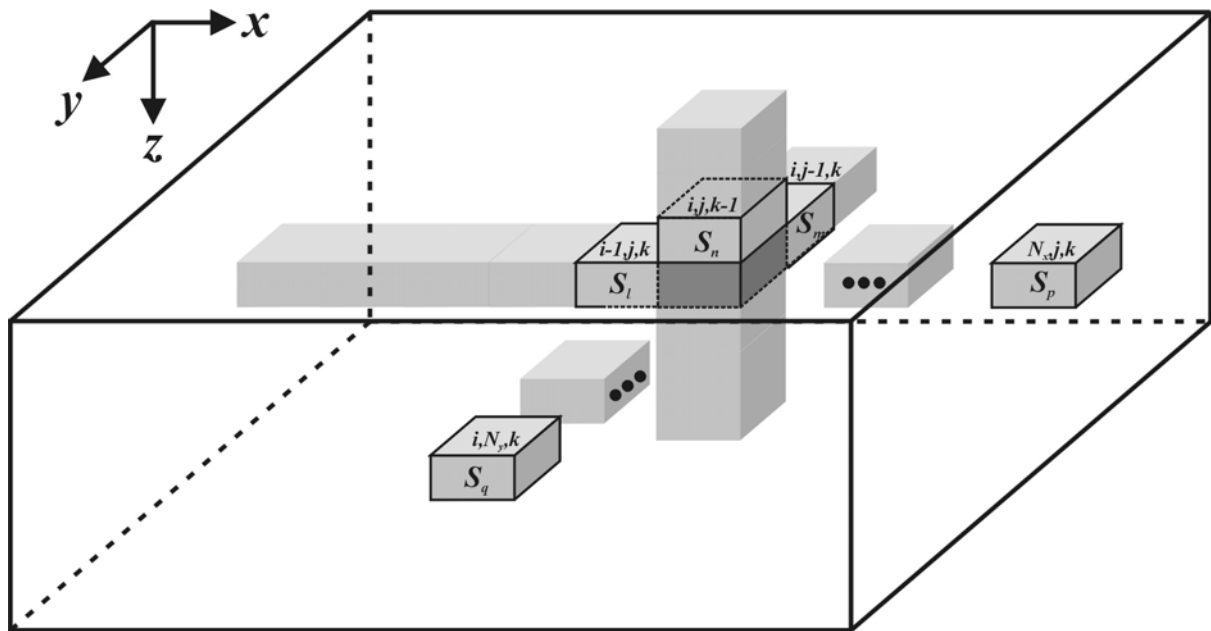
Figure 1. Computational cubic lattice of 3D CMC model.

In practical calculation of the 3D CMC model, all one-, two-, and three- dimensional CMC equations must be applied together (Park et al., 2002). The application of each CMC equation depends on the location in the three-dimensional lattice. For the periphery of the top layer, the one-dimensional CMC equation is used. For the remaining part of the top layer and the four side layers, the two-dimensional CMC equation is applied. For the rest of the domain, the three-dimensional CMC equation (Eq. (6)) is applied.

Algorithm of 3D CMC model

The developed algorithm for determining the lithology of each cell in a single realization using 3D CMC is as follows:

Step 1: The three dimensional domain is discretized using proper sampling intervals: for the vertical direction, the thinnest lithology observed will be used as a discretizing unit; for the vertical direction, Walther's law is adopted for the decision of the horizontal discretizing unit from the available information.

Step 2: The borehole data are saved in their proper locations for conditioning.

Step 3: The Markovian vertical transition probabilities are calculated from the tally matrix of the transition lithology using borehole data. Horizontal transition probabilities are inferred from the vertical transition probabilities using Walther's law. In most of the presented simulation, we use for the horizontal transition probabilities the averages of the upward- (i.e. from bottom to top) and the downward- (i.e. from bottom to top) transition probabilities to avoid direction biasness.

Step 4: The one- and two-dimensional CMC probabilities are applied on the periphery of the top layer and the periphery of the cube respectively. The lithology for each corresponding cell is filled in according to its corresponding conditional transition probability distribution.

Simulated

Original Image

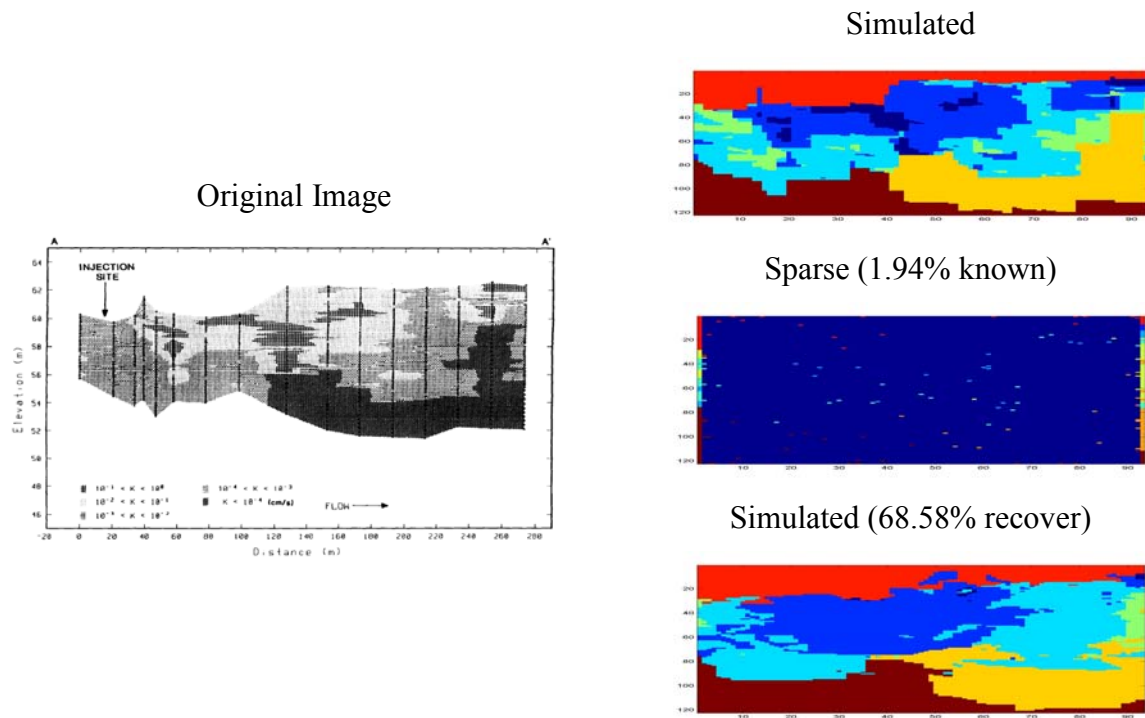Sparse (1.94% known)

Simulated (68.58% recover)

Figure 2. *LHS* – Hydraulic conductivity distribution along section A-A. Tick marks denote measurements of hydraulic conductivity using borehole flowmeter (from Adams and Gelhar, 1992), *RHS* – top, single realization by 2D CMC model, middle, reduced information down to 1.94% from simulated image, bottom, re-simulated image (conditioned on the 1.94% known data) recovering of 68.58% of the simulated image.

Step 5: The two- and three-dimensional CMC probabilities are applied on the top layer and inside the cube respectively. The lithology for each corresponding cell is filled in according to its corresponding conditional transition probability distribution.

Step 6: The procedure stops after having visited all the unassigned cells in the domain.

The Monte Carlo simulation is also possible by repeating the above-mentioned algorithm until desirable numbers of realizations are generated. After generation of enough single realizations, the final decision is made by the dominancy of the lithology in the corresponding cells. Calculation of the uncertainty is also possible through these multiple realizations.

4. Application of the 2D and 3D CMC Model

4.1 2D Application

As an illustration of the applicability of the CMC model, we consider an example based on data from the Alabama MADE Test Site in unconsolidated coastal plain sediments. Data from the site was available from 16 boreholes measurements along a transect (276m×12.2m – horiz.×vert.). The developed 2D CMC model is applied to the data (Figure 2, LHS). The input is prepared from the lithological data from the 16 boreholes located at the site. The data is discretized using the thickness of the finest lithology observed along the vertical direction, which is 0.1m. The transition probabilities along the vertical direction are calculated from the

16 boreholes using a descending sequence of lithologies in the boreholes. However, we use horizontal transition probabilities that are calculated by averaging the vertical ones in both descending and ascending sequences. This approach may take into account the various possibilities of the regional geologic setting in the horizontal direction at the site and avoid direction biasness [for details of this approach see Elfeki and Dekking (2003)].

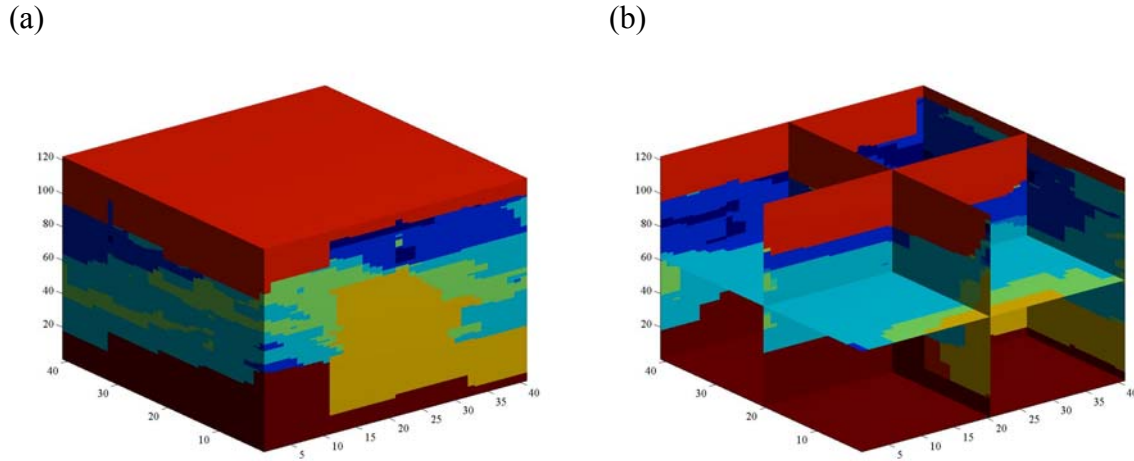(a)   (b)

Figure 3. Simulated single realization using the 3D CMC model: (a) outer slices; (b) inner slices.

Using our developed 2D CMC model, we generate a single realization image conditioned on 16 boreholes (Figure 2, RHS-top). Unaided visual comparison confirms the similarities between hand drawn geologic map (Figure 2, LHS) and the simulation result (Figure 2, RHS-top). We also performed a test using this simulated image by reducing the data down to 1.94% from the original image (Fig. 2, RHS-middle) as an extreme case, and re-simulate with this data (Figure 2, RHS-bottom) to see the effectiveness of the newly improved 2D CMC model. The single realization generated by this procedure reproduces 68.58% of the original image (Figure 2, RHS-top). From the structural point of view, the CMC model mimics very closely the structure of original one conditioned on the given sparse data.

4.2 3D Application

Based on the developed 3D CMC model, a single realization can be generated (Figure 3). In order to illustrate the 3D CMC model, we "bent" the previous transect data (at the MADE site) to form a cube. Therefore, the input data used for in the 3D CMC model is somewhat hypothetical where the boreholes are located on the periphery of the domain (Figure 3). On each side, there are four hypothetical boreholes where two boreholes are shared by neighboring sides. The domain size is 40m×40m×12.2m (hor.x×hor.y×vert.z) and discretized into 40 cells along the x- and y-directions, and 122 cells along the z-direction with uniform intervals.

The outside and the inside of the simulation results (single realization) by 3D CMC are shown in Figure 3(a) and 3(b), respectively. Each face of the block shows a geologically plausible 2D simulation profile as shown in the 2D CMC model. From the slice map (fence diagram) of the inside of the block, no unrealistic geologic changes are observed. Because the simulation is based on hypothetical data, our simulation results cannot be compared to real data. However, we can compare a number of realizations to see the stability of the model. In other words, if two realizations are totally different from each other, we can consider the model to be unstable. The instabilities can be, if any, predominantly observed from inside of the 3D block. Through our Monte Carlo analysis we confirmed that our model is stable and

the generated images have small variations from each other and also those variations happen around the lithologies boundaries (Figure 4).
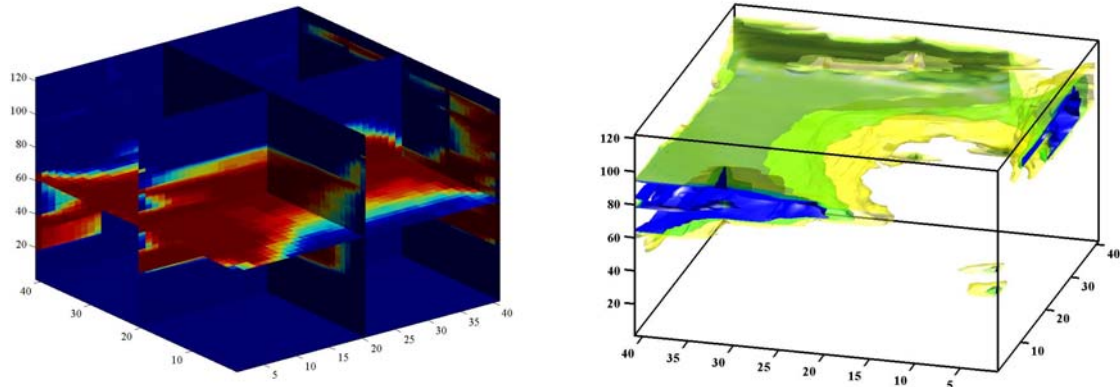

Figure 4. Statistical mapping of the EIF (Ensemble Indicator Function) in 3D space: (a) EIF of lithology 3 where red indicates high probability (~ 1.0) and blue indicates low probability (~ 0.0) of occurrence of the lithology; (b) superposition of the iso-EIF of encountering lithology 2 with a probability = 0.1 (yellow), 0.5 (green), and 0.9 (blue) in 3D space.
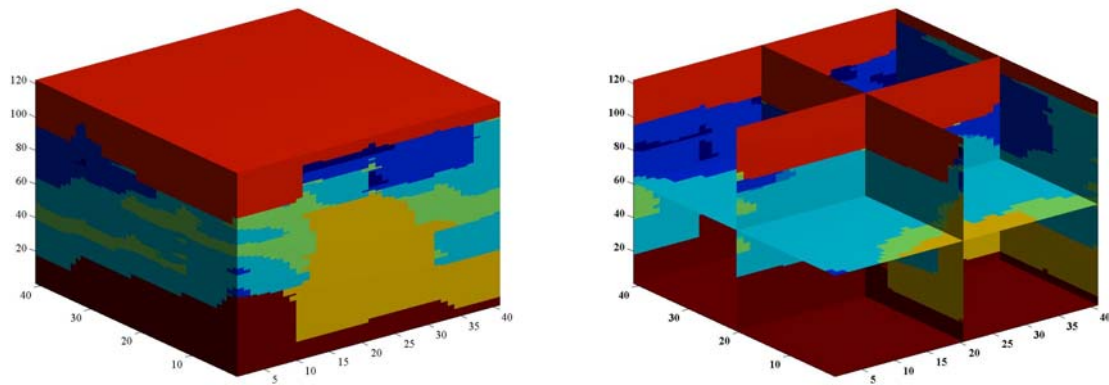

Figure 5. Final image generated from 30 realizations: (a) outer slices; (b) inner slices.

Through our Monte Carlo analysis, we generated a statistical map in 3D space of the ensemble indicator function (EIF) of each lithology. As a typical example, Figure 4(a) shows the statistical structure in 3D space of the EIF of lithology 3 and 2. In the figure, the EIF value of the dark red color is 1 and dark blue is 0. If an EIF is equal to 0, there is almost no probability of the appearance of the corresponding lithology, whereas an EIF is equal to 1 ensures that the corresponding lithology is present at that location in space. Figure 4(b) is an iso-EIF of lithology 3 for EIF values of 0.1, 0.5, and 0.9, respectively, which represent a 10% (yellow), 50% (green), and 90% (blue) degree of certainty of finding the corresponding lithology when the exploration reaches the inside of the EIF shell. The generated 3D statistical mapping can be used for engineering purposes.

Based on the EIF for each lithology, we extract an ensemble hypothetical geological map by assigning the lithology that has the maximum EIF at a given cell (Figure 5) [for details of this approach see Elfeki and Dekking (2003)]. There is no significant difference when we compare the final geological image to a single realization (Figure 3) except for some local variations at interfaces between lithologies that are smoothed out by the ensemble calculation.

This concludes that the 3D CMC model is a stable simulator honouring all the given conditioning data.

5. Summary and Conclusions

We developed a 3D CMC model and its corresponding software called CMC3D. This model is an extension the previous work of Elfeki and Dekking (2001). The software is available upon personal contact with the authors. A brief algorithm has been presented in this paper. We also provided improved schemes for the CMC2D conditioning. Monte Carlo simulations show that our model is stable for most cases. The 2D application on the MADE site shows that the model is promising in delineating the complex geological structure of the aquifer with sparse data (Figure 2).

Acknowledgement

References

Adams E.E. and Gelhar L.W., 1992, Field study of dispersion in a heterogeneous aquifer, 2 spatial moment analysis. Water Resour. Res., 28(12), 3293-3307.

Carle, S.F., and Fogg, G.E., 1996, Transition probability-based indicator geostatistics, Mathematical Geology, 28(4), 453-476.

Carle, S.F., and Fogg, G.E., 1997, Modeling spatial variability with one- and multidimensional continuous Markov chains, Mathematical Geology, 29(7), 891-917.

Carle, S.F., LaBolle, E.M., Weissmann G.S., VanBrocklin D., and Fogg, G.E., 1998, Conditional simulation of hydrofacies architecture: A transition probability/Markov approach, in Hydrogeologic Models of Sedimentary Aquifers, SEPM Concepts in Hydrol. Environ. Geol., vol. 1, edited by G. S. Fraser and J. M. Davis, pp. 147-170, Soc. For Sediment. Geol., Tulsa, OK.

Elfeki, A.M.M., 1996, Stochastic characterization of geological heterogeneity and its impact on groundwater contaminant transport. Ph.D.Thesis TU Delft, Balkema publishers, The Netherlands, 301p.

Elfeki, A.M.M., and Dekking, M., 2001, A Markov Chain Model for Subsurface Characterization: Theory and Applications, Mathematical Geology, 33(5), 569-589.

Elfeki, A.M.M and Dekking, F.M., 2003, Modeling Subsurface Heterogeneity by Coupled Markov Chains: Directional Dependency, Walther's law and Entropy. submitted to Water Resour. Res.

Krumbein, W.C., 1967, FORTRAN IV computer program for Markov chain experiments in geology: computer contribution 13, Kansas geological survey, Lawrence, Kansas, 38 p.

Park, E., Elfeki, A.M.M., and Dekking, F.M., 2002, Characterization of subsurface heterogeneity: Integration of soft and hard information using multi-dimensional Coupled Markov chain approach, Annual Progress Report, Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands, 30 p.

Ross, S., 2000, Introduction to probability models, 7th ed., 693 pp., Academic, San Diego, California.