Groundwater Quality: Remediation and Protection (Proceedings of the Prague Conference, May 1995). IAHS Publ. no. 225, 1995.

Stochastic simulation of heterogeneous geological formations using soft information, with an application to groundwater

AMRO M. M. ELFEKI, GERARD J. M. UFFINK & FRANS B. J. BARENDS

Delft University of Technology, Department of Civil Engineering, Geotechnical Laboratory, PO Box 5048, 2600 GA Delft, The Netherlands

Abstract A practical stochastic methodology for the generation of heterogeneous fields is presented. It uses soft data from geological surveys. The method is based on the extension of Markov chain theory in two dimensions through the use of two transition probability matrices. The results demonstrate that the methodology may simulate different types of realistic geological formations. In some selected formations groundwater flow is studied. The simulations yield realistic results. In the second half of the article, a Monte Carlo approach is followed to perform a series of numerical experiments on two types of structures: horizontal stratified formations and formations with inclined bedding. These experiments prove the following: (a) the probability density function of effective conductivity fits the normal or lognormal density function in case of small variability. In case of high variability the lognormal density fits better; (b) the effective conductivity is anisotropic due to the layered structure; (c) the tensorial property of the effective conductivity is established by the present model.

INTRODUCTION

In the last decade, modelling of groundwater flow and contaminant transport has been carried out within the stochastic framework, because of the natural heterogeneity of geological formations and the limited information about it. Heterogeneity is important for the prediction of the hydrological response and transport of contaminants. The predictions are indispensable for aquifer management, remediation strategies of contaminated aquifer systems, and for decision making. At present a variety of techniques is available to describe formation heterogeneity. Most of these are based on "hard data" such as measurements of hydraulic conductivity, porosity, dispersivity, etc. There is often insufficient hard information about heterogeneity of geological formations, but indirect qualitative geological "soft data" is usually available from geological surveys such as geological maps, well logs and borehole data. In general, acquisition of soft data is easier and less expensive. This paper presents a practical methodology for generating heterogeneous structure of geological formations using soft and hard data.

THE PROPOSED METHODOLOGY

Many geological processes display a Markovian property (it is the aspect, in which a subsequent event "remembers" a past event) (Harbaugh & Bonham-Carter, 1970). A one-dimensional Markov chain has been used by Krumbein (1967) for the simulation of geological sequences. The sequence of the strata is random but conditional to the preceding stratum. The presented methodology is an extension of Krumbein's method for two-dimensional situations. Geological formations often consist of stratifications deposited over thousands of years. According to that fact, anisotropy in the formation appears in the form of long extensions in the horizontal and small in the vertical direction. Two transition probability matrices suffice to describe this kind of anisotropy. The horizontal change between different geological materials in the formation is described by a horizontal transition probability matrix which is sampled over a proper sampling interval in the horizontal direction. Similarly, the vertical change between different geological materials is described by a vertical transition probability matrix.

Markov transition probability matrices

Transition probabilities may be defined as the relative frequency of a transition from a certain state to another state. These probabilities are expressed by p_{lk}^{d} , where p_{lk} is the probability of transition from state l to state k, and the superscript d indicates the transition direction (horizontal *1*, vertical *2*). Consider a rectangular two-dimensional domain of cells. Each cell has a row number i and a column number j. The geological system consists of a number of geological materials, say N, such as, sand, clay, peat, etc. The transition probabilities are in fact the conditional probabilities as known from the classical probability theory. A general form for the conditional probability in the upper horizontal boundary is $f(z_{i,1}|z_{i-1,1})$ where, $i = 2,3, \ldots, N_x$. Similarly, for the left vertical boundary we write $f(z_{1,j}|z_{1,j-1})$ where, $j = 2,3, \ldots, N_y$. p and f are related as follows:

$$\begin{aligned} f(z_{i,1} \mid z_{i-1,1}) &= p_{lk}^{(1)} \\ f(z_{1,j} \mid z_{1,j-1}) &= p_{mn}^{(2)} \end{aligned} \tag{1}$$

where, l is state $z_{i-1,1}$, and k is state $z_{i,1}$ in horizontal direction, and m is state $z_{1,j-1}$ and n is state $z_{i,1}$ in vertical direction. The state of the cells that are not on the boundary depends on the states of two neighbouring cells, one in the vertical and one in the horizontal direction. A general notation for the conditional probability is $f(z_{i,j}|z_{i-1,j},z_{i,j-1})$ where, $i = 2,3, \ldots, N_x$, and $j = 2,3, \ldots, N_y$. This term can be approximated as:

$$f(z_{i,j} \mid z_{i-1,j}, z_{i,j-1}) \approx f(z_{i,j} \mid z_{i-1,j}) \cdot f(z_{i,j} \mid z_{i,j-1}) = p_{lk}^{(1)} \cdot p_{mk}^{(2)} \tag{2}$$

In practice, the transitional probabilities of a geological system can be estimated either from well logs and surface maps or from a geological image synthesized by information from geologically similar areas. The vertical transitions are sampled at equidistant points with a suitable interval. The transition probability matrix is estimated from this observation as:

$$p_{lk}^{d} = \frac{m_{lk}^{d}}{\sum_{k=1}^{n} m_{lk}^{d}} \tag{3}$$

where, m_{lk}^{d} is the number of observed transitions from state l to state k in direction d.

The cumulative transition probability matrices are computed by adding each probability value to each succeeding value, moving from left to right within each row as

$$P_{lk}^{d} = \sum_{m=1}^{k} p_{lm}^{d} \tag{4}$$

The horizontal transition probability matrix can be estimated in a similar way along the horizontal direction from surface maps. The choice of the suitable sampling intervals is subjective. A proper sampling interval in the vertical direction is less than or equal to the minimum thickness of the geologic unit found in the well log. Similarly, the sampling interval in the horizontal direction should be less than or equal to the minimum length of a geological unit. The same methodology can be applied to handle inclined bedding of the formation structure. Two tests have been performed to validate this methodology: one uncorrelated field and one perfectly correlated field. The results show confidence.

GEOLOGICAL APPLICATIONS WITH GROUNDWATER FLOW SIMULATION

Some hypothetical geological formations with realistic characteristics are presented. One of the applications shows a comparison between the statistics used as input to the model

Table 1 Geological application, for Fig. 2.

Input data for geological model:

Length of the section (m) = 120

Depth of the section (m) = 60

Sampling interval Dx (m) = 1.0

Sampling interval Dy (m) = 0.50

Number of states = 4

Seed = 119 995

Input statistics				Calculated statistics			
Horizontal transition probability matrix							
0.900	0.030	0.030	0.040	0.924	0.017	0.047	0.013
0.010	0.970	0.010	0.010	0.022	0.960	0.016	0.002
0.010	0.020	0.960	0.010	0.024	0.010	0.963	0.003
0.040	0.040	0.010	0.910	0.093	0.053	0.099	0.755
Vertical transition probability matrix							
0.970	0.010	0.010	0.010	0.939	0.020	0.038	0.004
0.040	0.900	0.030	0.030	0.022	0.940	0.033	0.006
0.020	0.010	0.960	0.010	0.017	0.007	0.972	0.005
0.040	0.040	0.010	0.910	0.134	0.043	0.062	0.761

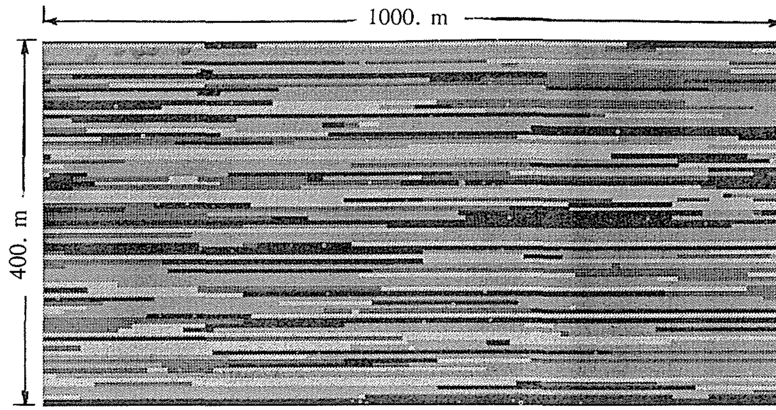

Fig. 1 Stratified formation of four lithological units with relatively small thickness.

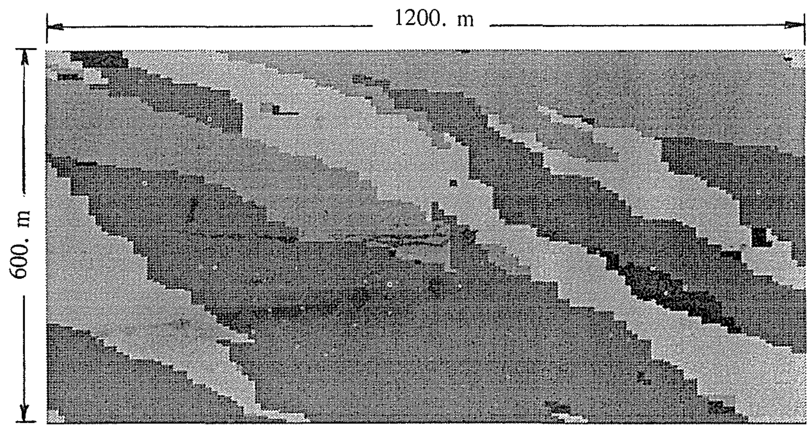

Fig. 2 Large scale geological formation of four lithological units.

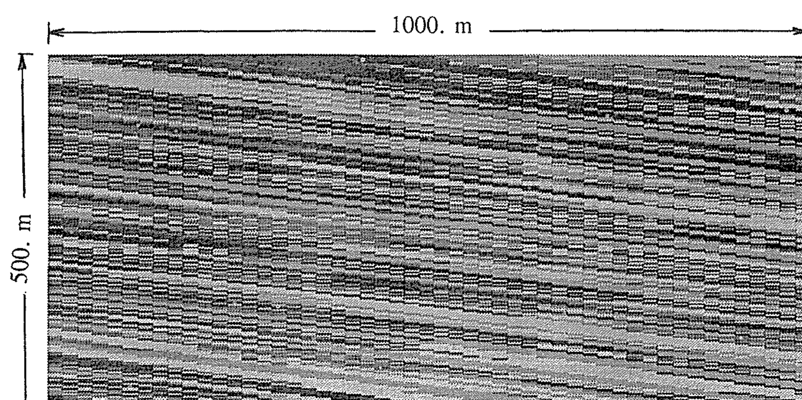

Fig. 3 Geological formation with inclined bedding at 7°.

and the statistics derived from the generated images in Table 1. Figures 1 to 3 display some examples of geological formations. Figure 1 displays a four-states-stratified geological system with small thicknesses representing a sedimentary basin. Figure 2 shows coarse formation structure. Figure 3 shows a fine geological system with a bedding of 7°.

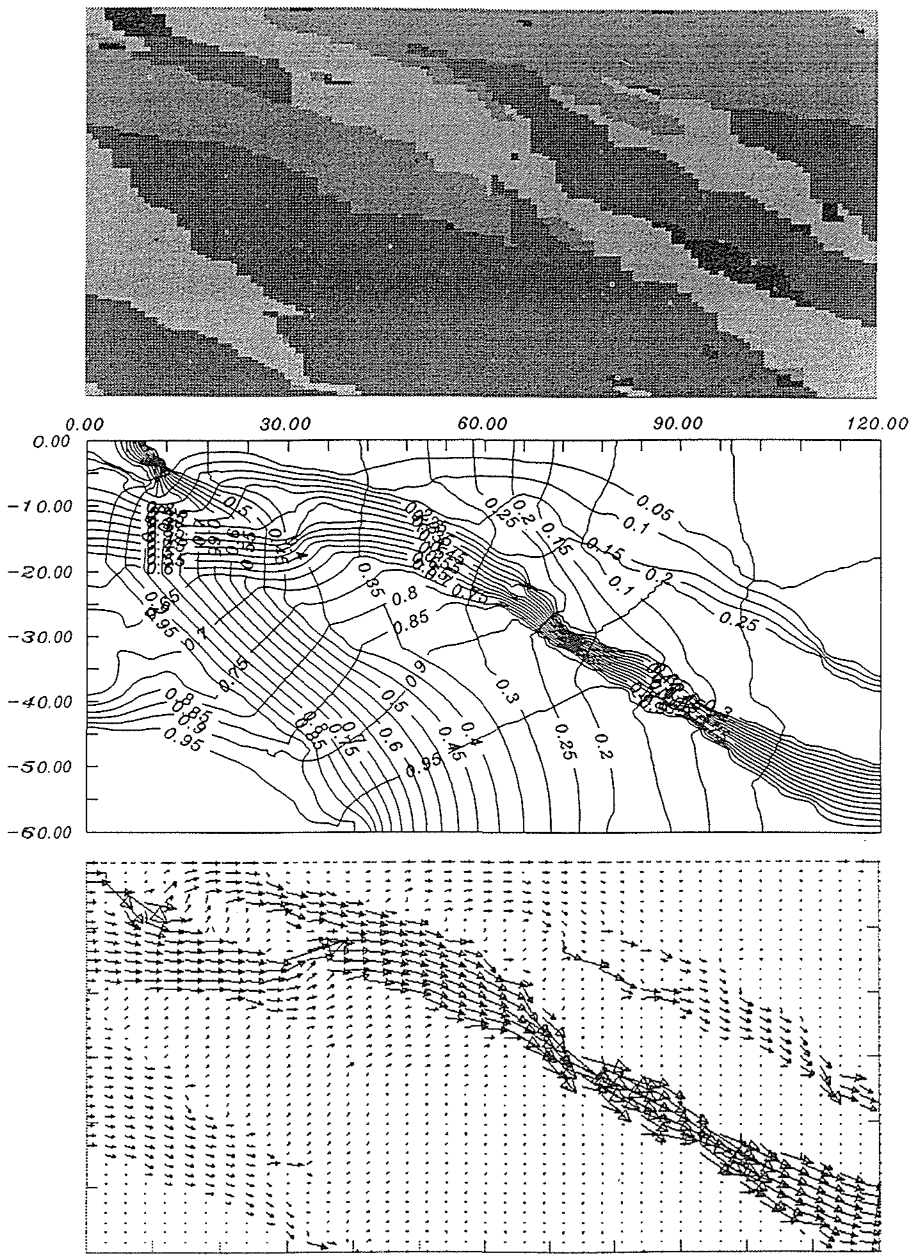

Fig. 4 Flow simulation in a single realization of large scale geological formation.

The geological structure in Fig. 2 has been used as an example for flow simulation. The corresponding statistics are given in Table 1. The resulting flow field is presented in Fig. 4. The procedure starts with superimposing a grid over the geological image. Each nodal point is assigned a hydraulic conductivity according to the corresponding geological material. Four types are considered. The conductivity values are 100, 10, 1 and 0.1 m day^{-1} which are distinguished by shades of grey in Fig. 4. Two types of boundary conditions are used, prescribed head and no-flow boundaries. Horizontal flow from left to right has been simulated by a constant head boundary on the left and right side, and a no-flow boundary at top of the flow domain.

The results can be verified by checking three conditions. First, the orthogonality of the flowlines and the equipotential lines. Second, the water particles must use the shortest path-lowest resistance, i.e. the flowlines should follow the regions of high conductivity values and take the shortest pass in low conductivity regions. Third, the mass conservation, which means that the quantity of seepage calculated at any vertical section must be the same. In the present model the quantity of seepage is calculated at two sections on the middle of the formation and at a section on the first third. The quantity of seepage is 4.524 and 4.552 m^2 day^{-1} respectively. The results of the present model fulfil the three conditions mentioned above.

NUMERICAL EXPERIMENTS

Numerical experiments may be used to replace laboratory experiments. They can easily be applied to estimate the effective conductivity of a heterogeneous sample. Two types of structures are investigated: horizontal stratifications and inclined bedding with 45° (see Fig. 5). The geological structure of both patterns possess the same statistical properties in terms of transition probabilities. A multi-realization approach has been adopted using the Monte Carlo method. Monte Carlo analysis relies on repetitive generation of replica of hydrogeological parameters and subsequent solution of the set of deterministic flow problems. The test cross-section is square in shape. A uniform grid is used with 10 000 cells. Four flow scenarios are applied: (1) horizontal flow in horizontal stratifications; (2) vertical flow in horizontal stratifications; (3) horizontal flow in inclined bedding with 45° with the mean flow direction; and (4) vertical flow in inclined bedding with 45° with the mean flow direction. The boundary conditions for the potentials are one on the left (on the top) and zero on the right (on the bottom) in case of horizontal (vertical) flow respectively. For 1600 numerical experiments the effective conductivity is determined. Two cases are investigated, one for relatively small variability (case 1) and one for a relatively large variability (case 2). The conductivities in case 1 (small variability) are 80, 60, 40 and 20 m day^{-1}. In case 2 (large variability) the conductivities are 100, 10, 1 and 0.01 m day^{-1}. The conductivities are distinguished by the same shades of grey. The effective conductivity $K_{xx_{eff}}$ is determined by:

$$K_{xx_{eff}} = -\frac{L_x}{L_y[\Phi_L - \Phi_R]} \sum_{j=1}^{N_y} \frac{(K_{xx_{k,j}} + K_{xx_{k+1,j}})}{2} \frac{[\Phi_{k+1,j} - \Phi_{k,j}]}{\Delta x} \Delta y \tag{5}$$

where, L_x and L_y are the domain dimensions in X and Y directions respectively, Φ_L and Φ_R are the left and right boundary heads respectively, and k is an indicator to define the section at which the horizontal seepage flow is computed. It could be any section

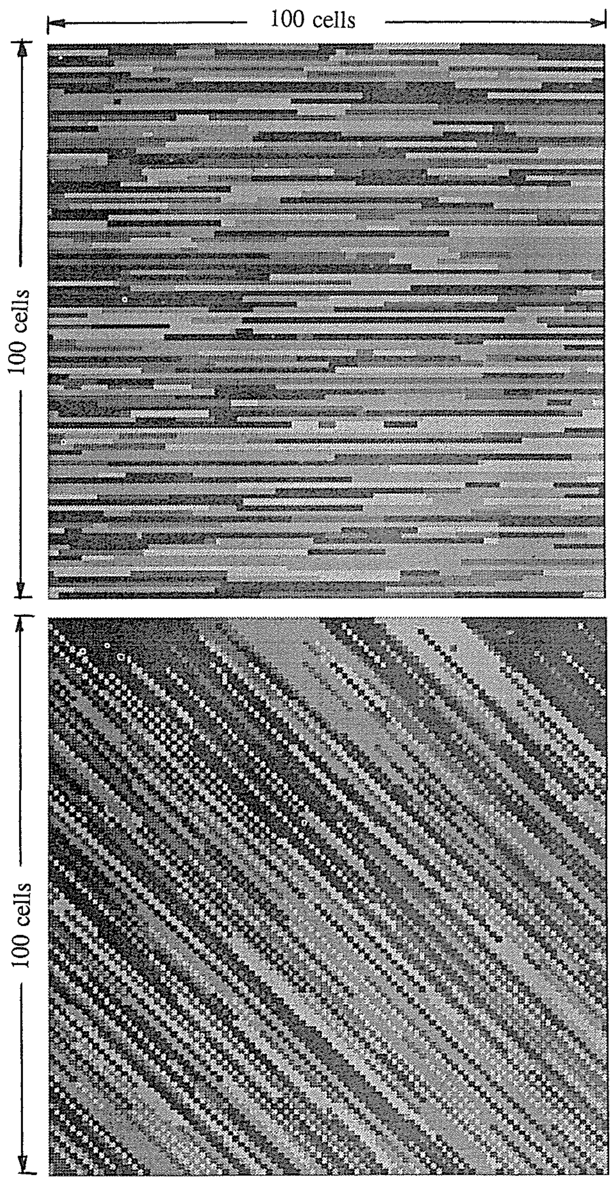

Fig. 5 A single realization used in numerical experiments (horizontal stratification, and inclined bedding with 45°).

between 1 and N_x. In the present study, a section at the middle of the aquifer has been taken. For vertical flow a similar formula has been used.

The probability density function of conductivity

The numerical experiments are used to obtain the probability density function of the effective conductivity. The results of case 1 (small variability) are displayed in Fig. 6.

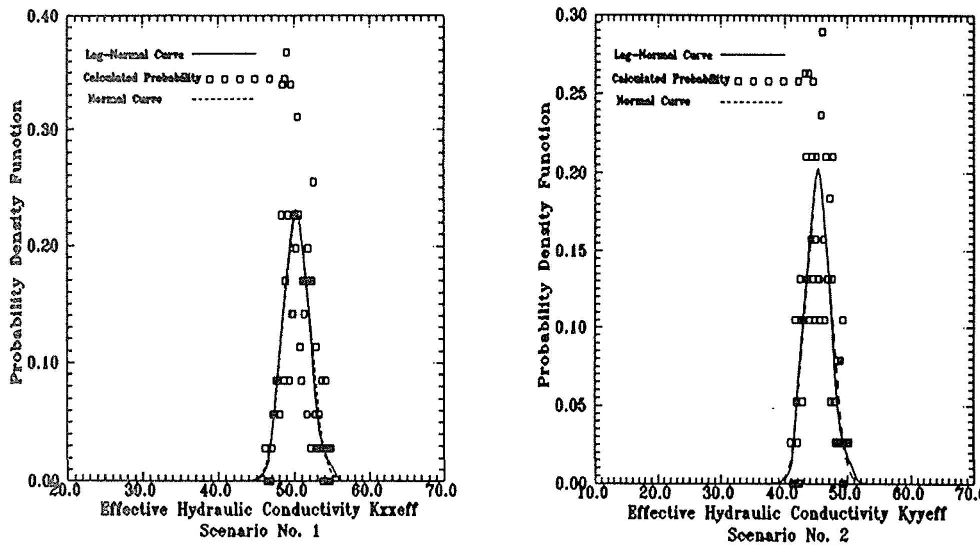

Fig. 6 Results of numerical experiments in Case 1 (scenarios 1 and 2).

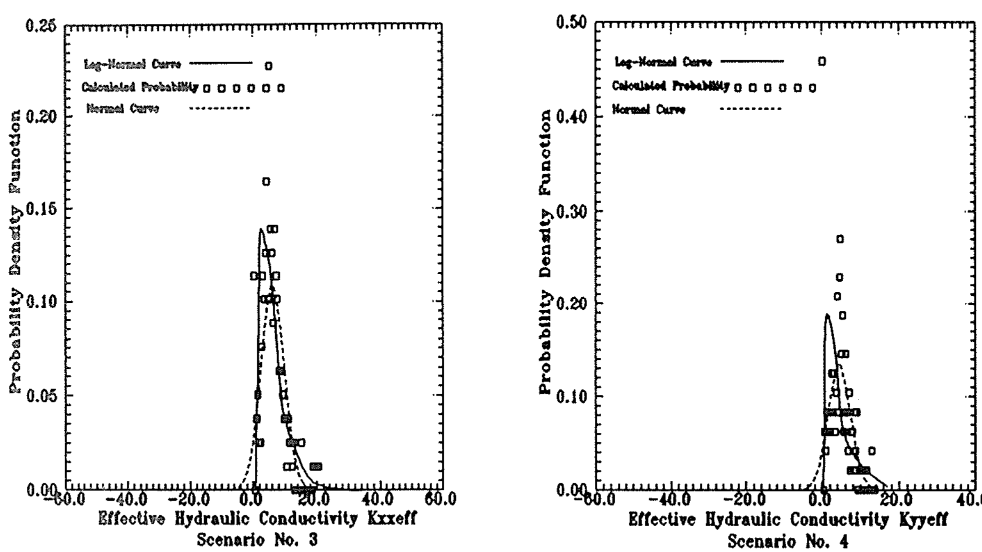

Fig. 7 Results of numerical experiments in Case 2 (scenarios 3 and 4).

It shows the calculated probability density function of the effective conductivity and the normal and lognormal densities. In Fig. 6, there is no discrepancy between fitting normal and lognormal density to the experimental points. However, the results of case 2 (large variability), presented in Fig. 7, show that the lognormality fits the experimental points better.

Tensorial property of the effective conductivity

The tensorial property of the effective conductivity is investigated. The anisotropy of the formation appears from the calculations, since obviously the ensemble mean effective horizontal conductivity $<K_{xxeff}>$ is not equal to its corresponding value $<K_{yyeff}>$ in vertical direction, as shown in Table 2. The anisotropy is stronger in case 2, because of

Table 2 Statistical ensemble measures of effective hydraulic conductivity.

Case study	No. 1 (small variability)		No. 2 (large variability)	
Scenario	$<K_{eff}>$ (m day^{-1})	σ_{Keff} (m day^{-1})	$<K_{eff}>$ (m day^{-1})	σ_{Keff} (m day^{-1})
1	50.33	1.74	20.89	2.86
2	45.33	1.98	1.63	0.94
3	47.55	3.35	6.11	3.67
4	47.33	3.39	4.47	2.95

the higher contrast in conductivity of various geological units. Comparison between ensemble statistics and spatial statistics of single realization shows that the value of $<K_{xxeff}>$ is close to the arithmetic mean K_a (50.33 and 49.70 respectively for case 1 and 20.89 and 26.80 respectively for case 2). This was expected because this case is close to flow parallel to a perfectly layered medium where the effective conductivity is the arithmetic mean. The value of $<K_{yyeff}>$ is close to the harmonic mean K_h (45.33 and 37.35 for case 1 respectively and 1.63 and 0.03 for case 2). This was expected as well, since this case is close to flow normal to a perfectly layered medium. From the present model, it was found that the effective conductivities in scenarios 1 and 2 transformed by a rotation of 45°, are in acceptable agreement with the effective conductivities from scenarios 3 and 4. The result is achieved using the following transformation:

$$\begin{Bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{Bmatrix} \begin{Bmatrix} <K^{(1)}_{xxeff}> & 0 \\ 0 & <K^{(2)}_{yyeff}> \end{Bmatrix} \begin{Bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{Bmatrix} = \begin{Bmatrix} <K^{(3)}_{xxeff}> & <K_{xyeff}> \\ <K_{xyeff}> & <K^{(4)}_{yyeff}> \end{Bmatrix} \quad (6)$$

where, θ is the rotational angle, $<K_{xxeff}{}^{(1)}>$, $<K_{yyeff}{}^{(2)}>$, $<K_{xxeff}{}^{(3)}>$, and $<K_{yyeff}{}^{(4)}>$ are the effective conductivity from scenarios 1, 2, 3, and 4 respectively, and $<K_{xyeff}>$ is the cross term.

Comparison between the results of the right hand side and left hand side of this transformation shows that in case 1 (small variability) the left hand side values (47.83 m day^{-1}) are within the central part of 95% confidence intervals (i.e. $<K_{eff}> \pm 2\sigma_{Keff}$). In this case we have $<K_{xxeff}{}^{(3)}> = 47.55 \pm 6.68$ m day^{-1} and $<K_{xxeff}{}^{(4)}> = 47.33 \pm 6.78$ m day^{-1}. However, in case 2 (large variability) the left hand side values (11.2 m day^{-1}) are close to the upper bound of the confidence intervals $<K_{xxeff}{}^{(3)}> 6.11 \pm 7.34$ m day^{-1} and $<K_{xxeff}{}^{(4)}> = 4.47 \pm 5.9$ m day^{-1}. The discrepancies between the values are due to the boundary conditions which force the flow to move in the main flow direction horizontally or vertically especially in highly heterogeneous mediums.

CONCLUSIONS

The following conclusions can be drawn from this study:

(1) The presented stochastic methodology is capable of simulating different patterns of natural geological formations, stratifications, layering systems or perfectly stratified systems, horizontal and inclined bedding with different angle of inclinations to the horizontal. The results are comparable to what is expected to be found in natural formations based on geological experience. All these patterns are controlled by the choice of the input parameters (transition probabilities).

(2) The technique calls for few input parameters (transition probabilities) which can be estimated from the geological surveys.

(3) The method reproduces pattern statistics. The given and simulated statistics show acceptable agreements with only minor fluctuations in some cases. The discrepancies are due to the insufficient number of transitions for one or more states in the geological system.

(4) The numerical simulation shows acceptable results of the flow field in terms of potentials, stream functions, and Darcy velocities in highly heterogeneous formations.

(5) The results of the numerical experiments proved that there is no distinction between fitting normal or lognormal density function of the effective conductivity in case of small variability. However, for large variability the lognormal density function is more likely.

(6) The hydraulic anisotropy of the formation structure appears from the results.

(7) The tensorial property of the effective conductivity could be proved from the present model.

The suggested methodology has the following advantages: (a) its theoretical background is simple, (b) it uses soft data beside the available hard data also, (c) its implementation is simple, (d) conditional simulation to local data is simple and straightforward, (e) it is efficient in terms of computer time and storage, (f) it can be extended to three dimensional problems. This approach will be used in future research to investigate the transport characteristics of heterogeneous deposits.

Acknowledgements This work is financially supported by the Committee Beek of the Delft University of Technology. The authors would like to acknowledge Dr Gerard Kruse for his constructive ideas.

REFERENCES

Bear, J. (1972) *Dynamics of Fluids in Porous Media.* American Elsevier, New York.

Harbaugh, J.W. & Bonham-Carter, G. (1970) *Computer Simulation in Geology.* J. Wiley, New York.

Krumbein, W.C. (1967) Fortran computer program for Markov Chain experiments in geology. *Computer Contribution 13, Kansas Geological Survey, Lawrence, Kansas.*