Probabilistic assessment of travel times in groundwater modeling

J. J. Gómez-Hernández and X.-H. Wen
Departamento de Ingeniería Hidráulica y Medio Ambiente, Universidad Politécnica, 46071 Valencia, Spain

Abstract: A Monte Carlo approach is described for the quantification of uncertainty on travel time estimates. A real (non synthetic) and exhaustive data set of natural genesis is used for reference. Using an approach based on binary indicators, constraint interval data are easily accommodated in the modeling process. It is shown how the incorporation of imprecise data can reduce drastically the uncertainty in the estimates. It is also shown that unrealistic results are obtained when a deterministic modeling is carried out using a kriging estimate of the transmissivity field. Problems related with using sequential indicator simulation for the generation of fields incorporating constraint interval data are discussed. The final results consists of 95% probability intervals of arrival times at selected control planes reflecting the original uncertainty on the transmissivity maps.

Key words: Indicator kriging, stochastic simulation, soft data, Walker Lake, sequential simulation, scaling-up

1 Introduction

Travel times estimates to be used as decision parameters in highly sensitive projects as site selection for nuclear waste repository should not be derived from a deterministic model built on limited information. Because petrophysical variables needed to model travel times present complex heterogeneity, the consequent uncertainty in predicting travel times must be assessed.

This paper discusses the modeling of uncertainty in travel time estimates and how to reduce such uncertainty by incorporating imprecise information in the modeling process. The paper also discusses the problem of using interpolated maps of transmissivity for transport predictions, and some other problems related to using sequential indicator simulation with constraint interval.

2 Statement of the problem

This paper presents a methodology to perform probabilistic assessment of groundwater advective travel times. The methodology is demonstrated using an exhaustive data set on which a prior exact answer is available.

Two sample data sets are available, one contains transmissivity measurements taken at a small scale. We will consider that this data set is free of measurement errors - we will refer to it as the "hard" data set. The other one comes from a dense 3-D geophysical survey which, once interpreted, provides upper and lower bounds for the transmissivity value - we will refer to this data set as the "soft" data set. These two data sets provide information on hydraulic conductivity values at a scale much smaller than the scale of the grid-blocks generally used for the numerical solution of the partial
differential equations describing groundwater flow and mass transport.

The scope of the study is to perform flow and transport predictions within a given area. We face two problems, first, the building of a numerical model for the spatial variability of hydraulic conductivity defined as averages over grid-block volumes, a volume much larger than that of the data used; second, the incorporation of soft data as obtained from geophysics in building such model.

The solution to the first problem calls for a scale-up algorithm. The second problem can be tackled using an indicator coding of soft information. The uncertainty in the spatial distribution of hydraulic conductivity is approached through stochastic simulations [e.g. Freeze, 1975; Delhomme, 1979]: several, alternative, equally likely, numerical models of hydraulic conductivity are built conditional to the information available, both hard and soft; each of these models is processed for travel time and the resulting ensemble solutions provide a probabilistic assessment of travel times uncertainty at various control plane locations.

3 The data set

The Walker Lake data set, variable U, described in Isaaks and Srivastava [1989, p. 542] is based on a digital model of elevation and measures the roughness of a topographic map. It has already been used as a surrogate for transmissivity by Desbarats and Srivastava [1991]. Because handling zero transmissivity values in numerical models for groundwater flow can create some problems, we have modified the original variable U by incrementing it in one unit. Spatial variability of the Walker Lake data set is due to natural processes not suspect of any artifact introduced by a synthetic generation process. We wish to stress this fact because most “reference” data, sets [e.g. Varljen and Sharer, 1991; Zimmerman and Gallegos, 1993] were generated using some kind of stochastic technique, which tends to bias favorably the results if the algorithm used builds on properties of that same stochastic model.

The data set contains 78,000 values arranged in a 2-D array with 300 rows and 260 columns with grid size considered as 1 m x 1 m. The histograms of both T and log T are shown in Figure 1, the exhaustive data set exhibits four orders of magnitude variation. About 10% of the data is equal to 1, which, for this analysis, is considered the minimum conductivity value. The mean log T is 1.66 and the variance of log T is 0.99 (variance of ln T of 5.2). The data set is not lognormal nor is it normal, which implies that any analysis calling for prior normal-related hypothesis is bound to be inaccurate. The use of log T in the display of the histograms and in the previous discussion is only for illustration purposes.

Notice that no specific unit was considered for transmissivity. In fact, no specific units will be used for any variable in this paper, except for distance. The aim of the study is to present a methodology and to compare results from different approaches.

To provide a better picture of the spatial variability of the data set, a gray scale map is shown in Figure 2. The shades were chosen such as to correspond to approximately the same proportion of data values, i.e., with thresholds equal to the data minimum, maximum and nine deciles. The heterogeneity of the data set is evident. Note the streak of high permeabilities that connects the Northwest border to the South border, and the curvilinear features in the Northeast quarter of the figure. As will be shown later, a proper characterization of these features is necessary to describe accurately the behavior of this aquifer.

The level of two-point spatial correlation of transmissivity can be measured by the variogram. The surface variogram of transmissivity, as shown in Figure 3, reveals a strong anisotropy with principal axis of major continuity at N14W. Cross-sections of this surface variogram along the two principal directions, i.e., traditional 1-D variogram curves, are displayed in Figure 4. The short scale variability represented by the discontinuity at the origin (nugget effect) represents 40% of the total variance. A zonal anisotropy up to about 50 m is also apparent. A zonal anisotropy indicates that more variability is present along any given strip in the N76E direction than along a similar strip in the N14W direction. The hole effect (decrease) of the variogram in the N76E direction at around 70 m can be explained by the pseudo periodic behavior observed in this direction; indeed, consider a cross-section of the transmissivity map of Figure 2 along the N76E direction, one can observe a sequence from left to right of intermediate-high-low-intermediate transmissivity values. Points which are further apart than 70 m in the N76E direction can take both similar transmissivity values hence their variogram would be smaller than the variogram for points at 40 m interdistance, where one point may take an intermediate value while the other would be either high or low.

Indicator variograms were also computed. The indicator transform for the variable T at threshold t and location x is given by