Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 217 (2006) 166–175
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Multivariate sensitivity analysis of saturated flow through simulated highly heterogeneous groundwater aquifers

C.L. Winter [a,*], A. Guadagnini [b], D. Nychka [a], D.M. Tartakovsky [c]

[a] *National Center for Atmospheric Research, Institute for Mathematics Applied to Geoscience, Boulder, 1850 W. Table Mesa Dr., CO 80305, United States*
[b] *Politecnico di Milano, DIIAR, Piazza L. Da Vinci, 32, 20133 Milano, Italy*
[c] *University of California, San Diego, La Jolla, CA, United States*

Received 2 August 2005; received in revised form 9 January 2006; accepted 19 January 2006
Available online 23 March 2006

Abstract

A multivariate Analysis of Variance (ANOVA) is used to measure the relative sensitivity of groundwater flow to two factors that indicate different dimensions of aquifer heterogeneity. An aquifer is modeled as the union of disjoint volumes, or blocks, composed of different materials with different hydraulic conductivities. The factors are correlation between the hydraulic conductivities of the different materials and the contrast between mean conductivities in the different materials. The precise values of aquifer properties are usually uncertain because they are only sparsely sampled, yet are highly heterogeneous. Hence, the spatial distribution of blocks and the distribution of materials in blocks are uncertain and are modeled as stochastic processes. The ANOVA is performed on a large sample of Monte Carlo simulations of a simple model flow system composed of two materials distributed within three disjoint blocks. Our key finding is that simulated flow is much more sensitive to the contrast between mean conductivities of the blocks than it is to the intensity of correlation, although both factors are statistically significant. The methodology of the experiment – ANOVA performed on Monte Carlo simulations of a multi-material flow system – constitutes the basis of additional studies of more complicated interactions between factors that define flow and transport in aquifers with uncertain properties.

Keywords: Sensitivity analysis; Parametric uncertainty; Groundwater flow; Stochastic partial differential equations

1. Introduction

Many geophysical applications deal with system parameters that are highly heterogeneous in space and are difficult to measure. This is certainly the case in hydrogeology where hydraulic conductivity, K, the principal system parameter in continuum representations of groundwater flow, varies greatly from one point of a groundwater aquifer to another. An n-dimensional aquifer, $\Gamma \subset \Re^n$, is a simply connected volume of geologic

* Corresponding author. Tel.: +1 303 497 1108.
E-mail address: lwinter@ucar.edu (C.L. Winter).

0021-9991/$ - see front matter © 2006 Published by Elsevier Inc.
doi:10.1016/j.jcp.2006.01.047

materials that is saturated with water. It has become common in models of groundwater flow to represent conductivity as a random field, $K = K(x,\omega)$, to quantify the uncertainty that arises from sparsely sampling a highly heterogeneous parameter. Here, ω is a member of Ω_Γ, the probability space defining the ensemble of random conductivity fields associated with an aquifer, and $x \in \Gamma$.

At macroscopic levels of description many aquifers can be viewed as block-structured, that is they are composed of a finite number of disjoint volumes, $\Gamma_i \subset \Gamma$, made from different materials. We call these volumes *blocks*, although they can have arbitrary shape. Upper Grand Canyon and Yucca Mountain, the proposed site of the US high-level nuclear waste repository, are typical examples. Each is composed of alternating layers of permeable and relatively impermeable materials, sandstone and limestone formations in the case of upper Grand Canyon, and nonwelded and welded tuffs at Yucca Mountain.

We base our analysis of block-structured systems on the random domain decomposition (RDD) model introduced by Winter and Tartakovsky [15,17]. In RDD an aquifer composed of n_B blocks is described by

$$\Gamma = \bigcup_{i=1}^{n_B} \Gamma_i + \sum_{i \neq j} \partial\Gamma_i \cap \partial\Gamma_j, \quad \Gamma_i \cap \Gamma_j = \emptyset \quad \text{if } i \neq j. \tag{1}$$

The block edges, $\partial\Gamma_i$, form boundaries, $\beta_{ij} = \partial\Gamma_i \cap \partial\Gamma_j$, between blocks and the set of boundaries, $\beta = \{\beta_{ij}\}$, defines an aquifer's block geometry. In RDD the points in a given Γ_i are all of the same material type, but their properties like conductivity vary from point to point, $K_i = K_i(x)$ for $x \in \Gamma_i$. Uncertainty about conductivity is modeled as a two-scaled stochastic process in RDD. The small-scale process K_i specifies conductivity at points within a given block, while the large-scale boundary process, $\beta = \beta(\omega)$, specifies the uncertain geometry of the blocks [17]. Although the geometry is usually uncertain, in this paper we keep it fixed in order to concentrate on other properties of block-structured aquifers. Recent advances in aquifer characterization techniques make it possible to estimate the number, n_B, and approximate the spatial extent of blocks in some aquifers [11,6]. Methods for measuring block boundaries are still in their development stage, and it seems likely that they will significantly improve in the coming years.

The small-scale conductivity process is defined by the finite-dimensional distributions $p_i(k(x_1), \ldots, k(x_\ell))$ for arbitrary points $x_1, \ldots, x_\ell \in \Gamma_i$, where $k(x) = K(x,\omega_0)$ is a specific value of K. The finite dimensional distributions can be assumed stationary in space, i.e. $p_i(k(x_1), \ldots, k(x_\ell)) = p_i(k(x_1 + \Delta), \ldots, k(x_\ell + \Delta))$ so long as the translated point $(x_1 + \Delta, \ldots, x_\ell + \Delta) \in \Gamma_i$, because each block is usually formed by a uniform geologic process during a single geologic period. Hence, the mean of conductivity can be assumed constant within a given block, i.e. $\overline{K_i}(x) = \overline{K}_i$, its variance, $\sigma^2_{K_i}$, is also constant and usually small, and the correlation $\rho_{K_i}(x,y) = \overline{K_i'(x,\omega)K_i'(y,\omega)} = \rho_{K_i}(\|x-y\|)$. The "overbar" indicates the operation of taking the ensemble average. We suppose $\rho_{K_i} = \rho_i$ has a finite correlation length, λ_i. These are standard assumptions of stochastic hydrology when describing flow through an aquifer formed of a single material [18,19].

When $\omega = \omega_0$ is fixed and the aquifer system can be described by (1),

$$K(x,\omega_0) = \begin{cases} K_1(x,\omega_0) & \text{if } x \in \Gamma_1, \\ \vdots & \\ K_n(x,\omega_0) & \text{if } x \in \Gamma_n. \end{cases} \tag{2}$$

Each $K_i(x,\omega_0) \in C^1(\Gamma_i)$. We use the Reynolds decomposition $K_i(x) = \overline{K_i} + K_i'(x)$ to represent random fields like $K_i(x)$ as the sum of a constant mean $\overline{K_i}$ and a zero-mean random fluctuation $K_i'(x)$. It is common in hydrogeology, to assume that the conductivity of a single geologic material is log-normally distributed, i.e. $\ln K_i(x,\omega) = Y_i(x,\omega) \sim N(\overline{Y}_i, \sigma^2_{Y_i})$. We use this assumption to generate synthetic realizations of K in a Monte Carlo framework, although RDD does not require it. From here on we will drop dependence on ω, but it will be implicit every time we perform a probability operation such as taking an average.

In saturated porous media, conductivity is related to macroscopic groundwater flux, $\vec{q}$, through a generalized form of Darcy's law,

$$\vec{q} = -K\nabla h. \tag{3}$$

When combined with conservation of mass, (3) yields the main state equation of groundwater flow,

$$S\frac{\partial h}{\partial t} = \nabla \cdot K\nabla h + f, \tag{4}$$

in which S is specific storage and $f(x,t)$ represents sources and sinks. The hydraulic head $h \in H^1(\Gamma)$, the Sobolev space of weakly differentiable functions on Γ. To simplify the discussion, we suppose (3) is a steady-state system without sources,

$$\nabla \cdot K\nabla h = 0. \tag{5}$$

Estimating the first few moments of h and $\vec{q}$ from estimates of the low-order statistics of K is a principal task of stochastic groundwater hydrology.

Various authors have adopted stochastic models in which natural porous media are represented as compositions of disjoint volumes, or blocks, of different materials whose individual properties vary uncertainly from one point to another. Suk and Lee [14] and Guler and Thyne [5] have used multivariate analysis based on chemical properties to divide aquifers into discrete zones. Conductivity is non-stationary, or heterogeneous, in space in such models. When conductivity is sampled in a multi-material aquifer without regard to the spatial distribution of materials, its probability density is multimodal [13,12,7,17]; and (Ritzi et al., 2004). Neuman [9] has demonstrated that the mathematical model of Di Federico and Neuman [3], which describes the conductivity field within the framework of a multiscale mode-superposition, is consistent with a representation of multimodal spatial variability in which space is filled by a discrete number of juxtaposed materials such as soil types, each having its own attributes. Unconditioned sampling without regard to the approximate locations of blocks generally leads to very high variances, σ_K^2, of conductivity [17], which in turn increases uncertainty about predictions of groundwater flow. In particular, $\sigma_K^2 > 1$ renders approximations of the system state questionable because most approximation techniques depend on small σ_K^2 in one way or another. Given statistics of the block locations and extents, the RDD model addresses the problem of small σ_K^2 by explicitly taking account of the (uncertain) spatial distribution of blocks [4,15–17]. This allows approximation of the moments of groundwater flow in strongly heterogeneous.

This paper reports the first in a series of studies that will use multivariate statistics to investigate the relative sensitivity of groundwater flow to the effects of different aquifer properties in block-structured aquifers. A number of factors can affect the state of flow in block-structured aquifer systems. These include the geometry of the blocks, stochastic variation of conductivity within each block including its mean, variance, correlation structure and correlation length, the roughness of the boundaries, and the degree of correlation between conductivities across materials and its length scale. Here we compare the effect of the degree of cross-correlation between materials to the effect of differences between the means of conductivities in different materials. Hence we hold all other factors fixed, including β as noted earlier.

According to Neuman [9] cross-correlations between material attributes can arise because the classification of soils and other natural porous media on the basis of percent sand, silt, and clay, is often subjective and therefore somewhat arbitrary. Data used in facies classification, including grain size distribution curves, are typically derived from core samples and are often poorly differentiated, which further prevents a clear distinction between different hydrofacies. Correlations between blocks may also occur in systems where lenses of one material are laid down simultaneously with a base material for a while so that uncertain boundaries result [10]. Winter and Tartakovsky [17] derive approximate expressions that relate flow statistics, including cross-correlations between block properties, to that of hydraulic conductivity, but do not explore their effects. Lu and Zhang [7] discuss correlation between materials and then provide examples that deal with the prediction of flow in media characterized by the lack of cross-correlation between conductivities in different materials. Uncertainty about the spatial distribution of materials can yield apparent cross-correlations even if two materials are themselves independent [17]. The effect of such cross-correlation has been an outstanding issue in groundwater hydrology and to our knowledge it has not been statistically compared to the effect of variations in mean conductivity.

Our goals in this paper are methodological as well as analytical. We evaluate the ability of a classical statistical technique, the Analysis of Variance (ANOVA), to measure sensitivity of groundwater flow to the cross-correlation and mean conductivity factors. Although ANOVA is a classical approach to evaluating sensitivity, our setting is not. Instead of a laboratory sample of physical subjects that have been treated

systematically to reflect the influence of different levels of factors, we examine a collection of simulated aquifers drawn by Monte Carlo sampling from an ensemble designed to reflect the differential effects on flow of systematically defined aquifer properties like correlation and type. While it is expensive to construct even a small number of experimental aquifers, generating a large number of realizations is relatively cheap. Furthermore the probabilistic properties of such an ensemble can be known exactly and constructed specifically to meet the assumptions of ANOVA.

Before getting to the ANOVA-based sensitivity analysis, we provide additional background and definitions for the probabilistic flow model (Section 2). We evaluate the sensitivity of flow through model aquifers to the degree of correlation on one hand and aquifer type on the other (Section 3).

2. Problem setting

The ensemble average of (5) gives the mean pressure head equation

$$\nabla \cdot (\overline{K}\nabla\bar{h}) - \nabla \cdot \bar{r} = 0 \tag{6}$$

which includes a second-order statistic, the mean "residual" flux $\bar{r}(x) = \overline{-K'\nabla h'}$ representing cross-correlation between hydraulic conductivity and head gradient fluctuations. This system is typically closed by perturbation approximations that depend on small conductivity variances [19].

2.1. Perturbation approximations

In block-structured systems the first-order approximation of $\bar{h}(x)$ depends on

$$\hat{r}_j^{(1)}(x) = \kappa_j(x)\sum_{i=1}^{2}\int_{\Gamma_i}\kappa_i(y)\rho_{ij}(x,y)\nabla_y\nabla_x^T\widehat{G}^{(0)}(y,x)\nabla_y\hat{h}_i^{(0)}(y)\mathrm{d}y, \tag{7}$$

which is a first-order approximation of $\bar{r}_j(x)$ conditioned on fixed β [17]. The "hat" indicates a mean taken with respect to the conditional density, $p(K|\beta)$, of K given β, i.e. $\bar{r}_j^{(1)}(x) = \int \hat{r}_j^{(1)}(x)p(\beta)\mathrm{d}\beta$. Note that the approximation (7) is equally valid for statistically inhomogeneous blocks. The effect of cross-correlation appears in $\rho_{ij}(x,y) = \overline{Y_i'(x)Y_j'(y)}$, which depends on $Y_i'(x) = \ln K_i'(x)$. The term $\kappa_i = \exp(\overline{Y}_i)$ is the geometric mean of K_i, $\widehat{G}^{(0)}$ is the zeroth-order approximation of the conditional mean Green's function, and $\hat{h}_i^{(0)}$ is the zeroth-order approximation of the conditional mean of h. Clearly $\hat{r}_j^{(1)}(x)$, and hence $\bar{h}(x)$, depends in a complicated way on factors such as κ_i, $\rho_{ij}(x,y)$, and β. Multivariate statistics, in particular ANOVA, provide a systematic way to investigate their relative effects, and in this study we concentrate on the relative effects of κ_i and $\rho_{ij}(x,y)$.

2.2. Cross-correlation and model flow geometry

Cross-correlation does not affect flow in some important cases. Consider a stratified medium in which the direction of seepage flow is perpendicular to the stratification. If the lateral boundaries are impermeable, constant head $h = 0$ is maintained at the lower boundary, and water enters the system at a constant rate $K_1\mathrm{d}h/\mathrm{d}y = -q$ through the upper boundary, then mean flow becomes one-dimensional. In this case direct integration of the flow equation and subsequent ensemble averaging leads to an *exact* solution for the conditional mean hydraulic head

$$\hat{h}(x) = -q\begin{cases}\frac{x}{K_{h_1}} & 0 < x < \beta_{\mathrm{L}},\\ \frac{\beta_{\mathrm{L}}}{K_{h_1}} + \frac{x-\beta_{\mathrm{L}}}{K_{h_2}} & \beta_{\mathrm{L}} < x < \beta_{\mathrm{U}},\\ \frac{\beta_{\mathrm{L}}}{K_{h_1}} + \frac{\beta_{\mathrm{U}}-\beta_{\mathrm{L}}}{K_{h_2}} + \frac{x-\beta_{\mathrm{U}}}{K_{h_1}} & \beta_{\mathrm{U}} < x < 1.\end{cases} \tag{8}$$

Here K_{h_i} denotes the harmonic mean of the hydraulic conductivity of the ith block. It is obvious from (11) that $\hat{h}(x)$ does not depend on $\rho_{ij}(x,y)$, so $\bar{h}(x)$ is not influenced by cross-correlations between the materials in such a flow system. Since the effect of cross-correlation on flow is our focus, we base the ANOVA on practically the opposite flow geometry to this case (Fig. 1).

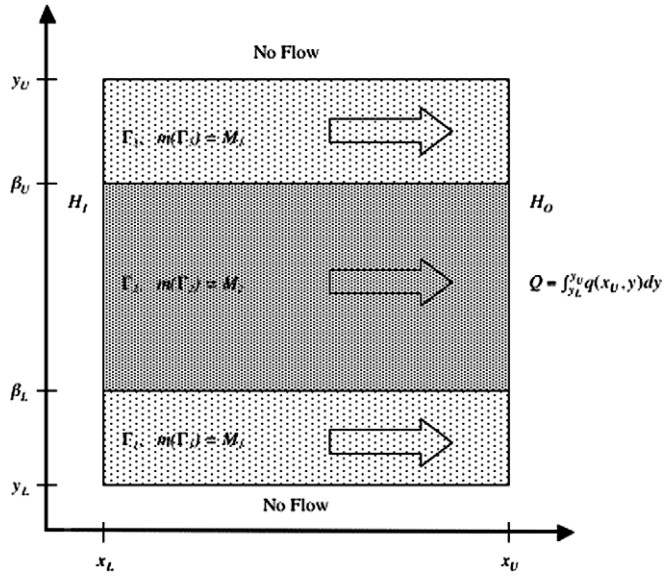

Fig. 1. Basic flow system. The internal boundaries, β_i, are fixed. The arrows indicate the mean flow direction.

2.3. Model flow system

The model flow system is block-structured with fixed (i.e. deterministic) internal boundaries and external boundary conditions (Fig. 1). This system is a simplification of composite aquifers that consist of two subaquifers separated by a less permeable aquitard. The model domain is a square with sides of length L aligned along the coordinate axes. Two kinds of materials, M_1 and M_2, are separated by internal boundaries into three blocks, hence the number of materials is $n_M = 2$ and the number of blocks is $n_B = 3$. The boundaries between them are $\beta_L = 0.29L$ and $\beta_U = 0.71L$. The material forming the inner layer, M_2, is generally less permeable than M_1, the material of the two outer layers. The statistics in our Monte Carlo simulations are consistent with systems composed of two sand layers separated by a layer of clay. The thickness of the aquitard, the middle layer, is $0.42L$. No flow is allowed across the top and bottom boundaries at y_U or y_L. Hydraulic head is held constant across the inlet at x_I, $h(x_I, y) = H_I$, and is held to $h(x_O, y) = H_O$ across the outlet at $x_O > x_I$. Additionally, $H_I > H_O$, so the basic direction of flow is from left to right.

These simplifications do not much alter the basic problem of estimating sensitivity of the flow system to different sources of variability in $K(x)$. Random external boundary conditions (and random sources and random initial conditions when they apply) enter (4) and (5) as additive sources of noise that can be dealt with relatively easily when compared to the multiplicative effects of K. We further simplify the model system by representing K as a scalar in the simulations, so the conductivity field is locally isotropic. This is a fairly common assumption in hydrogeology. The grid discretization is set to 0.2λ in order to have at least five grid elements per correlation scale within each material.

We base our simulations on Darcy's original method for determining hydraulic conductivity. $K(x)$ is a macroscopic parameter that reflects variability in the network of microscopic pores that actually make up an aquifer. To obtain K, Darcy forced water through a sample of natural porous medium obtained from an aquifer and measured the flux of water, $\vec{q}$, and the hydraulic pressure gradient, ∇h. In the simplest experiments a constant hydraulic pressure difference, Δh, is maintained between the inlet and the outlet of a core of unit cross-sectional area and of length L, and a uniform flux of water, Q, is known at the core's outlet, so $K = -QL/\Delta h$. Since the volume of the core is much smaller than the extent of the aquifer, K obtained in this way is essentially a point measurement. We use total volumetric flow rate across the outlet,

$$Q = \int_{y_L}^{y_U} q(x_O, y)\mathrm{d}y, \tag{9}$$

to measure system response to differences between material means and to the degree of correlation between the materials in the ANOVA.

3. Sensitivity analysis

We combine ANOVA with a Monte Carlo experiment to assess the sensitivity of the model flow system to the degree of correlation between conductivities of different materials and the relative difference between average conductivities of the materials. Given a sample of aquifer realizations based on the model system, the ANOVA evaluates relative sensitivity to the two factors by decomposing the overall variability of total outflow, Q, into components that are due to each of the factors, to their interaction, and to residual error not attributable to the factors. The result is a statistical assessment of the significance of the factors.

3.1. Monte Carlo realizations

Correlated conductivity fields are generated with the method of circulant embedding [1]. Given a stationary covariance kernel, circulant embedding produces a set of correlated log conductivity fields modeled as Gaussian random fields, a common assumption in groundwater hydrology [19] and one that is technically required for significance testing in ANOVA. In our simulations the covariance kernel is $k(r) = e^{-r/\lambda}$ with r denoting the distance between two points and λ the correlation length. In general, the cross-covariance of $Y_\ell(x) = \ln K_\ell(x)$ between materials is

$$C_{ij}(x,y) = \overline{Y_i'(x)Y_j'(y)} = L_\rho \int_0^\infty \gamma_i(x-\zeta)\gamma_j(y-\zeta)\mathrm{d}\zeta, \tag{10}$$

where γ_i is the square root kernel of K_i. The parameter L_ρ determines the maximum strength of the correlation and is one of the factors in the ANOVA.

3.2. Aquifer types

The parameters determining Q in the model system are listed in Table 1. We make several simplifying assumptions about these parameters that are compatible with our emphasis on comparisons between cross-correlation and mean conductivity contrasts. Roughness of the internal boundaries is set to zero, i.e. the internal boundaries are flat and parallel to the x coordinate. We also set $\sigma^2_{Y_1} = \sigma^2_{Y_2} = 1$, $\lambda_1 = \lambda_2 = 0.135 \times \beta_1$, and use the same exponential covariance function $e^{-\lambda r}$ for log-conductivity in the two materials. With these assumptions, the functional dependence of the normalized flow rate, Q/κ_1, on the aquifer parameters can be reduced to the dimensionless form,

$$\frac{Q}{\kappa_1} = f\left(\frac{\kappa_2}{\kappa_1}, L_\rho\right). \tag{11}$$

The only dimensionless parameters which act as sources of variation within our model system are 1) the level of correlation, L_ρ, between the logarithms of hydraulic conductivities of each material and 2) the relative contrast, κ_2/κ_1, between the geometric mean conductivities of the two materials. The L_ρ parameter determines the maximum correlation allowed between the two materials; we use five levels of $L_\rho = 0$, 0.25, 0.5, 0.75, 1 to de-

Table 1
System parameters

Parameter	Description
κ_i	The geometric mean of conductivity within the ith material
$\sigma^2_{Y_i}$	The dimensionless variance of $Y_i = \ln K_i$, the logarithm of hydraulic conductivity in the ith material
λ_i	The correlation scale of Y_i
β_j	The location of internal boundary β_j ($j = 1,2$)
r	The roughness of β_j
L_ρ	The maximum level of correlation between the conductivities of the two materials, which is dimensionless

fine the correlation factor in the ANOVA. The ratios $\kappa_2/\kappa_1 = 1$, 0.018, 0.007 describe average conductivity contrasts of zero, one and two orders of magnitude, in agreement with frequently observed properties of natural porous materials (cf. [2]). The simulated values are: (1) $\kappa_1 = \kappa_2 = 1.16 \times 10^{-5}$ [cm/s] corresponding to a relatively impermeable material like silty-sand; (2) $\kappa_1 = 6.32 \times 10^{-4}$ [cm/s] representing a sandy material and $\kappa_2 = 1.16 \times 10^{-5}$ [cm/s], a silty-sand; (3) another simulated sand/silty-sand system, but with a more conductive sand, $\kappa_1 = 1.72 \times 10^{-3}$ [cm/s] and the same conductivity of the silty-sand $\kappa_2 = 1.16 \times 10^{-5}$ [cm/s].

3.3. ANOVA

The combination of three mean conductivity ratios and five levels of correlation yields an experimental design with 15 subpopulations in the overall space of cross-correlated aquifers. We created 1000 realizations of each sub-population for our Monte Carlo simulations, resulting in a total sample size of 15,000 aquifer realizations. We calculated the total dimensionless flux at the outlet Q_{ijk} ($i = 1, \ldots, 3$; $j = 1, \ldots, 5$; $k = 1, \ldots, 1000$) for each realization and obtained the grand mean,

$$\widetilde{Q} = \frac{1}{15000} \sum_{i=1}^{3} \sum_{j=1}^{5} \sum_{k=1}^{1000} Q_{ijk}, \tag{12}$$

and the total variability of the sample,

$$S = \sum_{i=1}^{3} \sum_{j=1}^{5} \sum_{k=1}^{1000} \left(Q_{ijk} - \widetilde{Q} \right)^2. \tag{13}$$

The ANOVA decomposes S into components arising from the conductivity type, S_κ, from the degree of correlation between the materials, S_L, and from their interaction, $S_{\kappa L}$,

$$S = S_\kappa + S_L + S_{\kappa L} + S_\varepsilon. \tag{14}$$

The residual error, S_ε, contains the part of the variability in the sample of Q_{ijk} that is not accounted for by the two factors. The terms on the right of (14) have the same form as S, that is, they are sums of squared differences between various means of the Q_{ijk}. For instance,

$$S_\kappa \equiv \sum_{i=1}^{3} \left(\widetilde{Q}_i - \widetilde{Q} \right)^2 \tag{15}$$

depends on means that summarize the aquifer type factor, $\widetilde{Q}_i = \frac{1}{3000} \sum_{j=1}^{3} \sum_{k=1}^{1000} Q_{ijk}$ and the total sample mean, $\widetilde{Q}$. Similar expressions define S_L and $S_{\kappa L}$.

3.4. Sensitivity by ANOVA

The issue is whether dimensionless total flow responds significantly to variations in either correlation level, or conductivity contrast, or both. Fig. 2 reveals large differences among the different aquifer subgroups. The box-plots are organized by the three conductivity contrast groups and within these are plotted the individual responses to the five different levels of correlations. The boxes have limits at the 25 and 75 percentiles of the data with a line segment within the box locating the median. The vertical lines extending from the boxes indicate a range for the distribution and extreme data values outside this range are plotted separately. In this figure there is substantial variability due to the contrast in conductivity. Visually the first group has a different mean level and there is also different variability among the groups. The two sand/silty-sand aquifers (groups 2 and 3) perform about the same.

However, within a group the response to correlation is subtle. Fig. 3 isolates the effect of the correlation factor by first subtracting a mean response, $\widetilde{Q}_i$ due to the different contrast in conductivity groups. The variation in the mean due to different correlation levels and different cases is apparent once the large differences among conductivity contrast subgroups are removed. These adjusted means for each group will add to zero and so the interest is the deviation from zero as a function of the strength of the correlation. The broken lines

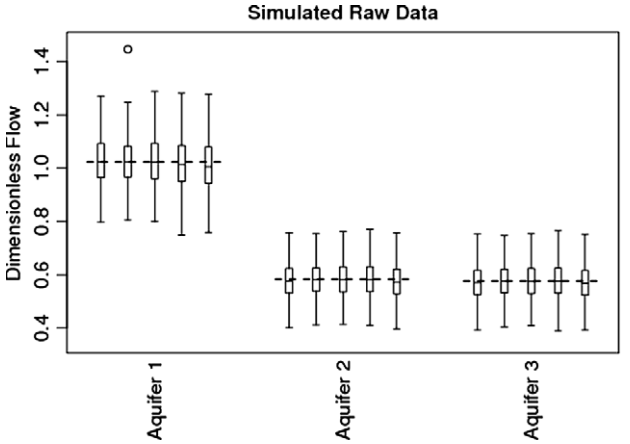

Fig. 2. Samples means $\widehat{Q}_{ij}$ and ranges classified by the factors of flow regime, A, and cross-material correlation, ρ. The basic grouping is by the three subgroups of A. Data within a subgroup is further subdivided by correlation level.

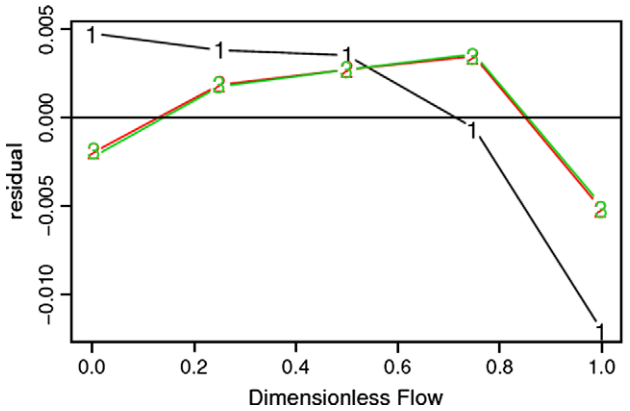

Fig. 3. The effect of the correlation factor.

connect the adjusted means for the same group and so trace out the pattern of how the response may change as a function of correlation for a given group. Although there is no obvious pattern, the response to different correlation levels does appear to deviate from zero.

The ANOVA results (Table 2) are a formal test for whether the deviations in these values can be explained purely by chance. If that hypothesis can be rejected, the natural inference is that the variation is due to a changing mean response due to the correlation levels. To insure that the assumptions for the ANOVA are satisfied, the data in each conductivity group have been standardized to have a standard deviation of one. The application of the model of (14) to our sample partitions the variability among the 15,000 values of Q_{ijk} as shown in Table 2.

Table 2
Analysis of variance

Source of variability	n_l	S_l	S_l/n_l	F Value
Aquifer type (κ)	2	657.635	328.818	53484.687
Correlation (L)	4	0.222	0.056	9.028
Interaction (κL)	8	0.078	0.010	1.586
Residuals (ε)	14,985	92.126	0.006	

The ANOVA consists of comparing the variability arising from residual error to the variability due to a given factor. A factor is insignificant if it does not account for more variability than the residual does. This is tested by the F statistic,

$$F_\ell = \frac{S_\ell/n_\ell}{S_\varepsilon/n_\varepsilon}, \tag{16}$$

which compares the variability of the ℓth factor ($\ell = \kappa, L, \kappa L$) to the variability that is left unaccounted by any factor, essentially S_ε. Note that the sums in (16) are over different numbers of elements and so must be scaled. The scaling constant is the factor's degrees of freedom, $n_\kappa = 3 - 1 = 2$, $n_L = 5 - 1 = 4$, $n_{\kappa L} = (3 - 1)(5 - 1) = 8$, and $n_\varepsilon = 3 \times 5 \times (1000 - 1) = 14{,}985$. The term S_ℓ/n_ℓ is called the mean square of the ℓth factor.

The first line of Table 2 indicates that the means describing the conductivity contrast groups are different, which comes as no surprise after our discussion of Fig. 2. The second line indicates that the average response across the different correlation levels is also statistically significant ($p < 0.001$). Note, however, the large difference between the mean squares of the correlation factor and of the conductivity contrast factor. This further quantifies the subjective impression given by Fig. 2. The third line reflects variation in the mean for each of the 15 combinations after accounting for the average for each conductivity contrast group and each correlation level. The overall interaction is not significant ($p = 0.122$), but a more specific test of the individual interactions suggests a heterogenous response to correlation across the conductivity groups. In particular, the groups appear to responding differently when the correlation factor is 1.0.

Finally, we note that the ANOVA depends on approximate Gaussian distributions for the responses to the extent that the F statistics computed in the ANOVA table follow an F distribution. In this case the assumptions are justified based on histograms of the observed responses and also the large number of Monte Carlo replicates (1000).

4. Summary

Flow in the model system is sensitive at a statistically significant level to the contrast between material means and to the maximum intensity of cross-correlation between materials. However the effect of correlation, while statistically significant, is far weaker than the contrast between material means. This accords with intuition and also agrees with concepts arising from issues related to connectivity (e.g. [8]), since the relative proportion of high conductivity material is the main factor controlling flow, the fact that is better reflected in the means than in the correlation structure of the field.

The exact response to the correlation factor appears to be complicated and could not be resolved completely based on the number of Monte Carlo samples used in this study. However, the cases when the correlation factor is 1.0 (perfectly correlated) appear to respond differently than other levels of correlation. The method was not too computationally expensive, although part of that was due to the simple geometry of the model flow system.

We intend to expand this work in several directions. For one thing intensity is not the only feature of cross-correlation that affects flow. The shape of the correlation function and especially the correlation length may have greater influence. We can test their effects against κ using a multifactor ANOVA. Additionally, we will investigate the relative effects of geometrical properties like roughness of the material interface and variability in β, the system geometry. In that way we can test the relative effects of each scale of uncertainty in the RDD model. Based on the effectiveness of ANOVA summaries used in this work, we anticipate that this statistical technique will be useful in interpreting and organizing the results of more comprehensive Monte Carlo based studies.

References

[1] G. Chan, A.T.A. Wood, Simulation of stationary Gaussian vector fields, Statistics and Computing 9 (4) (1999) 265–268.
[2] G. De Marsily, Quantitative Hydrogeology: Groundwater Hydrology for Engineers, Academic Press, 1986.

[3] V. Di Federico, S.P. Neuman, Flow in multiscale log conductivity fields with truncated power variograms, Water Resources Research 34 (5) (1998) 975–987.
[4] A. Guadagnini, L. Guadagnin, D.M. Tartakovsky, C.L. Winter, Random domain decomposition for flow in heterogeneous stratified aquifers, Stochastic Environmental Research and Risk Assessment 17 (6) (2003) 394–407.
[5] C. Guler, G.D. Thyne, Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy c-means clustering, Water Resources Research 40 (12) (2004).
[6] S.S. Hubbard, J.S. Chen, J. Peterson, E.L. Majer, K.H. Williamson, D.J. Swift, B. Mailloux, Y. Rubin, Hydrogeological characterization of the South Oyster Bacterial Transport Site using geophysical data, Water Resources Research 37 (10) (2001).
[7] Z.M. Lu, D.X. Zhang, On stochastic modeling of flow in multimodal heterogeneous formations 38 (10), October 2002.
[8] P.M. Meier, J. Carrera, X. Sanchez-Vila, A numerical study on the relationship between transmissivity and specific capacity in heterogeneous aquifers, Ground Water 37 (4) (1999) 611–617.
[9] S.P. Neuman, Relationship between juxtaposed, overlapping, and fractal representations of multimodal spatial variability, Water Resources Research 39 (8) (2003).
[10] H. Rajaram, D. McLaughlin, Identification of large-scale spatial trends in hydrologic data, Water Resources Research 26 (10) (1990) 2411–2423.
[11] Y. Rubin, S.S. Hubbard, A. Wilson, M.A. Cushey, Aquifer characterization, in: J.W. Delleur (Ed.), The Handbook of Groundwater Engineering, CRC Press, 1998.
[12] Y. Rubin, Flow and transport in bimodal heterogeneous formations, Water Resources Research 31 (10) (1995) 2461–2468.
[13] Y. Rubin, A.G. Journel, Simulation of non-Gaussian space random functions for modeling transport in groundwater, Water Resources Research 27 (7) (1991) 1711–1721.
[14] H. Suk, K.K. Lee, Characterization of a ground water hydrochemical system through multivariate analysis: clustering into ground water zones, Ground Water 37 (3) (1999).
[15] C.L. Winter, D.M. Tartakovsky, Mean flow in composite porous media, Geophysical Research Letters 27 (12) (2000) 1759–1762.
[16] C.L. Winter, D.M. Tartakovsky, A. Guadagnini, Numerical solutions of moment equations for flow in heterogeneous composite aquifers, Water Resources Research 38 (5) (2002).
[17] C.L. Winter, D.M. Tartakovsky, Groundwater flow in heterogeneous composite aquifers, Water Resources Research 38 (8) (2002).
[18] C.L. Winter, D.M. Tartakovsky, A. Guadagnini, Moment differential equations for flow in highly heterogeneous porous media, Surveys in Geophysics 24 (1) (2003) 81–106.
[19] D. Zhang, Stochastic Methods for Flow in Porous Media: Coping with Uncertainties, Academic Press, 2001.