Application of a multiple-point geostatistical model for understanding groundwater dynamics in a structurally heterogeneous aquifer

Michael J. Ronayne\textsuperscript{1}, Steven M. Gorelick\textsuperscript{1}, and Jef Caers\textsuperscript{2}

For presentation at the 21st Annual Meeting
Stanford Center for Reservoir Forecasting
May 8-9, 2008

\textsuperscript{1}Department of Geological and Environmental Sciences
Stanford University, Stanford, CA 94305-2115
\textsuperscript{2}Department of Energy Resources Engineering
Stanford University, Stanford, CA 94305-2220

Abstract

Groundwater dynamics are controlled by discrete geologic structures in a variety of aquifer settings. In this work, we analyze data from the well studied alluvial-fan system that underlies the Lawrence Livermore National Laboratory. There, subsurface flow and transport behavior is strongly influenced by discrete channel deposits, which are embedded within less permeable floodplain deposits. We model the distribution of these deposits using a multiple-point geostatistical technique. By coupling the geostatistical model with a dynamic groundwater flow model in a simulation-inversion framework, we identify specific channel structures that are consistent with hydraulic data from a multiple-well aquifer test. We perform shortest-path analysis to characterize the geometry of the simulated channels and to locate critical high-permeability conduits that control the system’s response to pumping. In one region of the subsurface, all successful inverse solutions identify the same set of stacked channel deposits, a geologic feature that provides an explanation for anomalous response behavior observed during the aquifer test.
1. Introduction

Subsurface structures exert a primary influence on groundwater dynamics in many geologic settings. Examples of discrete structures characterized by high or low permeability include faults, fractures, and depositional features such as coarse-grained channel fills. The spatial arrangement of these features often represents the first-order (i.e., most hydrologically significant) heterogeneity since their presence tends to produce large discontinuities in the permeability distribution [Fogg, 1986; Western et al., 2001; de Marsily et al., 2005]. Identifying and quantifying the impact of dominant structures at all relevant scales represents an important, continuing, research challenge [Proce et al., 2004; Schulz et al., 2006].

In geologic environments where discrete high-permeability (k) features are present, the connectedness and geometry of these features is critical. Many recent studies in the groundwater literature have emphasized the fact that, when the most permeable subsurface material is highly connected, flow channeling is enhanced and solute transport behavior often belies predictions based on the classical advective-dispersive model [Zinn and Harvey, 2003; Zheng and Gorelick, 2003; Liu et al., 2004; Knudby and Carrera, 2005; Feyen and Caers, 2006].

Although the availability of a connected path of high-permeability material is key, the geometry of that structural pathway is also important. Ronayne and Gorelick [2006] investigated fluid flow in synthetic systems spanned by high-k channel networks and demonstrated that the degree of flow channeling was very sensitive to the network geometry. In the study of porous composites, the geometry of permeable inclusions is often found to be an important influence on effective transport properties [Torquato, 2002]. In fractured rock settings, the orientation and arrangement of fractures are strong controls on flow channeling and the extent of fracture-matrix fluid exchange [Taylor et al., 1999; Matthäi and Belayneh, 2004].

Percolation theory offers a general framework for studying connections and the geometry of connected material. Percolation models are commonly used to represent heterogeneous porous media [Sahimi, 1993; Hunt, 2001], and there are many useful results in the literature that relate fluid flow behavior to the geometrical characteristics of percolation clusters. One particularly important property is the shortest path; the length and geometry of the shortest path between sites
on a cluster has been studied for various percolation models [Hermann and Stanley, 1988; Dokholyan et al., 1998; Paul et al., 2002]. Knowledge of these properties is useful for understanding dynamics on permeable clusters. For example, it has been demonstrated that the minimum transport time for conservative solute migration between two locations on a cluster is strongly correlated with the shortest-path length and fractal dimension [Lee et al., 1999; Andrade et al., 2000; Araújo et al., 2002]. These and other related theoretical results provide insight for field studies that aim to find a linkage between observations of hydraulic response and the properties of subsurface structures.

In this work, we consider data from the Lawrence Livermore National Laboratory (LLNL) field site in northern California. There, subsurface flow occurs within an alluvial sequence containing channel deposits. The permeability of the channel deposits is significantly higher than that of the adjacent overbank and floodplain material. Carle [1996] was among the first to explore how the structural heterogeneity at LLNL impacts hydraulic response behavior. Fogg et al. [2000] investigated the connectivity of the channels and proposed a “connected-network paradigm” to describe the heterogeneity, which may be broadly applicable to aquifers of fluvial origin. In this work, we build upon those previous studies and present a new quantitative method for identifying channel structures that control subsurface hydraulic response. This study is different from previous studies in that we use multiple-point geostatistics to develop high-resolution numerical models of the subsurface heterogeneity at LLNL. Also new is our application of a Bayesian inverse modeling technique, which allows us to identify specific channels that explain complex hydraulic response data measured during a site aquifer test. Previous investigators, most notably Fogg et al. [2000], have concluded that the aquifer-test response behavior is a consequence of strong heterogeneity and a high degree of channel connectivity within the alluvial system. By applying an inverse model and performing shortest-path analysis, we go a step further with this work; namely, we locate specific high-$k$ conduits, formed by interconnected channel deposits, that control the system’s response to pumping. Since our inverse solutions demonstrate good agreement with the measured hydraulic heads at several observation wells, and because we identify specific structures that explain observations of anomalous hydraulic response, the results of this study enhance the understanding of groundwater dynamics at LLNL. The inverse modeling technique that we employ, coupled with the shortest-path analysis, may be
useful in a variety of geologic settings where the heterogeneity is dominated by discrete structures.

The remainder of this paper is organized as follows. In Section 2, we review the hydrogeology of the LLNL site and describe the aquifer-test data set. Section 3 covers the geostatistical simulation and inverse modeling techniques. In Section 4, we describe our groundwater flow model, which we use to analyze the aquifer-test observations. Results are presented in Section 5 and, in Section 6, we highlight our conclusions and discuss some avenues for future research.

2. Study Area and Site Data
Field data used in this study are from an alluvial fan deposit that underlies the Lawrence Livermore National Laboratory. Located in northern California (Figure 1a), LLNL is situated along the boundary between the Livermore Valley and the Diablo Mountains, part of the California Coast Ranges. In addition to ongoing hydrogeologic investigations to support remediation activities, the Livermore site has been the subject of numerous research studies. Of particular relevance to this work are previous efforts to model the subsurface heterogeneity using a discontinuous facies framework [Carle, 1996; Fogg et al., 2000]. In these studies, a transition probability Markov-chain method was used to obtain stochastic realizations of the subsurface architecture. Recently, Lee et al. [2007] compared permeability fields generated with this method to fields simulated using a multivariate Gaussian model. They found that models constructed using transition probabilities were better able to explain the general response behavior observed during a site aquifer test. However, the previous modeling efforts have thus far been unable to successfully reproduce the response data at multiple observation wells. Both the transition probability and multivariate Gaussian approaches rely on two-point statistical measures. The multiple-point geostatistical approach used in this study allows for more explicit representation of connected curvilinear features, such as channel bodies (see Section 3).

2.1 Geology
Sediments that make up the LLNL alluvial fan system were delivered by streams originating in the Diablo Mountains. Consequently, the present fan system consists of discrete channel deposits embedded within finer-grained and less permeable floodplain deposits [Weissmann et
The floodplain material occupies a significantly higher volume fraction than the channel material. Previous conceptual models have envisaged a labyrinth geometry [Weber and van Geuns, 1990] to describe the distribution of channels. This interpretation is supported by the borehole data and is also consistent with general expectations for the depositional setting.

Noyes [1990], Carle [1996], and subsequent investigators (e.g., Fogg et al. [1998]; Weissmann et al. [2002]) demonstrated that textural descriptions and borehole geophysical data could be used to differentiate channel deposits from other finer-grained deposits within the alluvial system at LLNL. Following the general approach taken in those previous studies, we analyzed core descriptions and geophysical logs (single-point resistance and gamma logs) from 50 boreholes throughout the study area (extent shown in Figure 1b). In addition to estimating the fraction of each material type, the borehole data allowed us to determine realistic dimensions for the channel deposits. Channels comprise the coarsest subsurface material at LLNL. From textural descriptions, these deposits are identified as gravels, sandy gravels, and gravelly sands. On geophysical logs, channels are characterized by high resistivities and relatively low gamma signatures. Sandy silts, clayey silts, and silty sands are the dominant textures in the floodplain areas. These finer-grained deposits are distinguished on geophysical logs by moderate to low resistivities and gamma ray levels that are higher relative to the channels. Figure 2 shows example logs for two site boreholes. By jointly analyzing the geophysical data and textural descriptions from core samples, discrete channel intervals are identified with a high degree of confidence. Our analysis is consistent with the aforementioned previous studies, indicating that the volume fraction of the channel material is in the range of 15 to 20 percent. For the subsurface area and sampling locations that we considered, the average channel thickness is 1.3 meters.

The boreholes shown in Figure 2 are separated by only 3.9 meters and are situated along a line that is transverse to the main paleoflow directions. It is noteworthy that similar channel intervals are observed along both boreholes. We found this to be a common occurrence when analyzing data from nearby (< 10 m apart) locations. This ability to correlate individual channel deposits over short lateral distances did not apply to the more significant distances (> 20 m) that separate most boreholes. However, the successful correlation of channels for several closely-spaced
borehole pairs was important - not only to corroborate the conceptual notion that the channels are key depositional features that occupy significant volumes, but also to approximate a reasonable lower bound on the typical channel width.

One other important characteristic of the subsurface heterogeneity is the presence of laterally continuous paleosols. Given their high clay content and extremely low permeability, paleosol deposits can create extensive barriers to flow. Recent hydrogeologic analysis has led to the determination of several different hydrostratigraphic units (HSUs) within the alluvial sequence at LLNL [Mansoor et al., 2002; Hoffman et al., 2003]. Distinct HSUs have limited hydraulic communication, likely a result of bounding paleosol surfaces [Weissmann et al., 2002; Bennett et al., 2006]. In our model of the subsurface architecture, we include major paleosols (see Section 4), which are key features affecting flow and transport behavior.

### 2.2 Hydraulic Response Data

In this study, we analyze subsurface hydraulic response data from an aquifer test conducted at the LLNL site in April 1994. Groundwater was extracted from a single pumping well for a duration of 24 hours. The initial pumping rate was 98.5 m$^3$/day and, after 400 minutes, was reduced and maintained at approximately 69.1 m$^3$/day. Pressure transducers were installed to record transient hydraulic heads at multiple observation wells. A more detailed account of the aquifer test is given by Carle [1996].

Our analysis is focused on six observation wells (well locations are shown in Figure 1b). We corrected the observed hydraulic head data for atmospheric pressure fluctuations, in order to obtain the head changes that are produced by well pumping. Hereafter, we present only the corrected observation data. We will adopt the term drawdown to refer to the amount of head decline relative to the initial hydraulic head measurement at the start of pumping.
2.3 Anomalous Response Curves

Spatially variable drawdowns that result from pumping are governed by a diffusion equation

\[
\frac{\partial h(x)}{\partial t} = \nabla \cdot (D(x)\nabla h(x,t)) + Q(x,t)
\]

where \( h \) is the hydraulic head and \( Q \) represents a source/sink term. For groundwater applications, the relevant diffusion coefficient is the hydraulic diffusivity, \( D = K/S_s \), where \( S_s \) is the specific storage and \( K \) is the hydraulic conductivity, \( K = (k/\rho g/\mu). \) In response to pumping, the propagation rate of a given magnitude of drawdown is proportional to the hydraulic diffusivity [Streltsova, 1988].

Composite drawdown plots for the six observation wells are shown in Figure 3. The individual response curves are separated on the basis of each well’s screened interval. Three observation wells are screened in the upper hydrostratigraphic unit, HSU-2, and three are screened in the lower unit, HSU-3A.

It is useful to compare the field observations shown in Figure 3 to theoretical response curves that are obtained using well-studied idealized models. In the case of a medium with uniform \( D \), a drawdown pulse moves through the system in a manner that is directionally invariant. For a confined aquifer system with uniform \( D \) and boundary conditions consistent with assumptions employed in the Theis [1935] solution, data points \( (s \text{ versus } tl^2) \) from observation wells at various distances, \( r \), would collapse onto the same curve. Inspection of Figure 3 reveals that this is clearly not the case for the LLNL aquifer-test data. Departure from ideal behavior is especially pronounced for the HSU-2 observation wells. Of particular interest is the quick response and relatively large drawdowns recorded at W264, the most distant observation well.

Analyzed together, we note that the observed drawdown curves may be regarded as “anomalous” given that the response behavior differs so markedly from classical predictions [Cortis and Knudby, 2006]. In this work, we present a new spatially distributed model that relates this interesting hydraulic response to the locations, geometry, and connectivity of key channel features.
3. Geostatistical Methodology

Multiple-point geostatistics provides an effective framework for characterizing subsurface hydraulic property distributions that are dominated by discrete structures. The essence of the method is the application of higher-order statistics to capture patterns in heterogeneous fields. The ability to identify and reproduce complex multi-point patterns permits accurate modeling of features characterized by some irregular geometry.

Recent field applications of multiple-point geostatistics have focused on spatially variable petrophysical properties in petroleum reservoirs. For example, Strebelle et al. [2003] and Liu et al. [2004] applied the technique to model sand body geometry in offshore clastic reservoirs.

The multiple-point method is well suited for modeling the subsurface channel distribution at LLNL given its emphasis on discrete structures and their geometry. Moreover, since the geometry of individual discrete objects affects the long-range connectivity properties for groups of those objects [Garboczi et al., 1995; Baker et al., 2002], the multi-point approach offers some advantages (compared to two-point indicator geostatistical techniques) for realistically modeling the dynamic effects of connectivity. The “groups of objects” that are of interest in this study are connected channel deposits (or channel clusters), which will be discussed in Section 5.3.

A general two-point correlation function (e.g., $R_2^{(a)}(\mathbf{u}, \mathbf{u}')$) considers the probability of finding the same facies identifier, $a$, at locations $\mathbf{u}$ and $\mathbf{u}'$. With the multi-point approach, higher-order correlation functions are assembled to characterize the spatial variability more completely. Information gathered from the subsurface is too sparse to directly infer these higher-order statistics. Instead, a so-called 3D training image is created by interpreting the 3D geologic structure from subsurface data. For example, if borehole data, geophysical records, and geologic interpretation indicate a fluvial depositional environment, then a training image model is built to describe general characteristics of the channel deposits. The training image only depicts the conceptual model and need not locally match the actual subsurface data. In addition to site-specific information gained from geologic sampling, other relevant data such as outcrop and analog data may be used to constrain the geometry and orientation of the channel features.
Several stochastic simulation methodologies have been developed to generate 3D realizations that reproduce the style of heterogeneity depicted in a training image and, at the same time, are locally constrained to data [Strebelle, 2000; Zhang et al., 2006; Arpat and Caers, 2007]. In this paper we follow Strebelle’s approach, where multiple-point data events (essentially patterns) are extracted from the training image and anchored to subsurface data. The concept of the data event is critical for practical implementations of the technique [Strebelle, 2002; Okabe and Blunt, 2005].

Our geostatistical simulation model for the LLNL site considers two dominant categories: (i) the channel facies and (ii) the floodplain matrix facies. This discontinuous facies framework requires a discrete space random function (SRF) to describe the heterogeneity [Johnson and Dreiss, 1989]. Here, we use an indicator SRF defined as

\[
I(u) = \begin{cases} 
1, & \text{if } u \in C \\
0, & \text{otherwise}
\end{cases}
\]  

(2)

where \( u \) is the space coordinate and \( C \) refers to the set of all channels. If the function \( I(u) \) takes on a value of 1, then the channel facies is present at that location. Otherwise, the floodplain facies is present. In this study, we use multiple-point statistics to describe the variability of the indicator SRF in (2). Notice that we do not include a separate paleosol facies in (2). Instead, we represent major paleosol deposits deterministically, which will be discussed in Section 4.

For a particular simulation domain, the complete set of model parameters is given by [Caers and Hoffman, 2006]

\[
m = \{ I(u_1), \ldots, I(u_N) \} = \{ I_1, \ldots, I_N \}
\]

(3)

where \( N \) is the total number of grid cells. The prior probability distribution, which describes the dependencies among those model parameters, can therefore be expressed as

\[
f(m) = \Pr\{ I(u_1) = i_1, \ldots, I(u_N) = i_N \} = \Pr\{ I_1 = i_1, \ldots, I_N = i_N \}
\]

(4)
In our case, the full multivariate prior is not explicitly known; only conditional probability distributions are specified. Indeed, using the exact decomposition

\[ f(m) = \Pr\{I_1 = 1\} \times \Pr\{I_2 = 1 | i(u_1)\} \times \ldots \times \Pr\{I_N = 1 | i(u_1), \ldots, i(u_{N-1})\} \]  

(5)

only the univariate conditional distributions, describing the probability of each facies at a specific location, are required. In the approach that we use, these conditional distributions are modeled directly from the 3D training image [Strebelle, 2000]. Sequential simulation is used to draw realizations. Realizations (or samples) drawn from this prior are termed unconditional realizations.

An important attribute of any practical geostatistical simulation method is the ability to condition to local site data. The following posterior probability distribution includes an observed data vector, \( d_1 \), which contains \( n \) locations (borehole intervals) where the facies is known

\[ f(m | d_1) = \Pr\{I_1 = i_1, \ldots, I_N = i_N | d_1\} \]  

where \( d_1 = \{i(u_\alpha), \alpha = 1, \ldots, n\} \).

(6)

Conditional realizations are generated by drawing samples from the posterior distribution in (6). Conditioning to direct observations of the indicator variable (observed data in \( d_1 \)) is performed using the same sequential simulation methodology [Deutsch and Journel, 1998]. This conditioning is straightforward since the direct data are treated the same as previously simulated values.

3.1 Integration of Dynamic Data

We also condition the geostatistical realizations to dynamic hydraulic response data. That is, we perform inverse modeling in an effort to find facies distributions that are consistent with transient hydraulic heads measured during a site aquifer test. These dynamic data, \( d_2 \), which are indirectly (nonlinearly) related to our indicator variable, can be incorporated within a Bayesian framework
A forward operator $F$ explains the functional relationship between the dynamic data and the model parameters; $d_2 = F(m)$ [Tarantola, 2005]. The last expression in (7) holds if we assume that $d_1$ and $d_2$ are conditionally independent.

Equation (7) is the full posterior distribution that includes both direct and indirect data types. The traditional sampling approach calls for explicit analytical specification of likelihood and prior. Except for the multi-Gaussian framework and linear data, such specification becomes analytically intractable or requires Markov chain Monte Carlo techniques [Mosegaard and Tarantola, 1995; Omre and Tjelmeland, 1996] that are prohibitively expensive when the forward model $F$ is a high-resolution flow simulator.

Instead, we follow an alternative approach specified in detail by Caers and Hoffman [2006] that relies on a decomposition of the posterior in terms of two “pre-posterior” distributions. The pre-posteriors describe the dependence of an occurrence of $I$ (the indicator variable of a grid cell) on each data type separately. We introduce some additional notation to further explain the method:

$$
Pr\{I(u_j) = 1 \mid i(u_1), \ldots, i(u_{j-1}), d_1, d_2\} = Pr\{A_j \mid B_j, d_2\}
$$

where $A_j = \{I(u_j) = 1\}$

$$
B_j = \{i(u_1), \ldots, i(u_{j-1}), d_1\}
$$

The distribution $Pr\{A_j \mid B_j, d_2\}$ in (8) is the local conditional probability distribution for the $j$th grid cell. It considers both data types, $d_1$ and $d_2$, as well as previously simulated facies values for the other $(j-1)$ cells. To facilitate estimation of this distribution during sequential simulation, two pre-posteriors are defined, $Pr\{A_j \mid B_j\}$ and $Pr\{A_j \mid d_2\}$. The pre-posterior $Pr\{A_j \mid B_j\}$ is modeled directly from the training image as described above. The pre-posterior for the dynamic data, $Pr\{A_j \mid d_2\}$, is estimated iteratively using the probability perturbation method [Caers, 2003; Caers and Hoffman, 2006].
The probability perturbation method simplifies the inversion problem by focusing on one (global) or a few (regional) perturbation parameters. The following equation explains the effect of the perturbation parameter:

\[
Pr\{A_j | d_2\} = [(1 - r_D) \times i^{(0)}(u_j)] + [r_D \times Pr\{A_j\}] \quad j = 1, \ldots, N \quad (9)
\]

where \(r_D\) is a perturbation parameter, \(i^{(0)}(u_j)\) is the simulated facies value corresponding to the previous best-matching realization, and \(Pr\{A_j\}\) is the marginal probability. An \(r_D\) value of 0 means that the existing facies should be maintained at location \(u_j\). An \(r_D\) value of 1 corresponds to a maximum perturbation, such that \(Pr\{A_j | d_2\}\) is set equal to the overall channel fraction. To find an optimal value of \(r_D\), flow simulations are required. After each simulation, an objective function is evaluated, and then the value of the perturbation parameter is updated in an effort to minimize the objective function. In our case, the objective function quantifies the mismatch between observed and model-simulated drawdowns.

To increase the efficiency of the inverse model, we optimize on two (regional) perturbation parameters simultaneously. We use one \(r_D\) value to perturb the local probabilities in HSU-2 and a separate \(r_D\) value for model cells in HSU-3A, the lower hydrostratigraphic unit. With the sequential simulation framework, the use of multiple perturbation parameters does not produce any artifacts [Hoffman and Caers, 2005]. That is, a sample from (7), without discontinuities between regions, can still be obtained.

We use the method suggested by Journel [2002] to combine the pre-posteriors and estimate \(Pr\{A_j | B_j, d_2\}\). In practice, \(Pr\{A_j | B_j\}\) is formulated during sequential simulation and combined with \(Pr\{A_j | d_2\}\) to obtain \(Pr\{A_j | B_j, d_2\}\). Journel [2002] and Hoffman [2005] provide additional details and further justification of this approach for similar applications.

4. Model Development
A key goal of this work is to provide a physical explanation for anomalous response behavior observed during the aquifer test. To accomplish this, we require a structural model of the subsurface heterogeneity and a dynamic model for simulating the response to pumping. The first
part of this section addresses the heterogeneity model. Example realizations of the subsurface architecture are presented and discussed. In the second part of this section, we describe our groundwater flow modeling approach.

4.1 Subsurface Architecture
We use a training image-based multiple-point geostatistics approach to model the spatial distribution of the two dominant facies. The training image constrains the prior model. For the alluvial deposit investigated here, our training image contains specific information on the following properties: (1) the relative fractions of the channel and floodplain facies; (2) the thickness and (3) width of channel deposits; (4) the channel orientation (paleoflow direction); and (5) the sinuosity of channels. The first two items are estimated directly from available geologic data. As noted in Section 2.1, the average channel thickness estimated from borehole samples is 1.3 meters. Since the boreholes arbitrarily intersect portions of channel deposits that in reality are characterized by some curvilinear cross-sectional geometry, the sampled thickness is expected to be less than the maximum channel thickness. We use 2 meters as a representative maximum thickness. Site data from closely-spaced boreholes indicate that the typical channel width is at least 5 meters. Drawing from an extensive review of data from the sedimentology literature, Gibling [2006] notes that many channel bodies preserved in alluvial fans are characterized by a “broad ribbon” geometry with width-to-thickness (W/T) ratios from 5 to 15. We assume W/T = 10, meaning that the maximum target width for an individual channel deposit is 20 meters. Based on modern drainage patterns around LLNL, it appears that two channel orientations are likely, east-to-west and southeast-to-northwest [Hoffman and Dresen, 1990; Weissmann et al., 2002]. For the work presented here, we focus on the latter direction (i.e., we specify channels that trend to the northwest). The presence of channels in this direction was strongly indicated by the aquifer-test response data that we studied. Perhaps the most uncertain of the properties listed above is the channel sinuosity. In our training image, we assume channel deposits with relatively low sinuosity. Given typical gradients that characterize a fan surface, highly sinuous channels are not anticipated, except possibly along the most distal reaches.

We used the Boolean technique of Deutsch and Tran [2002] to develop our 3D training image, shown in Figure 4. Note that the training image is not conditioned to any local geologic samples.
or hydrologic observations; it represents a conceptual description that includes some quantitative information regarding the facies geometry. The size of the full training image is 500 m × 500 m × 16 m in the x-, y-, and z-directions. In Figure 4, a subset of this volume is displayed (500 m × 375 m × 16 m). As demonstrated for the single layer (horizontal slice) shown in the figure, all channels in the training image are modeled as being laterally continuous. We permit some variability in the channel orientation, given the potential for local variations in paleoflow direction.

We generate 3D realizations of the facies distribution using the algorithm of Strebelle [2000], with the 3D training image as input. More specifically, multiple-point data events identified during a scan of the training image are used to inform simulated realizations of the facies distribution. In addition to data events (patterns) from the training image, other information used to guide the sequential simulation procedure includes direct conditioning data (observed facies from boreholes) and indirect (drawdown-response) data. At specific locations and depth intervals sampled by boreholes, discrete facies (i.e., channel or floodplain material) were identified using the geologic analysis procedure described in Section 2.1. These samples, which are used to condition the geostatistical realizations, constitute the \( \mathbf{d}_1 \) vector for our application. Sample locations are shown in Figure 1b. Along each sampled borehole, there are multiple depth intervals where the facies occurrence is prescribed. It should be noted that data were also available to support facies identification at the pumping well, W612, and at all but one of the observation wells. Accurate determination of the facies distribution is particularly important along the pumping well’s screened interval, since structural heterogeneity around the pumped location strongly controls early-time response dynamics throughout the affected volume [Leven and Dietrich, 2006]. The data for W612 indicate that the well screen intersects a channel deposit. In fact, given the emphasis on remediation and the need to implement efficient pump-and-treat methods, most wells at LLNL are intentionally screened across high-\( k \) channel material. The combined physical/engineered alluvial system could therefore be described as a “plumbed” network of interconnected channel deposits contained within less permeable sediments [Fogg et al., 2000]. This suggests that some characteristics of the aquifer-test response dynamics may be similar to those observed in fractured rock settings, where packers are used to isolate specific high-\( k \) fracture zones.
Two example conditional realizations from the geostatistical model are provided in Figure 5. We discretized the model space using a grid comprised of uniform cells with dimensions $\Delta x = \Delta y = 2.5$ m and $\Delta z = 0.4$ m. The overall simulation domain contains 7,296,000 cells. This large, high-resolution grid was required in order to explicitly represent individual channel deposits throughout the subsurface volume ($> 10^7$ m$^3$) that we consider. The volume selected for presentation in Figure 5 is a subset (~45 percent) of the total simulated volume.

One noticeable characteristic of the realizations shown in Figure 5 is that individual channels are less (horizontally) continuous as compared to those included in the training image. This apparent difference in channel continuity is primarily a result of ergodic fluctuations [Deutsch and Journel, 1998] and has been described in more detail for multiple-point geostatistics by Caers and Zhang [2004]. For our work, practical considerations of computational efficiency placed limits on the training image size; we used a training image that is slightly smaller than the simulation grid. Consequently, the simulation algorithm has difficulty reproducing the longest-range patterns because adequate sample statistics are not available to fully describe those patterns. However, as will be discussed Section 5.3, vertical connections among different channel deposits produce long-range connectivity of the high-$k$ material.

In addition to the ergodicity issue, differences between the training image and output realizations are also a consequence of the stochastic, pixel-based multiple-point algorithm that we use. Some variability in geometrical properties (beyond the variability contained in the training image) is introduced by the geostatistical simulator. This issue has been documented elsewhere [e.g., Caers and Zhang, 2004]. Most importantly, the output conditional realizations are realistic for the depositional environment that we consider, consistent with the local geologic data, and capture many of the key properties specified in the training image.

Also represented in Figures 5a and 5b is the contact between the two hydrostratigraphic units that comprise our modeled section. Previous work indicates that this contact is most likely formed by an extensive paleosol deposit. Using existing data and interpretations regarding its elevation at borehole locations, we contoured the paleosol surface. We incorporate this surface
into our model by assigning the lowest $k$ value to all grid cells that it intersects. However, at locations between the boreholes, we permit discrete erosional breaks in the paleosol surface, and we simulate those breaks stochastically. The aquifer-test data support the notion of a discontinuous paleosol surface; observation points in HSU-3A respond relatively quickly to pumping in HSU-2. Therefore, at least in this particular area, there appears to be some hydraulic communication between the two HSUs, beyond the typical slow-leakage response that would be observed with a competent and continuous low-$k$ unit. In the model, we permit this communication by allowing breaks in the bounding paleosol. Our approach is simply to specify a break (i.e., assign high-$k$ channel rather than low-$k$ paleosol material) at any location where a simulated channel deposit coincides with the (contoured) paleosol surface.

4.2 Groundwater Flow Model

We apply the finite-difference code MODFLOW-2000 [Harbaugh et al., 2000] with a multigrid matrix solver [Wilson and Naff, 2004] to discretize and numerically solve (1) for our problem domain. The grid spacing used for the groundwater model is identical to that of the geostatistical simulation grid. Thus, no upscaling or downscaling step is required prior to running flow simulations.

The modeled vertical section ranges from approximately 28 to 32 m in thickness and spans the two hydrostratigraphic units where pumping-induced hydraulic head declines were observed. Top and bottom model boundaries follow low-permeability paleosol deposits and are specified as no-flow [$\partial h(\mathbf{x})/\partial n_0 = 0$ for $\mathbf{x}$ on the boundary segment]. Boundary conditions prescribed along the model sides are shown on the plan-view map in Figure 1b. Inflow and outflow segments are treated using Cauchy boundary conditions; the computed flux into or out of the modeled region depends on the difference between the boundary head (permitted to vary) and a fixed head located outside the domain. Also shown in Figure 1b are two no-flow boundary segments that coincide with regional flowlines. For the aquifer-test conditions and hydraulic properties evaluated here, these external boundaries are sufficiently far from the pumping location and therefore have a negligible impact on simulated drawdowns.
For each realization of the subsurface architecture, we use the simulated facies distribution as a template and assign hydraulic property values accordingly. The values of hydraulic conductivity and specific storage for each material type are given in Table 1. The tabulated property values were obtained primarily through model calibration, although we used relevant field data to guide the calibration process. In particular, we considered previous estimates of equivalent hydraulic conductivity from more than 20 single-well pumping tests. As would be expected, these equivalent $K$ values are in the range between our end-member hydraulic conductivities, which are representative of the low-$K$ floodplain and high-$K$ channel material. We emphasize that the property values listed in Table 1 are well within the range of published values for each material texture [e.g., Freeze and Cherry, 1979].

We apply our groundwater model to simulate the 24-hour aquifer test described in Section 2. Computed drawdowns are compared to measured drawdowns at the six observation well locations. Since we use a very large grid, deal with strong permeability contrasts, and run the transient flow model within a broader simulation-inversion framework, computational efficiency is a key consideration. All simulation results presented in this paper were generated on a shared-memory Linux workstation. On this computing platform, both the geostatistical simulations and the forward flow simulations access over 1GB of physical memory. The memory requirement of the geostatistical model (nearly 4GB) is significantly higher than that of the flow simulation model. The run time to generate a single geostatistical realization of the facies distribution is approximately 20 minutes, while the flow simulation run time is typically in the range of 50 to 60 minutes. Therefore, each function evaluation (iteration) directed by the inverse model requires about 75 minutes. The inverse model performance is discussed in Section 5.2.

5. Results
In this section, flow modeling results are presented for five different realizations of the subsurface architecture. Each realization was obtained using the inverse procedure described in Section 3.1.
5.1 Flow Model Response
For all realizations, the dynamic model indicates that, in response to pumping, groundwater flow becomes focused within the high-\(k\) channel material. This is consistent with previous modeling results reported by Carle [1996] and Fogg et al. [2000]. Additionally, we find that the drawdown distribution is highly sensitive to the locations and geometry of specific channel deposits. This sensitivity enables the inversion approach that we use to identify individual channels. Because the hydraulic response behavior is highly sensitive, dynamic observations are expected to have good information content that can be exploited to resolve the channel structure.

Figure 6 shows the simulated drawdown distribution during the first hour of pumping for one particular realization. As revealed in the figure, early-time drawdown closely tracks a large channel body that is intersected by the pumping well screen. The channel structure preferentially induces drawdowns to the northwest and southeast, an effect that facilitates the rapid response observed at a distant well, W264. From Figure 6, it is clear that the drawdown propagation rate within the channel material is higher than the rate for the floodplain matrix. This is due to the contrasting hydraulic diffusivities of each facies. As a result of this behavior, the response timing at a given observation point is a function not only of its distance, but also its connectedness (via the network of channel deposits) with the pumped area.

5.2 Performance of the Inverse Model
In Figure 7, model-simulated drawdown is compared to the measured drawdown at observation well locations. On each individual well plot, simulated results are presented for all five inverse solutions. The model generally performs better at the HSU-2 observation wells. Difficulty in reproducing drawdown observations for the deeper wells may be partly related to uncertainties in the data. Within HSU-3A there was a trend of declining hydraulic heads prior to the start of the aquifer test. Although we used available measurements to correct for this trend at W712 and W267, the pre-test variability is still a complicating factor that may introduce some (data) errors.
We use an overall sum of relative errors as our objective function to direct the inverse model

\[
\text{Obj} = \sum_{w=1}^{N_{\text{well}}} \sum_{t=1}^{T_O} \frac{\left( s_{\text{sim}}^w(t) - s_{\text{obs}}^w(t) \right)}{s_{\text{obs}}^w(t)}
\]  

(10)

where \( s_{\text{obs}}^w(t) \) is the observed drawdown for well \( w \) at time \( t \), \( s_{\text{sim}}^w(t) \) is the model-simulated drawdown for well \( w \) at time \( t \), \( T_O \) is the total number of observation times (for a given well), and \( N_{\text{well}} \) is the total number of wells. Application of a relative error measure is important since the smaller drawdowns (measured at early times) are often most informative about the subsurface variability. As demonstrated in Figure 6, dynamic head observations that immediately follow an applied stress can be quite sensitive to detailed heterogeneity and connectivity properties. In comparison, late-time measurements depend on larger-scale average properties [Meier et al., 1998].

In Figure 8, we show the objective function (as the mean of the relative errors) at each iteration of the inverse model for an example run. The results indicate a relatively high number of nonproductive realizations. This reflects the discontinuous nature of the problem and the large permeability contrasts that are modeled. Due to the discontinuity (i.e., discrete facies framework), the application of a traditional gradient-based inverse technique is not feasible. Research is underway to explore new approaches for enhancing the efficiency of the probability perturbation method (e.g., Johansen et al. [2007]). Most important for the present work, however, is that the results in Figure 8 reveal a downward trend, showing clear progress towards an inverse solution. Through Monte Carlo sampling from (6), we attempted to find a realization that adequately reproduced the drawdown observations without performing the inversion. This effort was not successful, which indicates that the inverse modeling is necessary to obtain a good solution within a reasonable period of time. Caers and Hoffman [2006] show that the sampling properties of the probability perturbation method (PPM) are similar to those of a classic rejection sampler; however, PPM is far more efficient.

Figure 5c further demonstrates the efficacy of the inverse model. Realizations sampled from (6), which are conditioned to the direct \( d_1 \) data but not the dynamic data, consistently underpredict
both the timing and magnitude of the measured drawdowns at well W264. It requires a specific high-$k$ connection, identified by the inverse model, to reproduce the data at this location. This is true despite the fact that all realizations have the same bulk anisotropy behavior (resulting from similar channel fraction, size, orientation, etc.). In Section 5.3 below, we characterize the specific geologic feature that explains the response behavior at this observation well.

To further evaluate the model performance at each observation well, we consider three different statistical measures, including the mean absolute error (MAE), the mean relative error (MRE), and the Nash-Sutcliffe [1970] coefficient of efficiency (NSE)

\[
NSE = 1 - \frac{\sum_{t=1}^{T_{0}} (s_{\text{obs}}(t) - s_{\text{sim}}(t))^2}{\sum_{t=1}^{T_{0}} (s_{\text{obs}}(t) - \bar{s}_{\text{obs}})^2}
\]  

Values of MAE, MRE, and NSE for each observation well are provided in Table 2. The tabulated results show that relative errors are generally below 25 percent except at well W267. The larger errors at this location may be related to data uncertainty (see above discussion regarding the pre-test trend within HSU-3A).

**5.3 Connectivity properties and shortest-path results**

We now focus on the structural properties that give rise to the complex flow behavior discussed above. We perform cluster counting [Stauffer and Aharony, 1994; Pardo-Igúzquiza and Dowd, 2003] to identify sets of connected channel cells for each realization of the facies distribution. Two locations are connected (i.e., on the same cluster) if there exists a continuous path of channel material between them. The probability of finding two connected channel locations decreases with distance and, for anisotropic models such as ours, is also a function of direction. For the conditional realizations that we generate, the cluster counting routinely finds a spanning cluster. A spanning (or percolating) cluster is a set of connected channel cells that spans the simulation domain in each major grid direction. Our results indicate that vertical connections created by stacked or overlapping channel deposits are key in producing long-range connectivity (both vertical and horizontal) of the high-$k$ material.
The inverse solutions consistently identify a set of vertically stacked channel deposits that provide a direct connection between the pumped interval and observation well W264. Figure 9 depicts this channel structure, which is part of a larger channel cluster, for the example realization presented in Figure 5a. Although in Figure 5a the two wells appear to intersect the same individual channel, their screened intervals are offset by over 3 m (after correcting for the regional stratigraphic dip). Therefore, the high-$k$ connection between these locations requires vertically intersecting channel deposits.

We perform shortest-path analysis to further investigate the significance of the connectivity event involving wells W612 and W264. Using coordinates for each well screen, we define a 3D vector $\mathbf{h}$ that describes the separation of these locations. $\|\mathbf{h}\| = 206.7$ m is the Euclidean distance between the two observation points. An important quantity is $l$, the length of the shortest path between points on the same channel cluster separated by $\mathbf{h}$. The only constraint is that all grid cells along the path must be occupied by channel material. Clearly, $l \geq \|\mathbf{h}\|$. We use Dijkstra’s algorithm to identify the shortest among all connected paths. For the W612↔W264 connection in particular, we find that $l$ assumes its lowest possible value (for the structural grid that we employ) in four of the five calibrated models. In the fifth model, the shortest path is only slightly (< 5 cm) longer. The direct high-$k$ path between these locations helps explain the quick response to pumping that was recorded at W264. We note that this relationship between quick response behavior and a short path is consistent with results presented by Knudby and Carrera [2006] for two-dimensional synthetic transmissivity fields.

We expand upon the above analysis to investigate the distribution of shortest-path lengths for randomly selected points separated by $\mathbf{h}$. After labeling the channel clusters for a particular realization, we use a random search technique to identify connected point pairs that are offset by $\mathbf{h}$. 500 such pairs were selected from 20 different geostatistical realizations drawn from the posterior distribution in (6). We then determined the shortest connected path for each set of points and recorded its length. Figure 10 shows the distribution that is obtained from this exercise. The normalized histogram presented in Figure 10 serves as an approximation of the conditional probability distribution function, $P(l \mid \mathbf{h})$, which has been studied for random
percolation models [Bunde and Havlin, 1996; Dokholyan et al., 1998]. It is interesting to note that, similar to those fractal percolation models, our estimated distribution appears to exhibit power-law behavior over much of its range. However, a larger domain size and more samples (>> 500) would be required to verify actual power-law scaling.

The very short path lengths that we identify for the W612 ↔ W264 connection are significantly below the mean value of $P(l | h)$ for our structural model. However, as indicated by the distribution presented in Figure 10, short paths (corresponding to the separation vector $h$) are relatively common. Therefore, although the inverse modeling results demonstrate that a high-$k$ path is required to reproduce observed drawdowns at W264, we find that the converse is not necessarily true. That is, the existence of a short path is no guarantee that the proper response will be simulated. Other aspects of the heterogeneity, including the overall channel body geometry and structural features outside the W612 ↔ W264 region, also influence the response behavior at this distant observation well.

6. Summary and Conclusions

In this work, we investigated aquifer-test response data from the heterogeneous alluvial-fan system that underlies Lawrence Livermore National Laboratory. The dominant structural heterogeneity at the site is produced by high-$k$ interconnected channel deposits, which are embedded within less permeable floodplain sediments. We used a multiple-point geostatistical technique to simulate the distribution of channels. The generated stochastic realizations are conditioned to both static (observed facies) and dynamic (observed drawdowns) site data. We achieve the dynamic data integration by coupling the geostatistical model with a groundwater flow model in a simulation-inversion framework. Our inverse procedure identifies specific channel structures that explain anomalous hydraulic response behavior observed during the site aquifer test. The model that we present, with key channel structures identified, is the first to closely reproduce the drawdown response data at multiple observation wells.

To further characterize the geometry of connected channel deposits simulated by our geostatistical model, we performed shortest-path analysis. Although not often used in hydrogeologic applications, this analysis can be quite useful for understanding the dynamic
effects of subsurface high- or low-\( k \) structures. The shortest-path results facilitated our interpretation of how specific heterogeneities influence drawdown response behavior. For example, we find that all inverse solutions contain a direct high-\( k \) connection between the pumped area and most distant \( (r > 200 \text{ m}) \) observation well, where rapid hydraulic response was recorded. The structural feature that offers this direct path is a set of vertically stacked channel deposits.

The training image-based geostatistical approach used in this study requires information on the relative abundance and geometry of each modeled facies. For the depositional environment that we considered, key site-specific data requirements included the width, thickness, and sinuosity of fluvial channel deposits. Although we relied on a large data set and a significant amount of prior geologic interpretation, there remains (as always) some uncertainty in the specified properties. Investigations like the one presented here, which attempt to understand the influence of structural heterogeneity, will benefit from new research developments on methods to characterize the geometry of subsurface structures. Regarding fluvial deposits in particular, modern channels offer important guidance. We echo the conclusions of Rubin et al. [2006] on the need to better understand the relationship between the geometry of surficial river channels and the geometry of channel deposits.

Aquifer tests have been and will likely remain an important tool in hydrogeologic investigations. Traditional interpretation of aquifer-test data provides estimates of average hydraulic properties. However, as demonstrated in this work and elsewhere in the literature [Raghavan, 2004 and references therein], when combined with geologic data, a critical analysis of aquifer-test results can yield additional insight into important subsurface details. A key research item moving forward will be to address the information content in anomalous response curves. For example, for a particular location and set of observations, what does such response behavior suggest about the presence and/or distribution of subsurface structures? Finally, we emphasize the importance of new test design and interpretation methods. Hydraulic tomography and similarly motivated techniques may significantly advance efforts to identify key structural features that control subsurface dynamics.
Acknowledgments

Funding for this work was provided by a National Defense Science and Engineering Graduate Fellowship awarded to M. Ronayne. Additional support was provided by the National Science Foundation under grant number EAR-0537668. We thank the Lawrence Livermore National Laboratory Environmental Restoration Department for sharing the field data. Computing resources for this study were provided by the Stanford Center for Computational Earth & Environmental Science.

Table 1. Hydraulic property values.

<table>
<thead>
<tr>
<th>Hydrostratigraphic Unit</th>
<th>Facies</th>
<th>$K$ (m/day)</th>
<th>$S_s$ (m$^{-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSU-2</td>
<td>Channel</td>
<td>90</td>
<td>$5 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>Floodplain</td>
<td>0.3</td>
<td>$1.8 \times 10^4$</td>
</tr>
<tr>
<td>HSU-3A</td>
<td>Channel</td>
<td>72</td>
<td>$4 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>Floodplain</td>
<td>0.24</td>
<td>$1.4 \times 10^4$</td>
</tr>
<tr>
<td>HSU-2 / HSU-3A</td>
<td>Paleosol</td>
<td>$6 \times 10^{-5}$</td>
<td>$1.8 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

Table 2. Statistical measures of model performance at all observation wells. Reported values for each well are obtained by averaging over the five realizations shown in Figure 7.

<table>
<thead>
<tr>
<th>Observation Well</th>
<th>MAE (m)</th>
<th>MRE (-)</th>
<th>NSE (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>W614</td>
<td>0.0058</td>
<td>0.22</td>
<td>0.63</td>
</tr>
<tr>
<td>W714</td>
<td>0.0028</td>
<td>0.16</td>
<td>0.91</td>
</tr>
<tr>
<td>W264</td>
<td>0.0080</td>
<td>0.13</td>
<td>0.84</td>
</tr>
<tr>
<td>W616</td>
<td>0.0190</td>
<td>0.25</td>
<td>0.57</td>
</tr>
<tr>
<td>W712</td>
<td>0.0086</td>
<td>0.22</td>
<td>0.78</td>
</tr>
<tr>
<td>W267</td>
<td>0.0149</td>
<td>0.38</td>
<td>0.24</td>
</tr>
</tbody>
</table>

References


Hoffman, B.T. (2005), Geologically consistent history matching while perturbing facies, Ph.D. dissertation, Stanford University, Stanford, California.


Noyes, C.D. (1990), Hydrostratigraphic analysis of the Pilot Remediation Test Area, LLNL, Livermore, California, M.S. thesis, University of California, Davis.


Theis, C.V. (1935), The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using ground-water storage, *Eos Trans. AGU*, 2, 519-524.


Figure 1. (a) Location of the Lawrence Livermore National Laboratory (LLNL) within California. (b) Base map for study site in the southwest corner of LLNL property. Conditioning data points identify borehole locations where observed facies are used to condition the geostatistical simulations. Arrows indicate approximate regional groundwater flow directions.
Figure 2. Geophysical logs for two neighboring site boreholes. Solid and dashed lines are the resistivity and gamma logs, respectively. Shaded intervals represent coarse-grained textures (sands and gravels) identified from core samples.

Figure 3. Composite drawdown plots for (a) observation wells screened in upper unit, HSU-2; and (b) observation wells screened in lower unit, HSU-3A. $x$-axis shows time (since the start of pumping) divided by distance squared; for each well, $r$ is the distance from pumped location. Drawdown, $s$, is plotted on the $y$-axis.
Figure 4. Training image.
Figure 5. (a) and (b) Two example realizations of the simulated facies distribution. Pumping well (circled) and observation well locations are shown in white. Only observation well W264 is screened across the displayed horizontal slice. The simulated volume shown here spans 525 m in the x-direction, 400 m in the y-direction, and 38 m in the z-direction. y is the north-south direction. (c) and (d) Simulated vs. observed drawdown at Wells W264 and W714, respectively. In these drawdown plots, the solid black line gives the simulated response for the facies distribution shown in (a), which is an inverse solution. The “other” response curves correspond to non-inverse realizations (i.e., samples from Equation 6 that are conditioned to the only direct data); the volume in (b) depicts one such realization.
Figure 6. Early-time propagation of drawdown in vicinity of pumped well. (a) Facies distribution; channel deposits are shown in blue. (b) Simulated drawdown after 10 minutes of pumping. (c) Simulated drawdown after 60 minutes.
Figure 7. Comparison of simulated and measured drawdowns at observation wells. Open circles are the observed data. Lines represent the simulated response for a given realization; M1 through M5 refer to five distinct realizations.
Figure 8. Objective function value at each iteration for an example inverse modeling run. Circles identify iterations where there was an improvement over the previous best realization. The simulated facies distribution corresponding to the final iteration was retained as an inverse solution; this particular realization of the subsurface architecture produced an acceptable match to the drawdown measurements (~ 20 percent error on average).

Figure 9. Main channel body (from example realization) that connects the pumping well and distant observation well W264.

Figure 10. Relative frequency of shortest-path lengths, $l$, for connected point pairs separated by the vector $\mathbf{h}$. Results based on 500 such pairs. Straight line is included to indicate possible power-law scaling.