Hydraulic conductivity imaging from 3-D transient hydraulic tomography at several pumping/observation densities

Michael Cardiff,[1,2] Warren Barrash,[1] and Peter K. Kitanidis[3]

[1] 3-D Hydraulic tomography (3-D HT) is a method for aquifer characterization whereby the 3-D spatial distribution of aquifer flow parameters (primarily hydraulic conductivity, K) is estimated by joint inversion of head change data from multiple partially penetrating pumping tests. While performance of 3-D HT has been studied extensively in numerical experiments, few field studies have demonstrated the real-world performance of 3-D HT. Here we report on a 3-D transient hydraulic tomography (3-D THT) field experiment at the Boise Hydrogeophysical Research Site which is different from prior approaches in that it represents a "baseline" analysis of 3-D THT performance using only a single arrangement of a central pumping well and five observation wells with nearly complete pumping and observation coverage at 1 m intervals. We jointly analyze all pumping tests using a geostatistical approach based on the quasi-linear estimator of Kitanidis (1995). We reanalyze the system after progressively removing pumping and/or observation intervals; significant progressive loss of information about heterogeneity is quantified as reduced variance of the K field overall, reduced correlation with slug test K estimates at wells, and reduced ability to accurately predict independent pumping tests. We verify that imaging accuracy is strongly improved by pumping and observational densities comparable to the aquifer heterogeneity geostatistical correlation lengths. Discrepancies between K profiles at wells, as obtained from HT and slug tests, are greatest at the tops and bottoms of wells where HT observation coverage was lacking.

1. Introduction

[2] Many hydrogeologic applications, particularly prediction of transport and design and operation of groundwater remediation systems, are crucially dependent on an understanding of subsurface aquifer heterogeneity. Variability in subsurface deposits includes a variety of factors that affect plume evolution, including heterogeneity in sediment/soil geochemistry, porosity, and hydraulic conductivity (K). In particular, our limited understanding of site-specific heterogeneity in K (or, in multiphase systems, intrinsic permeability) is continually and universally cited as a key impediment to improving contaminant transport

Additional supporting information may be found in the online version of this article.

[1]Department of Geosciences, Boise State University, Boise, Idaho, USA.

[2]Department of Geoscience, University of Wisconsin-Madison, Madison, Wisconsin, USA.

[3]Department of Civil and Environmental Engineering, Stanford University, Stanford, California, USA.

Corresponding author: M. Cardiff, Department of Geoscience, University of Wisconsin-Madison, 1215 W Dayton St., Weeks Hall, Madison, WI 53706, USA. (cardiff@wisc.edu)

model predictions [*NRC*, 2005; *Anderson and McCray*, 2011]. The large range of natural variability in hydraulic conductivity—which can be up to 13 orders of magnitude, and is often at least 1.5 orders of magnitude or more even at many relatively homogeneous sites [*Sudicky*, 1986; *Woodbury and Sudicky*, 1991]—means that even relatively simple predictions such as conservative tracer breakthrough may be subject to significant uncertainty without detailed characterization information.

[3] Because of the need for accurate information about 3-D hydraulic conductivity (K) variability, numerous aquifer characterization approaches have been advanced. The data source(s) used by different characterization approaches allows categorization into five main groups, as discussed in *Cardiff et al*. [2012]: (1) sample-based (core) methods; (2) pressure-based (hydrologic) methods; (3) tracer-based methods; (4) geophysically based methods; and (5) combination methods. Likewise, characterization approaches can be categorized by the way in which they analyze data. In common practice, data from field tests are often fit using analytical solutions that assume either homogeneity or simple heterogeneities (e.g., layering) within the region of influence of the test (we refer to these as analytical approaches). Given a particular experiment, analytical approaches return one "effective" parameter estimate per analyzed experiment, and heterogeneity is inferred as changes in effective parameters with testing location. A

more computationally intensive approach to data analysis, which has become more practical with the advent of cheap and powerful numerical computing, is to use what is known as a tomographic or "data fusion" approach. In these analysis approaches, data from a large number of tests are fit simultaneously by tuning parameter heterogeneity within a numerical model, and thus produce estimates of subsurface heterogeneity that are consistent with all collected data.

[4] Hydraulic tomography (HT), the focus of this work, is a pressure-based and tomographic aquifer characterization approach in which several pumping tests are performed at different locations within an aquifer and response data (head change) at several wells are analyzed through a tomographic approach. The premise of hydraulic tomography (HT) was originally examined almost 20 years ago [*Gottlieb and Dietrich*, 1995], and studied via a 2-D synthetic application where a constant-rate pumping test was used as the aquifer stimulation. Since that time, numerous advancements in data collection strategies and analysis approaches have been proposed for HT, resulting in a broad diversity of numerical, laboratory, and field testing approaches and a similarly broad diversity of analysis approaches (see the comprehensive summary in *Cardiff and Barrash* [2011]). As one very recent example, *Cardiff et al.* [2013] recently suggested a tomographic approach in which oscillating pumping, rather than constant-rate pumping, is used.

[5] The current "state-of-the-art" in HT research has focused in particular on effective 3-D hydraulic tomography (3-D HT) implementation, in which partially penetrating pumping and observation intervals are used to perform a series of fully 3-D aquifer pressure tests, and 3-D heterogeneity in aquifer parameters is estimated. For economy of space, we focus our review of prior published work only on fully 3-D studies of hydraulic tomography in which 3-D testing is performed and 3-D parameter distributions are estimated. A more comprehensive review including 1-D and 2-D HT applications has been performed previously by the authors and can be found in *Cardiff and Barrash* [2011]. 3-D HT investigations to date include both synthetic experiments and field experiments [*Yeh and Liu*, 2000; *Zhu and Yeh*, 2005; *Li et al.*, 2008; *Castagna and Bellin*, 2009; *Illman et al.*, 2009; *Bohling and Butler*, 2010; *Brauchler et al.*, 2011; *Berg and Illman*, 2011; *Cardiff and Barrash*, 2011; *Berg and Illman*, 2013; *Cardiff et al.*, 2012; *Schöniger et al.*, 2012; *Mao et al.*, 2013], and there has been a notable increase in frequency of 3-D HT applications particularly over the past 5 years with the increased availability of multicore and multiprocessor computing.

[6] While the computational overhead associated with 3-D HT data analysis has become less of an obstacle, application of 3-D HT in the field has been studied in only a few works [*Illman et al.*, 2009; *Brauchler et al.*, 2011; *Berg and Illman*, 2011; *Cardiff et al.*, 2012; *Berg and Illman*, 2013], and efforts to validate the results of field 3-D HT aquifer characterization have been mixed. *Illman et al.* [2009] used a transient model to analyze 35 response curves (choosing 218 total data points) from two pumping tests performed at the Mizunami Underground Research Site (Japan). Some sensors in this case were either too noisy to use, or did not respond to pumping (possibly due to sensor sensitivity limitations and/or the significant heterogeneity of this fractured rock system). Validation of the obtained heterogeneity esti-

mates was performed using qualitative comparisons with fault and lineament data, as well as through a qualitative comparison of prediction of 12 other available drawdown curves from the testing that were not inverted. Instead of pumping test data, *Brauchler et al.* [2011] inverted data from cross-well slug interference tests using an approximate "asymptotic" model of groundwater flow. Almost 400 source-receiver pairs were inverted by fitting the travel time and attenuation of the pressure response using an eikonal solver. The numerical model used for the 3-D inversion contained 600 grid cells and solved quickly ($\approx$1 min) due to the small scale of the numerical model and the fast eikonal solver used. Results were validated through qualitative comparison with prior site knowledge. *Berg and Illman* [2011] examined transient 3-D HT data from the North Campus Research Site (NCRS) at the University of Waterloo (Ontario, Canada) inverting data from four out of nine available pumping tests. About 160 pressure response curves were inverted, using a finite element model with about 30,000 elements. Using a 40-core computing cluster with a total of 192 GB of RAM, the computational demand for inverting the four pumping tests required "up to a week" of cluster computing time. Validations performed included prediction of responses from pumping tests that were not inverted (though these tests took place in the same wells as the inverted pumping tests), and qualitative comparison of *K* profiles against permeameter-obtained estimates. The same tests were later reanalyzed in *Berg and Illman* [2013] using a steady state numerical model. In *Cardiff et al.* [2012], data from a 3-D transient hydraulic tomography (3-D THT) field campaign were analyzed, consisting of 25 short-duration pumping tests from two different wells at the Boise Hydrogeophysical Research Site (BHRS) in Boise, ID, USA. However, again because of instrumentation issues and problems with pumping consistency, only 12 pumping tests were analyzed, and many transducer readings were eliminated due to sensors with significant noise or drift. The analysis in this work inverted about 250 drawdown curves and estimated hydraulic conductivity at over 100,000 grid cells. The inversion of all 12 tests was performed using six processor cores on a server with a total of 12 GB of RAM; inversion time excluding structural parameter optimization ranged from 48 to 72 h. The results of 3-D THT imaging in this work were validated via qualitative comparison with *K* profiles from slug testing.

[7] In this paper, we discuss a new field study of 3-D THT carried out during the Summer of 2011 at the Boise Hydrogeophysical Research Site (BHRS). This testing used temporarily emplaced equipment to obtain 3-D head change data; depth-discrete observations were implemented using packer-and-port strings in several fully penetrating wells within the test volume, while pumping took place at successive packed-off intervals in a central fully penetrating well. While similar to some of the prior studies listed above (particularly, the work of *Cardiff et al.* [2012]), the data analyzed in this study consist of a more complete set of pumping tests than those presented in studies to date, in the sense that pumping was performed at each 1 m interval throughout a single well, and high-quality data were obtained at every 1 m throughout five surrounding observation wells. In this work, we seek to more rigorously and quantitatively validate the results obtained
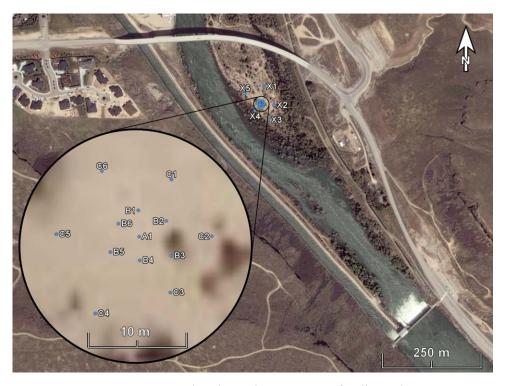
Figure 1. BHRS location and arrangement of wells on-site.

from a densely instrumented 3-D HT study, and to approximately quantify the loss in information that would occur if more sparse data collection is used. We accomplish this goal through two approaches. First, after attaining 3-D estimates of K throughout the site, we compare *for all 13 wells in the central area of the site K* profiles from HT analyses against K profiles estimated by partially penetrating slug testing. In the two prior works where K profiles from HT were compared against other K profiles for validation [*Cardiff et al.*, 2012; *Berg and Illman*, 2011], the comparison was only performed at wells that participated in the 3-D HT campaign (i.e., those wells that acted as pumping or observation locations). Second, we validate the predictive ability of K fields obtained from the inversions by simulating data from independent pumping tests that took place at other wells within the BHRS aquifer volume [specifically, the pumping tests from the 2010 field campaign analyzed in *Cardiff et al.*, 2012].

2. Field Site and Data Collection

[8] The BHRS is an uncontaminated hydrogeophysical field research site located on a gravel bar adjacent to the Boise River, roughly 15 km South-East from downtown Boise, ID, USA. The key infrastructure at the site is a set of 13 fully penetrating wells arranged in roughly concentric rings (Figure 1, A–C wells), surrounded by five boundary wells (Figure 1, X wells). The wells are fully screened through the cobble-and-sand aquifer, and the core-drive-drill emplacement method allowed natural collapse against well screens without an annular space or sand pack (see *Barrash et al.* [2006], for further information on well construction, and details about positive well skin). Stratigraphic units at the BHRS (Figure 2) have been defined

based initially on distributions of porosity estimated from neutron logs and grain-size characteristics from core [*Barrash and Clemo*, 2002; *Reboulet and Barrash*, 2003; *Barrash and Reboulet*, 2004], and similar structures have been recognized through analysis of ground-penetrating radar (GPR) [*Clement et al.*, 2006; *Clement and Barrash*, 2006; *Clement and Knoll*, 2006; *Irving et al.*, 2007; *Ernst et al.*, 2007; *Bradford et al.*, 2009; *Dafflon et al.*, 2011], seismic [*Moret et al.*, 2004, 2006], and capacitive conductivity [*Mwenifumbo et al.*, 2009] surveys.

[9] Several recent works have focused on the estimation of hydraulic conductivity heterogeneity at the BHRS using both traditional methods (partially penetrating slug tests) [*Cardiff et al.*, 2011; *Barrash and Cardiff*, 2013], and proof-of-concept 3-D THT methods [*Cardiff et al.*, 2012], with good correlation between these results. However, there is overall relatively poor or inconsistent correlation between K estimates and the porosity stratigraphy described above [*Barrash and Cardiff*, 2013]. Relative to other intensely monitored field research sites, the BHRS has relatively low to moderate heterogeneity; based on the slug test data set for the 13 central wells, as presented in *Barrash and Cardiff* [2013], the overall $\log_{10}(K)$ mean is -3.045 m/s (maximum is -1.80, minimum is -4.192) and $\log_{10}(K)$ variance is 0.093. For context, in Table 1 we compare the BHRS statistics to other well-known field sites, including those at which 3-D hydraulic tomography has been attempted (NCRS, GEMS, and Mizunami sites). Relative to the compared sites, the BHRS has overall relatively high $\log_{10}(K)$ mean and low to moderate $\log_{10}(K)$ variance, showing the greatest similarity to the GEMS site.

[10] The field testing analyzed in this work was designed to provide high-resolution coverage of aquifer response to successive, depth-discrete pumping tests throughout the
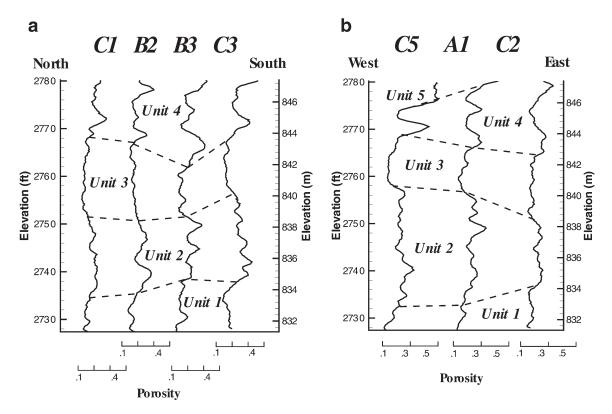
Figure 2. Porosity logs showing stratigraphic contacts between units at the BHRS that are recognized with porosity, lithology (core analysis), and geophysical methods. Unit 5 is a channel sand that pinches out in the center of the wellfield; Units 1–4 are cobble-and-sand units with lower porosity and porosity variance in Units 1 and 3, and higher porosity and porosity variance in Units 2 and 4. (a) Cross section roughly parallel to direction of river flow. (b) Cross section roughly West-to-East across site

aquifer thickness. The testing geometry consists of a series of pumping tests carried out at successive 1 m intervals in a central well (B1), with pressure responses observed at discrete depths using packer-and-port systems installed in five surrounding wells (B3, C3, C4, C5, and C6). Each packer-and-port string consisted of seven ≈1 m open intervals separated by a ≈1 m inflatable packer above and below. To obtain observations at successive 1 m intervals, we (1) performed all pumping tests with observational strings located in an "upper" configuration; then (2) we lowered all observational strings by ≈1 m to place observation intervals in locations formerly occupied by packers; and (3) repeated all pumping tests with observational strings in this "lower" configuration (see Table 2 for testing order and Figure 3 for test design geometry).

[11] A separate goal of the testing strategy was to provide a "baseline" data set for understanding what can be reasonably expected from 3-D HT performance when carried out under time and effort constraints by field practitioners. Since it is impractical for field investigators to implement many long-term pumping tests that reach or approximately reach steady state, we used a series of short-duration (15−20 min) partially penetrating (1 m interval) pumping tests and analyze transient response. This test duration was deemed appropriate for the unconfined BHRS aquifer based on the knowledge of average aquifer parameters and prior experience with the duration of pumping necessary to reach "late time" behavior. Overall, the pumping

tests carried out in this work required 5 days of field effort. Likewise, the testing arrangement, consisting of pumping and pressure observation equipment, was designed to investigate the aquifer at high resolution (1 m scale), while also minimizing the significant field effort that can be associated with equipment rearrangement.

[12] Compared to the 2010 3-D THT field campaign discussed in *Cardiff et al.* [2012], the testing analyzed here contained a few improvements. In terms of field hardware, pumping for all tests was carried out using a new, stable flow rate, in-well pump (Grundfos Redi-Flo3^{TM}), which allowed better test start-up and more consistent pumping

Table 1. Mean and Variance of $\log_{10}(K)$ Heterogeneity at Example Research Sites[a]

Site	$\log_{10}(K)$ (m/s) Mean	$\log_{10}(K)$(m/s) Variance
BHRS, Boise, ID [*Barrash and Cardiff*, 2013]	*−3.05*	*0.093*
NCRS, Waterloo, Ontario [*Berg and Illman*, 2011]	*−5.12*	*0.849*
GEMS, Lawrence, KS [*Bohling et al.*, 2010]	−2.82	0.108
Mizunami Research Site [*Illman et al.*, 2009]	*−6.93*	*0.377*
Borden Aquifer, Ontario [*Sudicky*, 1986]	−4.14	0.055
Cape Cod Site, Massachusetts [*Hess et al.*, 1992]	−3.45	0.026
MADE Site [*Rehfeldt et al.*, 1992]	−4.27	0.849

[a]Bolded sites are representative sites at which 3-D hydraulic tomography has been attempted.

Table 2. Summary of 3-D HT Pumping Tests Inverted

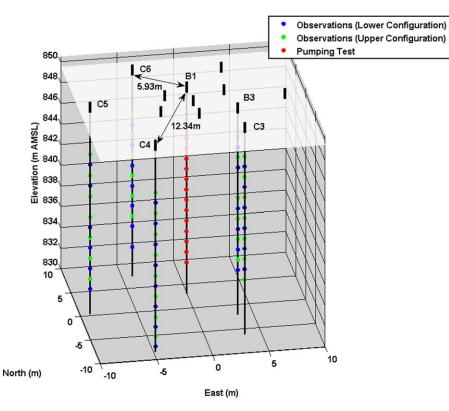| Test Name	Pumping Interval Elevation (m AMSL)	Observation Intervals Configuration	Number of Observation Locations Used (After Prescreening)	Avg. Pumping Rate (L/min)
081011 Test5	846	Upper	22	35.0
081111 Test1	845	Upper	27	25.6
081111 Test2	845	Upper	28	40.9
081111 Test3	844	Upper	27	36.0
081111 Test4	843	Upper	29	40.9
081111 Test5	842	Upper	30	40.6
081111 Test6	841	Upper	29	40.3
081511 Test1	844	Upper	27	36.4
081511 Test2	840	Upper	30	39.3
081511 Test3	839	Upper	31	37.4
081511 Test4	838	Upper	31	36.7
081511 Test5	837	Upper	32	37.2
081511 Test6	836	Upper	33	36.3
081511 Test7	835	Upper	33	36.0
081511 Test8	834	Upper	33	33.4
081511 Test9	833	Upper	33	32.3
081711 Test1	846	Lower	28	43.8
081711 Test2	845	Lower	28	42.5
081711 Test3	844	Lower	28	38.0
081711 Test4	843	Lower	29	42.3
081711 Test5	842	Lower	28	41.9
081711 Test6	841	Lower	28	42.0
081711 Test7	840	Lower	29	41.5
081711 Test8	839	Lower	29	39.6
081811 Test1	838	Lower	29	38.3
081811 Test2	837	Lower	29	38.6
081811 Test3	836	Lower	29	37.8
081811 Test4	835	Lower	29	37.4
081811 Test5	834	Lower	29	35.1
081811 Test6	833	Lower	29	34.4
Overall stats:	30 pumping tests, average time 15–20 min/test	1 m effective observation spacing	Up to 2628 observations inverted (drawdown at 3 times per obs. location)	37.9 L/min Avg. Pumping Rate

Figure 3. Pumping locations and observation locations during Summer 2011 BHRS 3-D THT testing.
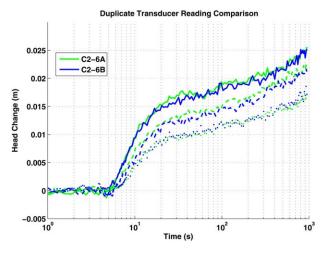
Figure 4. Head change curves obtained by two different transducers installed in the same observation interval, across several different pumping tests. Root mean square difference between transducer readings, in all cases, is less than 1 mm.

flow rates. Head changes were monitored using mainly the latest generation of small-diameter fiber-optic pressure transducers (FISO model FOP-MIV-NS-369D) located in observation wells C4, C5, and C6. These transducers, once stabilized for ambient water temperature, record pressure readings with errors that were verified to be as small as 1 mm water pressure (Figure 4). The remainder of the pressure change observations was recorded using standard strain-gage pressure transducers (Druck model POCR 1930–8388) located in observation wells B3 and C3. Data preprocessing and visualization in the field showed that, with few exceptions, data quality and pumping test quality were high, meaning that relatively few datapoints were removed during prescreening.

3. Data Analysis Strategy (Inversion)

[13] In order to convert the obtained data, i.e., head change curves from all pumping intervals and tests, into an image of aquifer K heterogeneity, we employ an inversion scheme developed in *Cardiff and Barrash* [2011] that uses (1) the standard, well-vetted MODFLOW groundwater flow model [*Harbaugh*, 2005] to simulate aquifer tests and act as a "forward model"; and (2) the Bayesian quasi-linear geostatistical algorithm of *Kitanidis* [1995] to solve the groundwater inverse problem. Below we briefly describe both of these components and address their efficacy as well as limitations.

[14] In our forward modeling (i.e., groundwater flow simulation), we utilize a modified version of MODFLOW-2005 developed in *Clemo* [2007] that integrates an "adjoint" process (ADJ) for calculating measurement sensitivities. As discussed in *Cardiff and Barrash* [2011], MODFLOW is a saturated flow model that is capable of simulating both confined and unconfined aquifer flow. However, in the case of unconfined flow, MODFLOW uses the instantaneous drainage assumption to simulate flow near the water table, meaning that suboptimal simulation results may be obtained from this model if used on slowly

draining systems. Based on prior analyses, the use of the instantaneous drainage assumption for the coarse-grained BHRS aquifer is appropriate when an S_y value representing "early time" drainage is employed [see *Cardiff et al.*, 2011, 2012].

[15] Inversion, in the context of aquifer imaging problems, is the process of finding reasonable heterogeneity distributions that are consistent with observed field data. To determine whether a given heterogeneity pattern is consistent with observed field data, a forward model is used to simulate the series of tests performed, and to produce synthetic measurements that are compared against their corresponding actual field measurements. In the Bayesian formulation, a parameter field's "consistency" with field data is determined by comparing the misfit (between synthetic measurements and real field data) against the expected magnitude of field measurement errors; parameter fields are tested for being "reasonable" by measuring their adherence to prior information. In the Bayesian geostatistical formulation developed by *Kitanidis* [1995], one minimizes the following objective function:

$$\min_{\mathbf{s}} NLAP(\mathbf{s}) = \frac{1}{2}(\mathbf{y} - \mathbf{h}(\mathbf{s}))^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{s}))$$
$$+ \frac{1}{2}(\mathbf{s} - \mathbf{X}\beta)^T \mathbf{Q}^{-1}(\mathbf{s} - \mathbf{X}\beta) \qquad (1)$$

where $\mathbf{y}$ is a ($n \times 1$) vector of field data, $\mathbf{s}$ is a ($m \times 1$) vector of values defining the heterogeneity pattern (for our case, K values at each node of a grid plus estimates of assumed-homogeneous S_s and S_y), and $\mathbf{h}()$ is the forward model $\mathbb{R}^m \rightarrow \mathbb{R}^n$ which converts a given heterogeneity model into a set of synthetic measurements. $\mathbf{R}$ is an expected data error covariance matrix ($n \times n$), representing the degree to which data misfit is expected. Similarly, $\mathbf{Q}$ is an expected spatial covariance matrix ($m \times m$) representing the spatial parameter variability that is expected. The final two terms, $\mathbf{X}\beta$, together represent the mean values expected throughout the aquifer volume, with $\mathbf{X}$ a ($m \times p$) known matrix defined to represent possible deterministic trends, and β a ($p \times 1$) vector of trend coefficients that are estimated. The objective function presented above is equivalent to maximizing the posterior probability of the heterogeneity given Gaussian measurement errors and a prior assumption of second-order stationary geostatistical parameter variability. Solution for the values of $\mathbf{s}$ and β, which is a nonlinear optimization problem, is performed using a linearization approach as discussed in numerous prior works [e.g., *Kitanidis*, 1995; *Cardiff and Barrash*, 2011; *Cardiff et al.*, 2012].

[16] Using the combination of forward modeling and inversion approaches discussed above, it is possible to perform 2-D or 3-D inversions of steady state or transient field data from either confined or unconfined aquifers. In *Cardiff and Barrash* [2011], the ability of this approach to estimate the spatial distribution of hydraulic conductivity (K), specific storage (S_s), and specific yield (S_y) was demonstrated, and for a large-scale problem with over 250 K unknowns required less than 3 days of computational time on a single multicore PC with 12 GB of RAM. However, through these numerical experiments it was also found that for reasonable ranges of variation in S_s and S_y (2 and 1 orders of

magnitude, respectively), assuming constant storage coefficients (with unknown values) does not highly impact K estimates obtained, but reduces inversion run-time.

[17] For the application in this work, data used in the inversion consist of three measurements chosen per drawdown curve, taken from the early, intermediate, and late-time response sections (at roughly 10, 90, and 550 s), similar to the analysis already presented in *Cardiff et al.* [2012]. The numerical models span a volume of 60 m × 60 m × 18 m, with maximum cell dimensions of 1 m × 1 m × 0.6 m and telescoping refinement near pumping locations. In MODFLOW, the model is oriented with its coordinate system roughly parallel/perpendicular to the Boise River (x/y, respectively). As prior information, we assume $\log_{10}(K)$ is a constant-mean random field with an exponential variogram, and that S_s and S_y are homogeneous values. As discussed above, our inversion routine is capable of estimating storage parameter variability in aquifer systems where such variation may be more significant and important, and we have performed inversions including storage parameter variability on the same computational hardware used in this work (albeit for a synthetic aquifer) [*Cardiff and Barrash*, 2011]. However, storage parameters for the BHRS aquifer (specific storage S_s and specific yield S_y) have not shown significant variability relative to K at this site (where the coarse sediments have virtually no silt or clay component). Similarly, as demonstrated in *Cardiff et al.* [2012], parameter field variances and correlation lengths can be estimated using a restricted maximum likelihood approach. However, for this work we assume these structural parameters are known in order to reduce computational burden, using parameter (K) covariance with horizontal and vertical correlation lengths of 10 and 2 m, respectively. These lengths are generally consistent with prior investigation results from the site [*Barrash and Clemo*, 2002; *Cardiff et al.*, 2011], and with structure dimensions observed in other high-energy fluvial deposits [e.g., *Jussel et al.*, 1994]. Error variances for measurements were assumed at 9×10^{-6} m^2 ($\sigma_y = 3$ mm), and variance for $\log_{10}(K[\text{m/s}])$ of 0.09 was assumed based on observed ranges of variability from prior inversions and other testing.

[18] The general process for the inversion is essentially a Gauss-Newton iteration with line search, a common gradient-based approach for nonlinear inverse problems [*Aster et al.*, 2005]:

[19] 1. To begin iteration, an initial guess is supplied consisting of a homogeneous K starting model using an appropriate "effective" value. Initial guesses for the assumed-homogeneous aquifer storage parameters are also supplied. These initial guesses are set as the current set of parameters ($\mathbf{s}_{curr}$), and the objective function $NLAP$ ($\mathbf{s}_{curr}$), determined by (1), is evaluated at this initial guess.

[20] 2. Using the adjoint sensitivity analysis, the Jacobian (i.e., a matrix representing the linear sensitivity of each observation to each parameter) is evaluated.

[21] 3. Using the quasi-linear geostatistical equations, which are equivalent to Gauss-Newton iteration, a new estimate of the parameters, $\mathbf{s}_{new}$, is obtained.

[22] 4. The objective function $NLAP$ ($\mathbf{s}_{try}$), where $\mathbf{s}_{try} = \mathbf{s}_{curr} + \alpha$ ($\mathbf{s}_{new} - \mathbf{s}_{curr}$), is evaluated at several values of α to find a suitable local decrease along the current search direction.

[23] 5. $\mathbf{s}_{curr}$ is set equal to the best $\mathbf{s}_{try}$ found (i.e., the line search result). Items 2 through 4 are then repeated until convergence.

[24] Convergence for our case was defined as obtaining a less than 2% change in any parameter value and a less than 1% change in the objective function value.

[25] We perform inversion for four different "Analysis Cases"—each of which uses all or a subset of the full set of field data, but represents progressive exclusion of data from the inversion—to aid in the examination of the incremental value of increased observational and/or pumping density for 3-D THT K resolution. Analysis Case 1 is an inversion of data from all pumping tests and all observation intervals (i.e., including both upper and lower observation configurations), resulting in an effective pumping and observation interval spacing of 1 m in the investigated portion of the aquifer. In Analysis Case 2, we invert all pumping tests with the observation well packers located in their "upper" configuration only, which increases the overall observation spacing to 2 m. Next, Analysis Case 3 eliminates several pumping tests from the analysis so that the effective spacing of both pumping and observation intervals is 2 m. Finally, Analysis Case 4 reduces the pumping test data set further to a set of pumping tests separated by 4 m intervals, while keeping 2 m spacing for observations. It may be noted that in Analysis Case 4, the number of tests inverted and the spacing of observation intervals is very similar to the field 3-D THT example presented in *Berg and Illman* [2011]. For all analysis cases, six processor cores on a high-end PC with 12 GB of RAM were used. Total computing time for each inversion was on the order of 2 days (Analysis Case 4) to 1 week (Analysis Case 1).

4. Results of Inversion

[26] In Figures 5–8, we show visualizations of the 3-D imaging results obtained along selected "slice-planes" between pumping and observation wells. Qualitatively, more detailed features are apparent in the Analysis Cases with more data inverted, though all four cases show similar overall features. If Analysis Case 1 is considered as a base case quantitatively, trends include (see Table 3): (1) roughly the same average $\log_{10}(K)$ value, but with (2) significantly decreasing variance (e.g., > 40% decrease in variance from Analysis Case 1 to Analysis Case (2) and (3) significantly increasing parameter root mean squared difference (RMSD) from the base case. In addition, if all 2011 tests are simulated using the parameter estimates from each case, a slight increase in data RMSE can be seen. These comparisons, especially the major reduction in parameter field variance, suggest that significant information about heterogeneity is lost when pumping/observational density is decreased beyond the scale of the aquifer heterogeneity geostatistical correlation lengths [see also, *Yeh and Liu*, 2000].

[27] In terms of comparison against existing K estimates, we show in Figure 9. 1-D profiles of the estimates obtained with slug testing [*Barrash and Cardiff*, 2013] against the estimates obtained with 3-D THT. Note that the profiles shown include both wells that are used during 3-D THT as pumping/observation wells, but also wells that were completely unused in 3-D THT testing. Overall there appears to be good correspondence between the major features
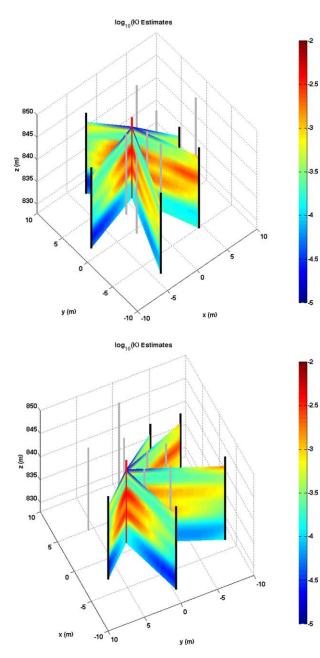
[28] While the issues with discrepancies near the surface are evident (in terms of negative correlations for two wells and quite low positive correlations (below 0.3) at four other wells, the correlations are much stronger when the top of the aquifer (i.e., the region above the highest observation zones in the 3-D THT testing) is excluded from analysis. For example, if only elevations below 845 m AMSL (which excludes the top ≈15% of the aquifer, i.e., the region above the top-most pressure sensors in these 3-D THT experiments), positive correlation is observed between slug test K estimates and 3-D THT K estimates at all wells, with significant correlations at 9 of the 13 wells. If elevations of 843 m AMSL and below are considered, 12 out of the 13 wells (C5 being the lone exception) in the central BHRS
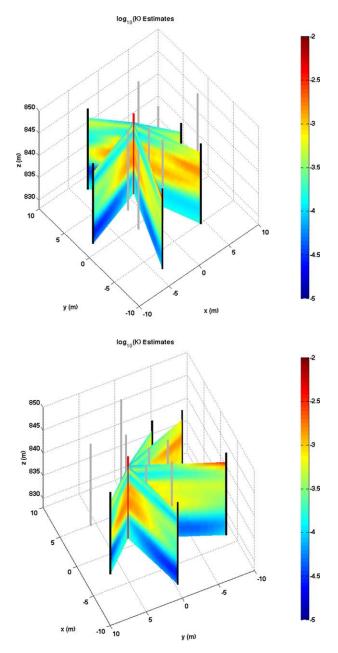

Figure 5. Analysis Case 1 results of inversion along well slice-planes, viewed from (top) south-west and (bottom) north-west.
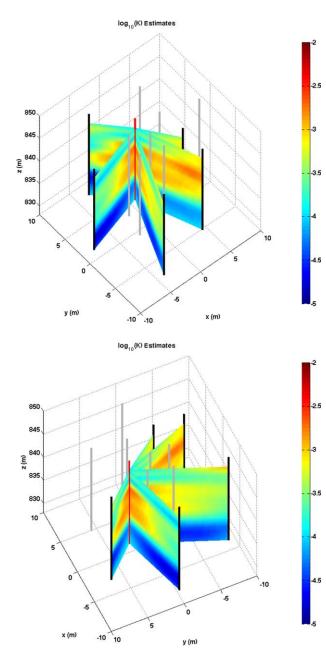
identified with these two methods, though it is notable that some discrepancies are present especially near the top (and to a lesser extent the bottom) of the aquifer, where slug K measurements were collected but 3-D THT observational coverage was missing or limited. A similar observation was seen by *Liu et al.* [2002] during experimentation in 2-D sandbox THT setups. A more detailed quantitative examination of the correlations between the K estimates generated through slug testing and the full inversion (Analysis Case 1) is shown on a well-by-well basis in Table 4. These calculations represent correlation between the 1-D K profiles produced by slug testing and comparable 1-D K profiles obtained via 3-D THT data inversion (i.e., as shown in Figure 9).

Figure 6. Analysis Case 2 results of inversion along well slice-planes, viewed from (top) south-west and (bottom) north-west.

Figure 7. Analysis Case 3 results of inversion along well slice-planes, viewed from (top) south-west and (bottom) north-west.

by-well correlation coefficients and overall (all well) correlation coefficients as data are successively removed from consideration. While all four cases maintain positive correlations at all wells, a general decrease in the number of statistically significant correlations is seen, from 12 out of 13 in Analysis Case 1, to only 8 out of 13 in Analysis Case 4 (likewise, generally, a decrease in correlation coefficient values across all wells is seen with decreasing observation/pumping density).

[30] Another quantitative examination of the performance of each Analysis Case is presented in Figure 11, which shows the ability of the inverted K field to accurately simulate pumping tests from the previous (2010) round of field experiments. Overall, all four Analysis Cases show
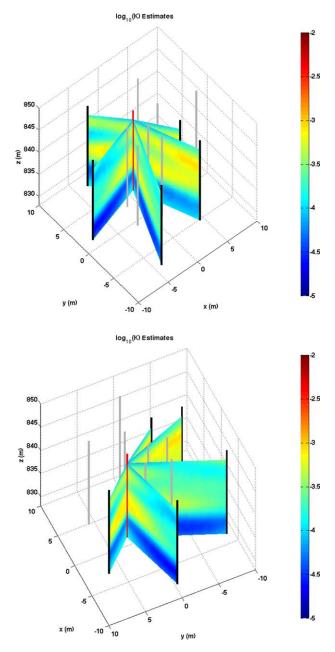
well area display moderate to strong, statistically significant correlations.

[29] In most wells, the top-most observation interval was centered around 844–845 m elevation, and thus it is not surprising that imaging results are less reliable above this elevation. Indeed, this is a common phenomenon seen in numerous tomographic approaches, e.g., GPR cross-well tomography. Considering the elevations 843 m AMSL and below—where coverage from pumping and observation intervals of 3-DTHT is consistently high, and where slug versus 3-D THT K estimate correlations are overall good to excellent—the four Analysis Cases provide an opportunity to examine the effect of reducing pumping or observation interval spacing. In Table 5, we show changes in the well-

Figure 8. Analysis Case 4 results of inversion along well slice-planes, viewed from (top) south-west and (bottom) north-west.

Table 3. Correlation Statistics Between Slug K Estimates and 3-D HT K Estimates at Well Profiles, Analysis Case 1[a]

Well	All Elevations		Elevations below 845 m AMSL		Elevations below 843 m AMSL	
	Correlation Coefficient	Significance Level	Correlation Coefficient	Significance Level	Correlation Coefficient	Significance Level
A1	−0.135	0.394	0.127	0.453	0.511	**0.003**
B1	0.437	**0.002**	0.743	**<0.001**	0.788	**<0.001**
B2	0.446	**0.002**	0.810	**<0.001**	0.846	**<0.001**
B3	0.250	0.093	0.684	**<0.001**	0.755	**<0.001**
B4	0.289	0.054	0.229	0.166	0.432	**0.014**
B5	0.523	**<0.001**	0.502	**0.002**	0.466	**0.008**
B6	0.679	**<0.001**	0.595	**<0.001**	0.606	**<0.001**
C1	−0.241	0.110	0.773	**<0.001**	0.706	**<0.001**
C2	0.520	**<0.001**	0.593	**<0.001**	0.590	**<0.001**
C3	0.566	**<0.001**	0.481	**0.002**	0.532	**0.001**
C4	0.177	0.223	0.268	0.086	0.435	**0.008**
C5	0.025	0.871	0.103	0.539	0.290	0.102
C6	0.817	**<0.001**	0.813	**<0.001**	0.789	**<0.001**

[a]Wells used for either pumping or observation are italicized, and statistically significant correlations are bolded.

relatively good, unbiased simulation of these independent pumping tests. However, using the densest set (Analysis Case 1) results in a substantial improvement in the sense that: (1) There is less simulation bias, with an overall calibration falling closer to the 1:1 perfect fit line; and (2) The RMSE of the simulations of independent pumping tests (Figure 11a, 3.14 mm) is reduced by almost 1 mm. The latter of these validations suggests that independent data misfits are the result of noise and modeling error, but that additional error due to lack of heterogeneity resolution is introduced as Analysis Cases with less dense data coverage are used.

[31] We briefly consider the K distribution in relation to established stratigraphy at the BHRS as noted above, and we refer to Figure 10, which shows two slice plots through the Analysis Case 1 K tomographic volume along with contacts between porosity stratigraphic units as an illustrative example. K structure is evident at two scales [*Barrash and Cardiff*, 2013]: a larger scale of three layers with higher K in the middle layer, and smaller scale with three lenses of relatively higher K within the middle layer that are recognizable in both plots. Similarities between porosity stratigraphy and K stratigraphy are limited to local coincidence of some porosity unit contacts with contrasts in relative K, and to a general coincidence of the contact between stratigraphic Units 1–2 and the break between the lower low-K unit and the middle higher-K unit. Such limited correspondence between $\log_{10}(K)$ and porosity has been noted previously [*Cardiff et al.*, 2011]; additional details are given in *Barrash and Cardiff* [2013].

5. Discussion

[32] The utility of 3-D HT for estimating aquifer heterogeneity has been a subject of some debate in the literature recently. This debate is perhaps best exemplified by quotes from two opposing works. In promoting hydraulic tomography, *Yeh and Lee* [2007] state that "...HT is merely an application of the concept of the CAT scan technology in medical sciences and tomographic surveys in geophysics to imaging subsurface hydraulic heterogeneity. This new way to collect and analyze data for aquifer characterization, we

are certain, will lead us to much [more] detailed subsurface characterization beyond the reach of traditional technologies." At the other end of the spectrum, *Bohling and Butler* [2010] discuss the "inherent limitations of hydraulic tomography" and state that "Given the expense and effort associated with performing such an extensive set of tests in the field, it is safe to say that no practically feasible number of tomographic pumping tests will ever produce anything approaching a unique estimate of the spatial distribution of aquifer hydraulic properties without incorporating other sources of data." Based on the current experimental results, we believe that the true value of hydraulic tomography lies somewhere between these two end-member opinions.

[33] The ability of HT to detect heterogeneities is dependent on the signal that can be measured, which is affected by both the overall contrast/variance in K values within the aquifer (which will determine the degree to which data reflect deviations from homogeneity) and the average K value (which will determine the overall magnitude of drawdowns that can be measured). That said, we believe the BHRS—a moderately heterogeneous, high conductivity aquifer—presents a relatively difficult case for HT analysis. Even given the difficulties of applying 3-D HT in this environment, we showed through comparison of 1-D K profiles that hydraulic tomography is capable of detecting the overall structure of subsurface deposits in 3-D. In quantifying the "expense and effort" that HT entails, it is important to note that the HT pumping tests discussed in this document required 5 days of field effort, while the slug test results they have been compared against required 30 days of effort, and non-negligible computational time to analyze. Overall, it is notable that in all four cases, K estimates at both 3-D THT observation wells and unused wells show good correlation with slug test K estimates. If the full aquifer thickness is examined, two wells (A1 and C1) show negative, but statistically insignificant, correlation. Below 845 m elevation, however, (the rough elevation at which the highest pressure transducers were located during experimentation) all wells show positive correlation across all Analysis Cases. As one note of caution, we point out that while it is tempting to treat the slug test K profiles as "true" values, they are in fact subject to significant uncertainty and possible biases
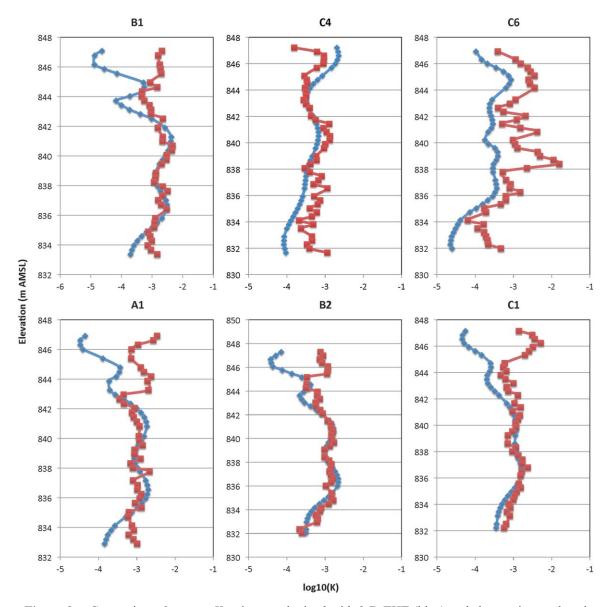
Figure 9. Comparisons between K estimates obtained with 3-D THT (blue) and slug testing analyzed using a skin value of 5e−4 m/s (red). (top) Three different wells used in 3-D THT testing, as pumping/ observational wells. (bottom) Three different wells not used in 3-D THT testing.

associated with assumptions that are made during data analysis— these include, among others, assumptions about wellbore skin or lack thereof, and assumptions or errors in the "effective radius" formulation used to scale wellbore inertial response [see discussion in *Cardiff et al.*, 2011; *Barrash and Cardiff*, 2013]. The relative magnitudes of the K values obtained through slug testing provide a useful cross validation for 3-D THT, though we caution that the lack of perfect correlation between these two estimates may be indicative of both errors and biases in slug test K estimates as well as errors and resolution constraints of 3- D THT imaging results.

[34] It has been noted in particular that sensitivity of 3-D HT data decreases away from vertical planes ("slice-planes") connecting pumping and observation wells. *Bohling and Butler* [2010] discussed the "lack of sensitivity to K variations outside [the vertical] plane" and pointed out— for an example with four coplanar wells— that many

images of heterogeneity consistent with data can be developed, with an especially large amount of uncertainty possible away from these slice-planes. By combining data from multiple *noncoplanar* wells—as done in this work— improved understanding of lateral variability can be gained. We note that, especially for Analysis Case 1, even wells not located on slice-planes show high correlations (e.g., wells C1, C2, and B2) to slug test K estimates. While indeed all imaging-type inverse problems are conceptually ill-posed, in practice our results show that the use of reasonable, geostatistically based prior information to regularize the inverse problem produces good maximum a posteriori estimates, measured in terms of high correlation with slug test estimates, even at distance from the vertical slice-planes.

[35] In addition to comparing our results with other estimates of the BHRS K field from other testing methods, another important measure of the 3-D THT imaging results
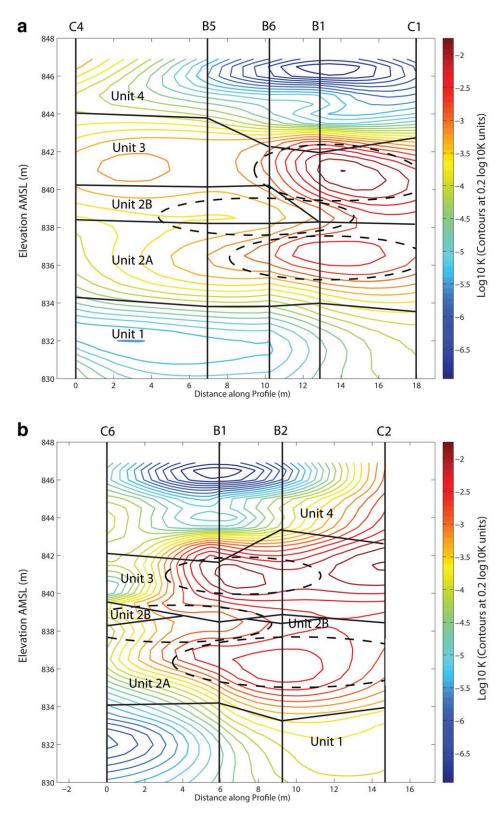
Figure 10. Two slice plots through the Analysis Case 1 tomographic volume showing $\log_{10}(K)$ distribution with $\log_{10}(K)$ contours at 0.2 intervals, and showing contacts between porosity-lithology-geophysical stratigraphic units for reference (see also Figure 2). (a) Approximately south-north plot through five wells. (b) Approximately west-east plot through four wells.

Table 4. Correlations Between Slug K Estimates and 3-D HT K Estimates Obtained Below 843 m[a]

Analysis case	1		2		3		4	
Pumping spacing	1 m		1 m		2 m		2 m	
Observation spacing	1 m		2 m		2 m		4 m	
Data points inverted	2478		1359		726		447	
Well[b]	Correlation Coefficient	Significance Level	Correlation Coefficient	Significance Level	Correlation Coefficient	Significance Level	Correlation Coefficient	Significance Level
A1	0.511	**0.003**	0.432	**0.013**	0.309	0.085	0.285	0.113
B1	0.788	**0.000**	0.654	**0.000**	0.671	**0.000**	0.594	**0.000**
B2	0.846	**0.000**	0.595	**0.000**	0.558	**0.001**	0.469	**0.007**
B3	0.755	**0.000**	0.776	**0.000**	0.797	**0.000**	0.816	**0.000**
B4	0.432	**0.014**	0.530	**0.002**	0.509	**0.003**	0.487	**0.005**
B5	0.466	**0.008**	0.505	**0.004**	0.462	**0.009**	0.413	**0.021**
B6	0.606	**0.001**	0.434	**0.024**	0.424	**0.028**	0.292	0.140
C1	0.706	**0.000**	0.385	**0.030**	0.283	0.116	0.167	0.361
C2	0.590	**0.000**	0.440	**0.010**	0.456	**0.008**	0.451	**0.008**
C3	0.532	**0.001**	0.526	**0.002**	0.526	**0.002**	0.553	**0.001**
C4	0.435	**0.008**	0.360	**0.031**	0.363	**0.029**	0.323	0.054
C5	0.290	0.102	0.184	0.306	0.221	0.217	0.296	0.094
C6	0.789	**0.000**	0.833	**0.000**	0.855	**0.000**	0.798	**0.000**
All Wells	0.473		0.414		0.444		0.349	

[a]Bolded cells signify correlation coefficients with significance above 95%.
[b]Italicized wells represent those wells that served as pumping or observation wells during the testing.

is its ability to produce accurate predictions under other aquifer stimulations. In particular, we showed how progressively including more data in the current inversion (using 2011 HT data) was able to improve predictions of the results of independent tests (2010 HT data). This again lends credence to the idea that HT data provide a useful source of information for improving predictions. Perhaps most impressively, our testing showed that the ability to predict independent tests was nearly as strong as the degree to which inverted tests were fit.

[36] While both of the lines of evidence above are encouraging, in no way does this mean that HT provides or will ever provide a "unique" estimate of the spatial distribution of heterogeneity. Rather, we believe the most useful perspective is to consider HT data as one component that can lead to continual refinement of aquifer understanding and continual improvement of predictions. Detailed HT studies, such as the one presented here, can help to reduce the feasible space of heterogeneity patterns, and especially in cases where geophysical methods do not provide useful information about K, HT may represent one of the few practical ways to reduce K uncertainty between wells. While the desired degree of imaging accuracy will be problem and site-specific, the Analysis Cases suggest that spac-

Table 5. Mean and Variance of $\log_{10}(K)$ Parameter Fields Across all Analysis Cases, and Root Mean Square Difference (RMSD) from Case 1

	Analysis Case			
	1	2	3	4
Pumping spacing	1 m	1 m	2 m	2 m
Observation spacing	1 m	2 m	2 m	4 m
Mean	−3.705	−3.620	−3.823	−3.686
Variance	0.174	0.098	0.089	0.068
RMSD from Case 1	—	0.171	0.207	0.283

ing in observation and pumping locations at distances comparable to heterogeneity correlation lengths (estimated, for this case, at 1–2 m in the vertical) is highly beneficial for obtaining statistically significant correlations (a result consistent with *Yeh and Liu* [2000]), as well as for providing predictive validity. This is important because it implies that inversion of only a few pumping tests, as presented in analyses to date [e.g., *Illman et al.*, 2009; *Berg and Illman*, 2011], may result in significant reductions in imaging resolution and accuracy (see, e.g., Tables 3 and 5). Likewise, this implies that forward models and inverse methods used to analyze useful HT field experiments will require the ability to handle both very large parameter spaces and large data sets. Indeed, further developments in improving hydraulic tomography will require careful data collection, clever methods for analyzing data, and advanced computational techniques.

[37] In terms of specific issues discovered in this study, two key questions arise. The first is why slug test K estimates and 3-D THT K estimates are poorly correlated at the top of the investigated volume, and the second is why poor correlation is observed at well C5, which was instrumented for observation. In regards to the former issue, there are several plausible hypotheses for this lack of correspondence, which will be investigated in the future. We believe the most likely possibility is that imaging above ≈845 m elevation is unreliable due to the fact that few observations were available above this elevation because of instrument positioning (a similar result to that obtained by *Liu et al.* [2002], in sandbox studies). This possibility is supported by the analysis of uncertainty for the imaging experiments, which can be derived through Bayesian geostatistical theory. As an example, plotting of the posterior standard deviation of K estimates (see Figure 12) shows that uncertainty increases especially at the bottom, but also at the top of the aquifer within the central measurement area. However, another hypothesis is that since
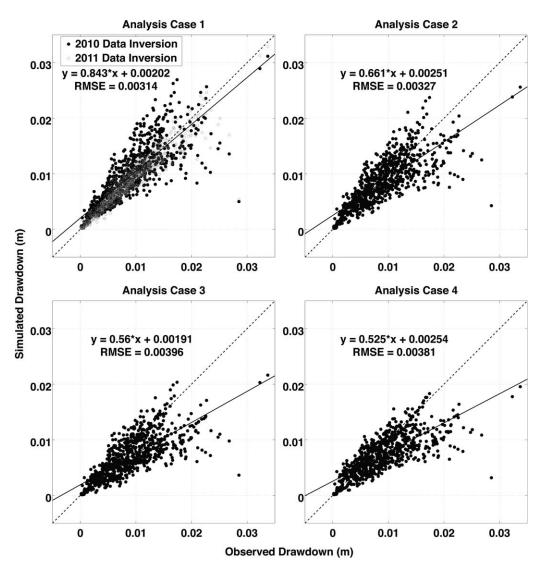
Figure 11. Results of simulating independent pumping tests using obtained *K* heterogeneity fields. Field data are from Summer 2010 pumping tests presented in *Cardiff et al.* [2012] (pumping from B4 and B5).

storage change effects are most prominent near the water table, the assumption of constant S_s and S_y values within our numerical model may manifest as "aliasing" of storage effects onto *K* values especially near the aquifer surface. Additionally, any inaccuracy in our numerical approximation of assuming instantaneous drainage may have some effect on the results. The question of correlations at well C5 is possibly more easy to address. While most wells on the site were instrumented with either recent-generation fiber-optic pressure transducers or reliable (but lower accuracy) strain-gage pressure transducers, well C5 was the one well instrumented entirely with a set of early prototype fiber-optic pressure transducers, simply due to instrumentation availability. These transducers were known to have lower reliability and higher "drift." We thus believe that the measurements at well C5 were perhaps the least reliable of those collected, which may be causing the lower correlations at well C5, and may also be reducing accuracy of imaging in other wells in the vicinity of its location.

6. Conclusions and Future Work

[38] In this work, we have presented a "baseline" study of 3-D hydraulic tomography in the field. The density of observations and pumping locations, and the quality of measurements for the tests analyzed in this work provide a useful base case for understanding the imaging resolution that can be obtained with this method, as well as the decreases in resolution that will occur if a testing regime is reduced in scope. The large set of data collected during our experiment and the analysis of different subsets of these data (Analysis Cases 1–4 above) provide a unique opportunity to examine the information content and imaging accuracy of 3- D THT. While the qualitative loss of resolution with decreasing data is a common feature of all inverse problems (and thus the results contained herein are not surprising), the degree to which decreasing 3-D HT data density reduces predictive ability has not been investigated thoroughly in the past.

[39] The work presented here provides an understanding of what can be expected from a relatively detailed,
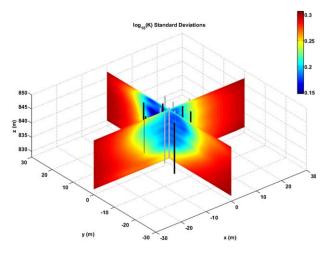
Figure 12. Uncertainty metrics (posterior standard deviation) for Analysis Case 1, along north-south and east-west slice-planes through aquifer volume. Note relatively greater uncertainty values obtained near top and bottom of aquifer in central testing area.

"single testing arrangement" HT investigation at a low to moderate-heterogeneity sedimentary aquifer. By including tests with different arrangements of pumping and observation wells, the overlapping volumes investigated by this type of 3-D THT analysis should further increase resolution of key features and their connectedness, and increase the spatial scales investigated. Another interesting possibility for investigation would be to attempt joint inversion of both 3-D THT data and slug testing data, which could help to improve imaging near wellbores while filling in details of connectedness with HT data between wells.

[40] As pointed out by *Bohling and Butler* [2010], data fusion approaches which use different data sources (e.g., hydrologic and geophysical) can provide a powerful method for reducing uncertainty in aquifer characterization, beyond what is possible with HT or other hydrologic methods. However, based on our finding of limited correspondence between stratigraphy and obtained K estimates, we caution that care must be taken in using geophysical data. If structural similarity is not present, at the given scale of investigation, between K fields and geophysically measured parameters such as resistivity or seismic velocity, then using such information to constrain hydraulic conductivity estimates could lead to erroneous results.

References

Anderson, M., and J. McCray (2011), Foreword: Lessons learned about contaminant hydrogeology from legacy research sites, *Ground Water*, *49*, 617–619.

Aster, R. C., B. Borchers, and C. H. Thurber (2005), *Parameter Estimation and Inverse Problems*, *Intl. Geophys. Ser.*, vol. 90, Elsevier Acad., Burlington, Mass.

Barrash, W., and M. Cardiff (2013), Hydraulic conductivity distribution from multi-level slug tests and multivariate facies associations in a conglomeratic fluvial aquifer, Boise Hydrogeophysical Research Site, *BSU CGISS Tech. Rep. 13-03*, 71 pp.

Barrash, W., and T. Clemo (2002), Hierarchical geostatistics and multifacies systems: Boise Hydrogeophysical Research Site, Boise, Idaho, *Water Resour. Res.*, *38*(10), 1196, doi:10.1029/2002WR001436.

Barrash, W., and E. C. Reboulet (2004), Significance of porosity for stratigraphy and textural composition in subsurface, coarse fluvial deposits: Boise Hydrogeophysical Research Site, *Geol. Soc. Am. Bull.*, *116*, 1059–1073.

Barrash, W., T. Clemo, J. J. Fox, and T. C. Johnson (2006), Field, laboratory, and modeling investigation of the skin effect at wells with slotted casing, Boise Hydrogeophysical Research Site, *J. Hydrol.*, *326*, 181–198.

Berg, S., and W. Illman (2011), Three-dimensional transient hydraulic tomography in a highly heterogeneous glaciofluvial aquifer-aquitard system, *Water Resour. Res.*, *47*, W10507, doi:10.1029/2011WR010616.

Berg, S. J., and W. A. Illman (2013), Field study of subsurface heterogeneity with steady-state hydraulic tomography, *Ground Water*, *51*, 29–40.

Bohling, G. C., and J. J. Butler (2010), Inherent limitations of hydraulic tomography, *Ground Water*, *48*, 809–824.

Bradford, J. H., W. Clement, and W. Barrash (2009), Estimating porosity with ground-penetrating radar reflection tomography: A controlled 3-D experiment at the Boise Hydrogeophysical Research Site, *Water Resour. Res.*, *45*, W00D26, doi:10.1029/2008WR006960.

Brauchler, R., R. Hu, P. Dietrich, and M. Sauter (2011), A field assessment of high-resolution aquifer characterization based on hydraulic travel time and hydraulic attenuation tomography, *Water Resour. Res.*, *47*, W10530, doi:10.1029/2011WR010698.

Cardiff, M., and W. Barrash (2011), 3-D transient hydraulic tomography in unconfined aquifers with fast drainage response, *Water Resour. Res.*, *47*, W12518, doi:10.1029/2010WR010367.

Cardiff, M., W. Barrash, B. Malama, and M. Thoma (2011), Information content of slug tests for estimating hydraulic properties in realistic, high-conductivity aquifer scenarios, *J. Hydrol.*, *403*, 66–82.

Cardiff, M., W. Barrash, and P. K. Kitanidis (2012), A field proof-of-concept of aquifer imaging using 3D transient hydraulic tomography with temporarily-emplaced equipment, *Water Resour. Res.*, *48*, W05531, doi:10.1029/2011WR011704.

Cardiff, M., T. Bakhos, P. K. Kitanidis, and W. Barrash (2013), Aquifer heterogeneity characterization with oscillatory pumping: Sensitivity analysis and imaging potential, *Water Resour. Res.,* *49*, 5395–5410, doi:10.1002/wrcr.20356.

Castagna, M., and A. Bellin (2009), A Bayesian approach for inversion of hydraulic tomographic data, *Water Resour. Res.*, *45*, W04410, doi:10.1029/2008WR007078.

Clement, W., and W. Barrash (2006), Crosshole radar tomography in a fluvial aquifer near Boise, Idaho, *J. Environ. Eng. Geophys.*, *11*, 171–184.

Clement, W. P., and M. D. Knoll (2006), Traveltime inversion of vertical radar profiles, *Geophysics*, *71*, K67–K76.

Clement, W. P., W. Barrash, and M. D. Knoll (2006), Reflectivity modeling of a ground-penetrating radar profile of a saturated fluvial formation, *Geophysics*, *71*, K59–K66.

Clemo, T. (2007) MODFLOW-2005 Ground Water model— User guide to the adjoint state based sensitivity process (ADJ), *SU CGISS Tech. Rep. 07-01*, 45 pp.

Dafflon, B., J. Irving, and W. Barrash (2011), Inversion of multiple intersecting high-resolution crosshole GPR profiles for hydrological characterization at the Boise Hydrogeophysical Research Site, *J. Appl. Geophys.*, *73*, 305–314.

Ernst, J., A. Green, H. Maurer, and K. Holliger (2007), Application of a new 2d time-domain full-waveform inversion scheme to crosshole radar data, *Geophysics*, *72*, J53.

Gottlieb, J., and P. Dietrich (1995), Identification of the permeability distribution in soil by hydraulic tomography, *Inverse Probl.*, *11*, 353–360.

Harbaugh, A. (2005), Modflow-2005, the U.S. Geological Survey modular ground-water model— The ground-water flow process, *U.S. Geol. Surv. Tech. Methods 6-A16*, 253.

Hess, K., S. Wolf, and M. Celia (1992), Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts. 3: Hydraulic

conductivity variability and calculated macrodispersivities, *Water Resour. Res.*, *28*, 2011–2027.

Illman, W., X. Liu, S. Takeuchi, T.-C. J. Yeh, K. Ando, and H. Saegusa (2009), Hydraulic tomography in fractured granite: Mizunami Underground Research Site, Japan, *Water Resour. Res.*, *45*, W01406, doi:10.1029/2007WR006715.

Irving, G. D., M. D. Knoll, and R. J. Knight (2007), Improving crosshole radar velocity tomograms: A new approach to incorporating high-angle traveltime data, *Geophysics*, *72*, J31–J41.

Jussel, P., F. Stauffer, and T. Dracos (1994), Transport modeling in heterogeneous aquifers. 1: Statistical description and numerical generation of gravel deposits, *Water Resour. Res.*, *30*, 1803–1817.

Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, *31*, 2411–2419.

Li, W., A. Englert, O. A. Cirpka, and H. Vereecken (2008), Three-dimensional geostatistical inversion of flowmeter and pumping test data, *Ground Water*, *46*, 193–201.

Liu, S., J. T.-C. Yeh, and R. Gardiner (2002), Effectiveness of hydraulic tomography: Sandbox experiments, *Water Resour. Res.*, *38*(4), 5-1–5-9, doi:10.1029/2001WR000338.

Mao, D., T. J. Yeh, L. Wan, J. Wen, W. Lu, C. Lee, and K. Hsu (2013), Joint interpretation of sequential pumping tests in unconfined aquifers, *Water Resour. Res.*, *49*, 1782–1796, doi:10.1002/wrcr.20129.

Moret, G. J. M., W. P. Clement, M. D. Knoll, and W. Barrash (2004), VSP traveltime inversion: Near-surface issues, *Geophysics*, *69*, 345–351.

Moret, G. J. M., M. D. Knoll, W. Barrash, and W. P. Clement (2006), Investigating the stratigraphy of an alluvial aquifer using crosswell seismic traveltime tomography, *Geophysics*, *71*, B63–B73.

Mwenifumbo, C., W. Barrash, and M. Knoll (2009), Capacitive conductivity logging and electrical stratigraphy in a high-resistivity aquifer, Boise Hydrogeophysical Research Site, *Geophysics*, *74*, E125–E133.

NRC (2005), *Contaminants in the Subsurface*, Natl. Acad. Press, Washington, D. C.

Reboulet, E. C., and W. Barrash (2003), Core, grain-size, and porosity data from the Boise Hydrogeophysical Research Site, Boise, Idaho, *BSU CGISS Tech. Rep. 03-02*, 84 pp.

Rehfeldt, K., J. Boggs, and L. Gelhar (1992), Field study of dispersion in a heterogeneous aquifer. 3: Geostatistical analysis of hydraulic conductivity, *Water Resour. Res.*, *28*, 3309–3324.

Schöniger, A., W. Nowak, and H.-J. H. Franssen (2012), Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography, *Water Resour. Res.*, *48*, W04502, doi:10.1029/2011WR010462.

Sudicky, E. (1986), A natural gradient experiment on solute transport in a sand aquifer: Spatial variability of hydraulic conductivity and its role in the dispersion process, *Water Resour. Res.*, *22*, 2069–2082.

Woodbury, A. D., and E. Sudicky (1991), The geostatistical characteristics of the Borden aquifer, *Water Resour. Res.*, *27*, 533–546.

Yeh, J. T.-C., and S. Liu (2000), Hydraulic tomography: Development of a new aquifer test method, *Water Resour. Res.*, *36*, 2095–2105.

Yeh, T.-C. J., and C.-H. Lee (2007), Time to change the way we collect and analyze data for aquifer characterization, *Ground Water*, *45*, 116–118.

Zhu, J., and J. T.-C. Yeh (2005), Characterization of aquifer heterogeneity using transient hydraulic tomography, *Water Resour. Res.*, *41*, W07028, doi:10.1029/2004WR003790.