Comparison of aquifer characterization approaches through steady state groundwater model validation: A controlled laboratory sandbox study

Walter A. Illman,[1] Junfeng Zhu,[2] Andrew J. Craig,[3] and Danting Yin[4]

Received 16 January 2009; revised 7 October 2009; accepted 28 October 2009; published 3 April 2010.

[1] Groundwater modeling has become a vital component to water supply and contaminant transport investigations. An important component of groundwater modeling under steady state conditions is selecting a representative hydraulic conductivity (K) estimate or set of estimates which defines the K field of the studied region. Currently, there are a number of characterization approaches to obtain K at various scales and in varying degrees of detail, but there is a paucity of information in terms of which characterization approach best predicts flow through aquifers or drawdowns caused by some drawdown inducing events. The main objective of this paper is to assess K estimates obtained by various approaches by predicting drawdowns from independent cross-hole pumping tests and total flow rates through a synthetic heterogeneous aquifer from flow-through tests. Specifically, we (1) characterize a synthetic heterogeneous aquifer built in the sandbox through various techniques (permeameter analyses of core samples, single-hole, cross-hole, and flow-through testing), (2) obtain mean K fields through traditional analysis of test data by treating the medium to be homogeneous, (3) obtain heterogeneous K fields through kriging and steady state hydraulic tomography, and (4) conduct forward simulations of 16 independent pumping tests and six flow-through tests using these homogeneous and heterogeneous K fields and comparing them to actual data. Results show that the mean K and heterogeneous K fields estimated through kriging of small-scale K data (core and single-hole tests) yield biased predictions of drawdowns and flow rates in this synthetic heterogeneous aquifer. In contrast, the heterogeneous K distribution or "K tomogram" estimated via steady state hydraulic tomography yields excellent predictions of drawdowns of pumping tests not used in the construction of the tomogram and very good estimates of total flow rates from the flow-through tests. These results suggest that steady state groundwater model validation is possible in this laboratory sandbox aquifer if the heterogeneous K distribution and forcing functions (boundary conditions and source/sink terms) are characterized sufficiently.

Citation: Illman, W. A., J. Zhu, A. J. Craig, and D. Yin (2010), Comparison of aquifer characterization approaches through steady state groundwater model validation: A controlled laboratory sandbox study, *Water Resour. Res.*, *46*, W04502, doi:10.1029/2009WR007745.

1. Introduction

[2] The characterization of subsurface heterogeneity of hydraulic parameters is of paramount importance to water supply assessment or protection and the remediation of groundwater throughout the world. Many communities rely on groundwater and in the case of Kitchener-Waterloo region of southern Ontario in Canada, over 80% of the water supply comes from groundwater. Unfortunately, in Walkerton, Ontario, Canada, over 2300 people have fallen ill and seven of them have passed due to the bacterial contamination of municipal wells [*O'Connor*, 2002]. Since this tragedy, a large effort has gone into delineating groundwater protection areas for water supply wells in Canada. The accurate delineation of such areas relies on detailed and reliable information on subsurface heterogeneities of hydraulic parameters such as hydraulic conductivity (K) and specific storage (S_s).

[3] A number of aquifer characterization approaches to delineate the heterogeneity in hydraulic parameters have been developed and tested over the last several decades [e.g., *Sudicky*, 1986; *Boggs et al.*, 1992; *Wolf et al.*, 1992; *Hess et al.*, 1992; *Sudicky et al.*, 2010]. The classical approach in delineating K heterogeneity is to collect a large number of

[1]Waterloo Institute for Groundwater Research, Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario, Canada.
[2]Kentucky Geological Survey, University of Kentucky, Lexington, Kentucky, USA.
[3]IIHR-Hydroscience and Engineering, University of Iowa, Iowa City, Iowa, USA.
[4]San Antonio, Texas, USA.

0043-1397/10/2009WR007745

core samples from multiple boreholes and to conduct grain size or permeameter analyses. Other approaches include the slug testing of large numbers of piezometers and/or monitoring wells [e.g., *Rehfeldt et al.*, 1992], flowmeter [e.g., *Hufschmeid*, 1986; *Molz et al.*, 1989; *Boman et al.*, 1997], steady state dipole flow [e.g., *Kabala*, 1993; *Zlotnik et al.*, 2001], or single-hole pumping or injection tests conducted at successive intervals in multiple boreholes to obtain a large number of point-scale K estimates, that are subsequently analyzed using geostatistical methods. More recently, geophysical methods [e.g., *Hyndman and Gorelick*, 1996; *Hubbard and Rubin*, 2000] and hydraulic tomography [e.g., *Neuman*, 1987; *Gottlieb and Dietrich*, 1995; *Yeh and Liu*, 2000; *Bohling et al.*, 2002; *Zhu and Yeh*, 2005, 2006] have been used to characterize subsurface heterogeneity. In particular, *Sudicky* [1986] obtained over 1279 K estimates from cores taken at the Borden site in Canada through a falling head permeameter and conducted geostatistical analysis to estimate the macrodispersion observed during a long-term tracer test at the site. He found good agreement between the macrodispersion coefficient determined from the long-term tracer test and those computed independently from stochastic theory using the statistical properties of K. This study was the first to validate the stochastic theory of *Gelhar and Axness* [1983] through the independent collection of K heterogeneity data and its application to predict macrodispersion of tracers at the Borden site.

[4] Several researchers have compared the validity of various characterization approaches in the field [e.g., *Lee et al.*, 1985; *Zlotnik and Zurbuchen*, 2003; *Butler*, 2005]. For example, *Butler* [2005] compared the K results from permeameter analysis of core samples, traditional slug tests, dipole flow tests, multilevel slug tests, borehole flowmeter tests, direct-push slug tests, pumping tests, and hydraulic tomography at the Geohydrologic Experimental and Monitoring Site (GEMS) in Kansas. He found that different characterization approaches yield reliable estimates of K estimates along the boreholes at the site. However, he found that most of the techniques examined cannot provide information on subsurface heterogeneity between wells.

[5] Recently, hydraulic tomography has been developed to obtain information on subsurface heterogeneity of K and S_s through sequential pumping tests. Hydraulic tomography is similar in concept to the Computerized Axial Tomography (CAT) scan technology, but the energy source is a change (decrease or increase) in hydraulic head achieved through water pumping or injection, and the sensors (pressure transducers) measure the propagation of head change at multiple locations throughout the aquifer. These hydraulic head data are then used to interpret the spatial distribution of K, S_s, connectivity of highly permeable zones, and to quantify uncertainties associated with the interpretations. The result is a highly resolved three-dimensional depiction of hydraulic parameters and their uncertainties.

[6] Various inverse methods have been developed for hydraulic tomography, which utilize pumping test data simultaneously or sequentially [e.g., *Gottlieb and Dietrich*, 1995; *Yeh and Liu*, 2000; *Bohling et al.*, 2002; *Brauchler et al.*, 2003; *Zhu and Yeh*, 2005, 2006; *Li et al.*, 2005; *Fienen et al.*, 2008; *Castagna and Bellin*, 2009; *Xiang et al.*, 2009]. Pneumatic tomography is similar in concept to hydraulic tomography, but the cross-hole tests are conducted with air in the unsaturated zone [*Illman and Neuman*, 2001, 2003] and interpreted using a stochastic inverse model [*Vesselinov et al.*, 2001a, 2001b] which interprets multiple tests simultaneously or sequentially.

[7] In the quest of quantitatively evaluating the performance of hydraulic tomography, various researchers [e.g., *Liu et al.*, 2002; *Illman et al.*, 2007, 2008; *Liu et al.*, 2007; *Yin and Illman*, 2009] have conducted laboratory scale studies in which the heterogeneity patterns can be prescribed and forcing functions controlled. In particular, *Illman et al.* [2007, 2008] showed that steady state hydraulic tomography [*Yeh and Liu*, 2000] is a promising technique in imaging subsurface heterogeneity patterns of K (from now on K tomograms) through laboratory sandbox studies. *Liu et al.* [2007] did the same, but used transient hydraulic tomography [*Zhu and Yeh*, 2005] to image K and S_s heterogeneity patterns (from now on K and S_s tomograms). Most recently, *Yin and Illman* [2009] showed that the hydraulic tomography based on the temporal moments of drawdown-recovery data [*Zhu and Yeh*, 2006] yield a satisfactory K tomogram but an unreliable S_s tomogram. In these previous studies, different methods and data collected at different scales were utilized to evaluate the tomograms. Among the various methods for tomogram evaluation, both *Illman et al.* [2007, 2008] and *Liu et al.* [2007] found that the best method was to simulate an independent cross-hole pumping test not used in the construction of the tomograms. In particular, *Illman et al.* [2007, 2008] were successful in predicting the steady state head response from a single cross-hole pumping test using the K tomogram from steady state hydraulic tomography. Likewise, *Liu et al.* [2007] were successful in predicting the transient head response of a single, independent cross-hole pumping test using the K and S_s tomograms from transient hydraulic tomography. These results were encouraging in that an independently conducted test can be predicted accurately if the heterogeneity pattern and forcing functions are all known. However, the experimental finding could have been construed to be weak and fortuitous because there was only one independent test from the experiments conducted by these authors previously.

[8] In the field, transient hydraulic tomography has also been applied in unconsolidated media [*Straface et al.*, 2007], while the hydraulic tomography based on the steady shape analysis [*Bohling et al.*, 2002] was reported for unconsolidated materials by *Bohling et al.* [2007]. More recently, *Li et al.* [2008] utilized their hydraulic tomography approach to image floodplain deposits. Likewise, *Cardiff et al.* [2009] conducted a 2-D depth-averaged hydraulic tomography analysis under steady state conditions in an unconfined aquifer. Most recently, *Illman et al.* [2009] used transient hydraulic tomography to image heterogeneity patterns of K and S_s in fractured granite at the kilometer scale, which appear to resemble connective fault zones and low K barriers. The numerical study by *Hao et al.* [2008] supports the notion that hydraulic tomography can be effective in imaging the connectivity of fractured rocks.

[9] There is a critical need to assess the capabilities of various characterization approaches in a controlled environment. There are a number of methods to assess these approaches, but we feel that this is accomplished best through the prediction of independently conducted drawdown inducing stresses induced by pumping tests or through

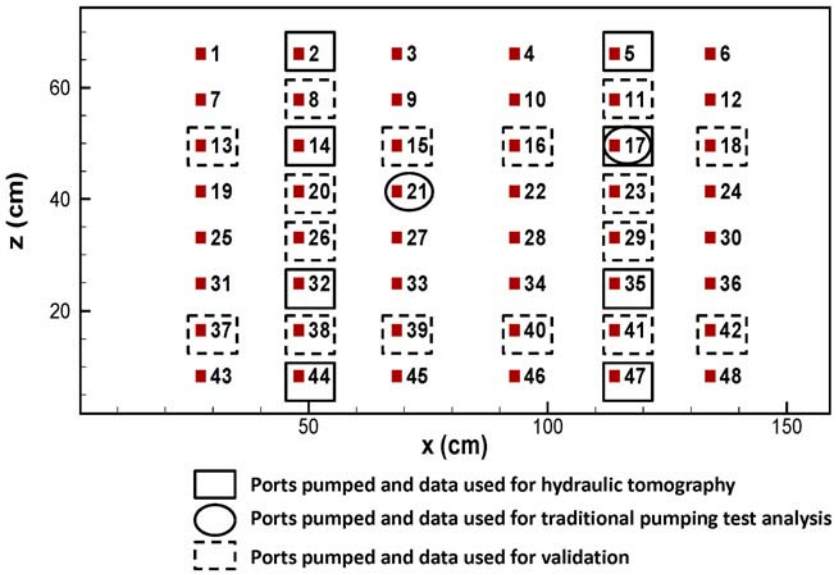

Figure 1. Schematic diagram of synthetic heterogeneous aquifer used for validation of steady state groundwater flow models. Numbers next to solid squares indicate port numbers, open squares around numbers indicate the eight ports (2, 5, 14, 17, 32, 35, 44, 47) used for hydraulic tomography, open ovals around numbers indicate the two ports (17 and 21) pumped and data analyzed using VSAFT2 by treating the medium to be homogeneous (i.e., traditional pumping test analysis), and the dashed open squares around the 16 other ports (8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42) indicate the pumping locations for the independent cross-hole pumping tests used for validation purposes.

the careful measurement of flow rates through the aquifer volume under question. The field assessment of various characterization techniques is of course most desirable, but field experiments can be affected by various external forcings, uncontrollable noise, and can be costly. In contrast, laboratory experiments can be efficient in assessing hydraulic tomography and other characterization approaches because both forcing functions and experimental errors can be controlled. The main objective of this paper is to assess K estimates obtained by various approaches by predicting independent cross-hole pumping tests and total flow rates from flow-through tests in a synthetic heterogeneous aquifer. Specifically, we (1) characterize the synthetic heterogeneous aquifer built in the sandbox through various techniques (permeameter analyses of core samples, single-hole pumping tests, cross-hole pumping tests, and flow-through tests), (2) obtain the mean K fields through traditional analysis of test data by treating the medium to be homogeneous, (3) obtain heterogeneous K fields through kriging and steady state hydraulic tomography, and (4) conduct forward simulations of 16 cross-hole pumping tests and six flow-though tests using these homogeneous and heterogeneous K fields and comparing them to actual drawdown and flow rate data from the laboratory experiments. The direct comparison of simulated and measured drawdowns as well as its statistical analysis yields a quantitative performance measure for each of these K fields. Moreover, the direct comparison of simulated and measured flow rates from the flow-through experiments allow for an additional test to quantify the performance of the various K fields. The comparison also amounts to the validation of various steady state groundwater models with different K fields obtained by various aquifer characterization methods.

2. Sandbox and Synthetic Heterogeneous Aquifer Construction

2.1. Sandbox Construction

[10] A synthetic heterogeneous aquifer was constructed in a sandbox to validate various fluid flow and solute transport algorithms. The sandbox is 193.0 cm in length, 82.6 cm in height, and has a depth of 10.2 cm. All materials used inside the sandbox are made of 316 stainless steel, brass, or Viton®. Forty eight ports, 1.3 cm in diameter, have been cut out of the stainless steel wall to allow coring of the aquifer as well as installation of horizontal wells. Each well was constructed by making six cuts spaced 1.46 cm apart in sections of brass tubing. The cuts were then covered with a stainless steel mesh that was bonded to the tubing with corrosion-resistant epoxy. Extreme care was taken to avoid the epoxy filling the mesh which could impede water flow. The wells fully penetrate the thickness of the synthetic aquifer. This allowed each location to be monitored by a pressure transducer, used as a pumping or an injection port and as a water sampling port. Figure 1 is a schematic drawing of the frontal view of the synthetic aquifer, showing the 48 port and pressure transducer locations.

[11] The flow system for the sandbox is driven by two constant head reservoirs, one at each end of the sandbox. The adjoining reservoirs are capable of supplying water throughout the length and thickness of the sandbox quickly and efficiently. In this sandbox, a series of 4-inch perforated

Table 1. Characteristics of Porous Media Used to Create a Synthetic Heterogeneous Aquifer

Sand Type	d_{50} (mm)	K (cm/s) Shepherd	K (cm/s) Darcy
16/30	0.872	1.32×10^{-1}	3.84×10^{-1}
20/30	0.750	1.03×10^{-1}	3.12×10^{-1}
20/40	0.578	6.68×10^{-2}	2.05×10^{-1}
#12	0.525	5.70×10^{-2}	2.05×10^{-1}
F32	0.504	5.33×10^{-2}	1.45×10^{-1}
#14	0.457	4.53×10^{-2}	1.21×10^{-1}
4030	0.355	2.99×10^{-2}	5.79×10^{-2}
F55	0.242	1.59×10^{-2}	2.80×10^{-2}
F65	0.204	1.20×10^{-2}	1.83×10^{-2}
F75	0.174	9.22×10^{-3}	1.73×10^{-2}
F85	0.151	7.29×10^{-3}	1.35×10^{-2}

plate/mesh combination was installed at each end of the sandbox to serve as a porous media/water interface and provide hydraulic control. The boundary head levels can be easily adjusted to be equal or to create a desired hydraulic gradient. The developed system is capable of maintaining three constant head boundaries simultaneously by ponding water at the top in addition to fixing the hydraulic heads in the two constant head reservoirs.

[12] The data acquisition system used for the laboratory experiments consisted of three major components. Pressure measurements were made with 50 *Setra* model 209 gage pressure transducers with a range of 0 to 1 psi, 48 of which measured hydraulic head in the aquifer and one in each constant head reservoir. These pressure transducers were installed at each of the 48 data acquisition ports in the stainless steel wall of the sandbox. The second component was a 64-channel data acquisition board from *National Instruments*. A hub that separates excitation and output currents for the transducers was also assembled. The third component was a dedicated PC with *National Instruments LabVIEW* software for automated data acquisition. Further details of the sandbox construction can be found in the work of *Craig* [2005].

2.2. Synthetic Aquifer Construction Through Cyclic Flux of Sediment-Laden Water

[13] Sandbox experiments have been conducted for a variety of purposes using different methods for aquifer construction. Early studies utilized uniform packing of sands to create a homogeneous medium and uniformly heterogeneous packing [e.g., *Silliman and Simpson*, 1987; *Schincariol and Schwartz*, 1990; *Illangasekare et al.*, 1995]. More recently, complex heterogeneity patterns have been packed by various researchers [e.g., *Welty and Elsner*, 1997; *Silliman et al.*, 1998; *Chao et al.*, 2000; *Barth et al.*, 2001; *Silliman*, 2001; *Danquigny et al.*, 2004; *Fernàndez-Garcia et al.*, 2005] in two- and three-dimensional small- and intermediate-scale sand tanks to mimic the more complex heterogeneity patterns with statistical properties including correlation lengths that are representative to those found in natural geologic media. The process involves generating a correlated K field using a random field simulator and carefully packing the resulting K distribution with different size sand blocks. *Jose et al.* [2004] created more natural deposits consisting of lenses rather than blocks through a settling procedure in standing water. This approach produced small-scale heterogeneities such as laminations, cross lamination, and rippled like structures that are similar to those found in natural fluvial deposits.

[14] In this study, the synthetic heterogeneous aquifer was created through the cyclic deposition of sediments under varying water flow and feed rates of sediments. Previous sandbox aquifer studies by our group [*Illman et al.*, 2007, 2008; *Liu et al.*, 2007; *Yin and Illman*, 2009] involved careful packing of heterogeneity patterns in a prescribed fashion. Packing of sand bodies by hand is a laborious procedure and complex heterogeneity patterns are hard to create. Our goal in relying on sediment transport was to create a more realistic heterogeneity pattern with various scales of heterogeneity in an efficient manner. The one drawback of this approach is that the heterogeneity pattern in the sandbox cannot be controlled precisely as in the case of packing a sandbox aquifer by hand.

[15] Various commercially available sands of variable grain sizes (Table 1) (Unimin Corporation and U.S. Silica) were used to create the synthetic aquifer. Examination of the grain size distribution (not shown here, but available from *Craig* [2005]) shows that the sands are well sorted and the d_{50}, which is the particle diameter for which 50% of the weight is finer, ranges between 0.151 mm to 0.872 mm. On the basis of *Shepherd*'s [1989] empirical power law relationship relating the d_{50} to the K, the estimated K for each class of sands ranges from 7.29×10^{-3} to 1.32×10^{-1} cm/s from the smallest to the largest d_{50}. We also estimated K for each sand type using a constant head permeameter yielding a range of 1.35×10^{-2} to 3.84×10^{-1} cm/s from the smallest to the largest d_{50} showing that the K estimates are in general consistent, but differences do arise from the two approaches.

[16] The sediment layers were deposited using a sediment suspension tank that lies in series with the sandbox and a 1/3 horsepower pump (Figure 2). The pump was operational in one direction only. Therefore the plumbing was designed to accommodate flow in either direction through the sandbox. The design of how the different sands of varying grain sizes were deposited heavily relied on flow rates into the sandbox from the suspension tank. The flow rates were determined for each sand type empirically by determining the mean velocities for under which a particular sediment type is transported, deposited, and scoured within the sandbox [*Craig*, 2005]. For each layer, a specific sediment type was chosen and introduced into the suspension tank (Figure 2). To introduce heterogeneity among the layers, care was taken to introduce sediments with variable d_{50} from one layer to the next. Deposition of each layer was accomplished by adjusting the flow rate of the sediment-laden water into the sandbox. Under uniform flow conditions, the coarser particles appeared to drop off first and progressively finer particles deposited on top of the coarser ones creating small-scale heterogeneity within each deposited layer. To create heterogeneity at the larger scale and interfingering of different sediment layers, the flow of water into the tank was alternated after the deposition of each layer. In total, 17 layers were deposited to mimic an interfingering fluvial deposit with the lowest deposit in the sandbox designated as layer 1 with additional deposits overlying labeled in an ascending numerical order. Table 2 summarizes the layers deposited, its sand type, and the volume of sand deposited for a particular layer. Because the deposition of the final layer was not possible as the height of the synthetic aquifer exceeded the top most orifice in the sandbox used for sediment transport,

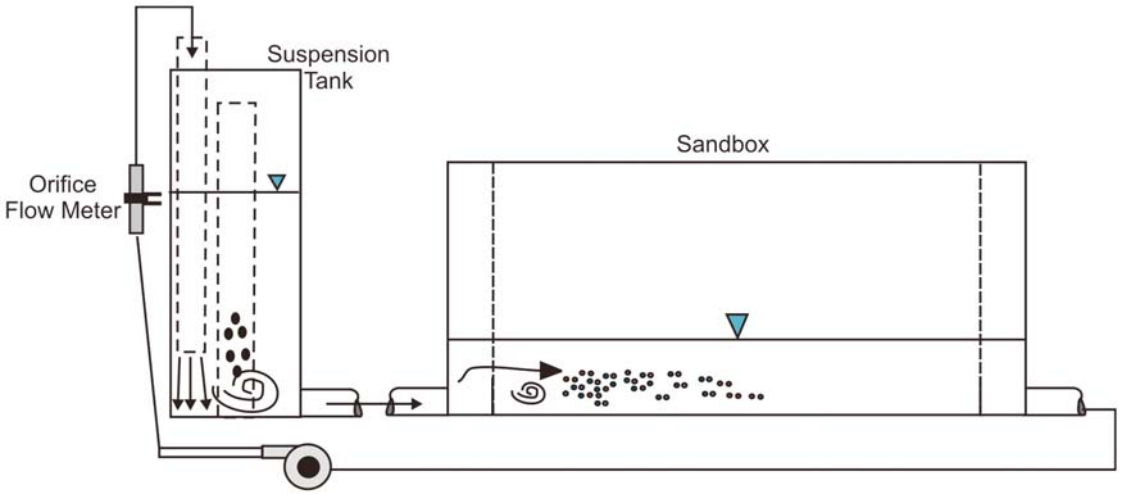

Figure 2. Schematic diagram of the sediment transport system for the creation of the synthetic heterogeneous aquifer.

20/30 sand was placed over layer 17 to fill the tank to the top. Figure 3 is a photograph of the sandbox showing each of the deposited layers. Further details to the creation of the synthetic heterogeneous aquifer are summarized in the work of *Craig* [2005].

3. Characterization of the Synthetic Heterogeneous Aquifer Treating the Medium to be Homogeneous

[17] The characterization of the synthetic heterogeneous aquifer was accomplished by various established aquifer characterization methods and with hydraulic tomography. Our main goal was to test the ability of these approaches to characterize the aquifer at different scales. Below, we describe the various characterization approaches and report the geometric mean of K values (K_G) determined from each test type as well as corresponding variance ($\sigma^2_{\ln K}$) estimates in Table 3. The K_G and $\sigma^2_{\ln K}$ were computed using

$$K_G = \exp\left(\frac{1}{n}\sum_{i=1}^{n} \ln K_i\right) \tag{1}$$

$$\sigma^2_{\ln K} = \frac{\sum_{i=1}^{n}\left(\ln K_i - \overline{\ln K}\right)^2}{(n-1)} \tag{2}$$

where n is the sample size, i is the index number, ln K is the natural logarithm of K, and $\overline{\ln K}$ is the arithmetic mean of ln K values.

Table 2. Sand Type and Deposited Volume for Each Layer in the Sandbox Aquifer

Layer	Sand Type[a]	Volume (m^3)
1	20/30	0.0051
2	4030	0.0034
3	F85	0.0031
4	20/40	0.0051
5	Mix	0.0034
6	Mix	0.0034
7	#12	0.0043
8	F32	0.0034
9	20/40	0.0031
10	F65	0.0043
11	#12	0.0019
12	16/30	0.0024
13	20/30	0.0051
14	F-75	0.0043
15	20/40	0.0043
16	Mix	0.0043
17	F-85	0.0043
18	20/30	0.0051

[a]Layers labeled "mix" consisted of equal volumes of #14, F75, and 16/30 sands.

3.1. Constant Head Permeameter Analysis of Core Samples

[18] Core samples from the synthetic aquifer were obtained by inserting a horizontal core tube through the 48 wells. The extracted cores had dimensions of 1.28 cm in diameter and 10.16 cm in length. After the collection of the core, a cotton cloth with permeability much higher than the core samples was placed on each end and prepared for their analysis in a constant head permeameter [*Klute and Dirksen*, 1986]. Because the extracted core was very small, a special

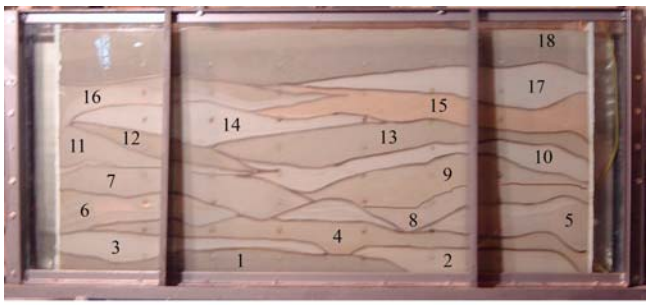

Figure 3. Photograph of synthetic heterogeneous aquifer created via cyclic flux of sediment-laden water.

Table 3. Summary of K Estimates From Various Approaches

Test Type	n	Min K (cm/s)	Max K (cm/s)	K_G (cm/s)	$\sigma^2_{\ln K}$
Core	48	0.0111	0.298	0.077	0.868
Single-hole	48	0.014	0.320	0.059	0.384
Cross-hole (Port 17)	48	0.021	0.315	0.076	0.301
Cross-hole (Port 21)	48	0.054	0.420	0.111	0.224
Flow-through	6	0.131	0.141	0.136	0.001

permeameter was built for testing purposes. Inspection of cored sands revealed that sands are quite uniform; thus we do not expect significant anisotropy in K measurements. Details to the core extraction method and the design of the constant head permeameter are provided in the work of *Craig* [2005].

3.2. Single-Hole Pumping Tests

[19] We then conducted single-hole pumping tests at each of the 48 ports. The tests were conducted by pumping water at each port and monitoring the transient head change within the pumped well. A constant pumping rate (Q = 1.25 cm^3/s) was set for each single-hole pumping test. For each test, data collection started without the pump running in order to obtain the initial hydraulic head in the sandbox at all measurement ports. A peristaltic pump was then activated at the pumping port and allowed to run at a constant rate until the development of steady state flow conditions. The pump was then shut off to collect recovery data. The steady state head data were analyzed using VSAFT2 [*Yeh et al.*, 1993] through manual calibration by treating the aquifer to be homogeneous. Details to the numerical modeling and calibration effort are provided by *Craig* [2005].

3.3. Cross-Hole Pumping Tests

[20] We then conducted cross-hole pumping tests at each port along columns 2 (ports 2, 8, 14, 20, 26, 32, 38, and 44) and 5 (ports 5, 11, 17, 23, 29, 35, 41, and 47) and nine additional pumping tests at various ports outside of these two columns (ports 13, 15, 16, 18, 21, 37, 39, 40, and 42) (see Figure 1). The cross-hole tests were conducted by pumping at rates ranging from 2.50–3.17 cm^3/s at 25 separate ports indicated by open and dashed squares on Figure 1. Prior to each cross-hole pumping test, all pressure transducers were calibrated to ensure accurate data collection. We next collected hydraulic head data for several minutes in all pressure transducers to establish a static, initial condition. After establishment of static conditions, we pumped from each port using a peristaltic pump, while taking head measurements in all 48 ports and a pressure transducer in each of the constant head boundaries. For each test, pumping continued until the development of steady state conditions, which was determined by observing the stabilization of all head measurements within the aquifer. The pump was then shut off to collect recovery head data until its full recovery. We note that we only utilized steady state head data for subsequent analyses to estimate K by treating the medium to be homogeneous and through steady state hydraulic tomography to estimate a K tomogram.

[21] An estimate of K obtained between the pumped and observation intervals when the medium is treated to be homogeneous is considered to be an equivalent hydraulic conductivity (K_{eq}) [*Renard and de Marsily*, 1997; *Neuman*, 2005]. To obtain an estimate of K_{eq} at each of the observation ports, we selected two cross-hole pumping tests among the 25 tests. One pumping test took place in port 17 and the other in port 21. Both ports were completed in layer 13. The core samples from ports 17 and 21 yielded K values of 1.64×10^{-2} and 2.46×10^{-1} cm/s, respectively.

[22] The observation head data from the two tests were analyzed by manually calibrating VSAFT 2 and assuming the aquifer is homogeneous. The numerical setup for the calibration was identical to the single-hole test analysis. Analysis of the two pumping tests yielded 96 estimates of K for the equivalent homogeneous medium (K_{eq}). Values of K_{eq} from each cross-hole pumping test were then averaged using equation (1) to estimate the K_G for the entire synthetic aquifer.

3.4. Bidirectional Flow-Through Tests

[23] We also conducted six bidirectional flow-through tests through the entire synthetic aquifer to obtain a large-scale K estimate under steady state flow conditions. Specifically, each of these six tests was conducted by changing the height of the constant head reservoirs on both sides of the sandbox. After the flow reached a steady state condition, we measured discharge from one side of the sandbox. We also measured the difference between the heights of the water column in the two constant head reservoirs to determine the hydraulic gradient. We then applied Darcy's law to obtain the effective hydraulic conductivity (K_{eff}) of the synthetic aquifer. This estimate is considered to be the K_{eff} value of the synthetic aquifer as flow takes place throughout the entire synthetic heterogeneous aquifer, the medium is treated to be homogeneous, and is independent of boundary conditions [*Renard and de Marsily*, 1997]. An average of the six different K_{eff} estimates was taken by applying equation (1) to obtain the K_G.

[24] The six flow-through tests were conducted by alternating the flow direction yielding three tests with flow taking place from the left to right and three additional tests with flow taking place from the right to left through the synthetic heterogeneous aquifer. In previous efforts [*Illman et al.*, 2007, 2008; *Liu et al.*, 2007] flow-through tests were conducted in only one direction using a different sandbox aquifer with a prescribed heterogeneity pattern.

3.5. Statistical Analysis of Test Data

[25] Table 3 provides a summary of key statistics from the various data collected in this sandbox. Results show that the K_G from various characterization approaches ranges from 0.059 to 0.136 cm/s, while the $\sigma^2_{\ln K}$ ranges from 0.001 to 0.868. Each test type yields a different K_G and $\sigma^2_{\ln K}$ indicating the uncertain nature of estimating these parameters in a heterogeneous aquifer. It is of interest to note that the K_G from the two cross-hole tests depends on the pumping location in this sandbox. A similar observation was noted by *Wu et al.* [2005] who used numerical simulations to study the behavior of effective hydraulic parameters during pumping tests, while *Illman and Neuman* [2001, 2003], *Vesselinov et al.* [2001a], and *Illman* [2006] observed a similar behavior, in which large-scale permeability esti-

Table 4. Geostatistical Model Parameters for Kriging Core and Single-Hole ln K Data

Data	Model	Nugget	Range (cm)	Sill	Anisotropy Ratio
Core	Exponential	0	17	0.90	2.2
Single-hole	Exponential	0	16	0.45	2.0

mates varied spatially at a field site consisting of unsaturated fractured tuffaceous rocks.

4. Geostatistical Analysis of Small-Scale K Estimates

[26] Geostatistical analysis of core and single-hole K data were conducted using the Surfer 8 software developed by Golden Software, Inc (www.goldensoftware.com). All 48 data were used for the kriging of core and single-hole data. We developed a grid that is compatible with the steady state hydraulic tomography analysis described later so that the estimated K fields from both approaches could be compared. The exponential variogram model was used to fit the experimental variograms in both horizontal and vertical directions, resulting in an anisotropic variogram model which can be readily incorporated into the Kriging function in Surfer 8 for interpolation.

[27] Table 4 lists the variogram parameters fit to the experimental variograms. Figures 4a and 4b show the kriged K fields derived from the core and single-hole K data, respectively. The results, in general, reveal smoother K fields in comparison to the interfingering layers shown in Figure 3 which is expected considering that there are only 48 data points used for kriging.

5. Steady State Hydraulic Tomography Analysis of Selected Cross-Hole Pumping Tests

5.1. Inverse Model Description

[28] The steady state hydraulic tomography analysis of cross-hole pumping tests in the sandbox was conducted using a sequential geostatistical inverse approach developed by *Yeh and Liu* [2000]. We only provide a brief description of the inversion approach here. The inverse model assumes a steady flow field and the natural logarithm of K (ln K) is treated as a stationary stochastic process. The model additionally assumes that the mean and correlation structure of the K field is known a priori. The algorithm essentially is composed of two parts. First, the Successive Linear Estimator (SLE) is employed for each cross-hole test. The estimator begins by cokriging the initial estimate of K_{eff} and observed heads collected in one pumping test during the tomographic sequence to create a cokriged, mean removed ln K (f, i.e., perturbation of ln K) map.

[29] Cokriging does not take full advantage of the observed head values because it assumes a linear relationship [*Yeh and Liu*, 2000] between head and K, while the true relationship is nonlinear. To circumvent this problem, a linear estimator based on the differences between the simulated and observed head values is successively employed to improve the estimate.

[30] The second step of *Yeh and Liu*'s [2000] approach is to use the hydraulic head data sets sequentially instead of including them simultaneously in the inverse model thus the term, "sequential successive linear estimator" (SSLE) is used to describe the inverse algorithm hereon. In essence, the sequential approach uses the estimated K field and covariances, conditioned on previous sets of head measurements as prior information for the next estimation based on a new set of pumping data. This process continues until all the data sets are fully utilized. Modifications made to the code for this study include its ability to account for variations in the boundary conditions with each pumping test as they are sequentially included and implementing the modified loop scheme described by *Zhu and Yeh* [2005].

[31] To obtain a K tomogram from the available cross-hole pumping tests, we solve an inverse problem for steady state flow conditions. The synthetic aquifer was discretized into 741 elements and 1600 nodes with element dimensions of 4.1 cm × 10.2 cm × 4.1 cm. A similar grid was also used previously for steady state hydraulic tomography analysis of a different sandbox aquifer by *Illman et al.* [2007, 2008], transient hydraulic tomography by *Liu et al.* [2007], and hydraulic tomography based on the temporal moments of drawdown-recovery data by *Yin and Illman* [2009].

[32] *Illman et al.* [2007] found that the variations in constant head boundary conditions from one test to the next can be critical in conducting a steady state hydraulic tomography survey. Therefore, we vary the boundary head values from one test to the next as they are included into the inverse model.

5.2. Input Parameters and Cross-Hole Tests Used

[33] Input data to the inverse model include initial guesses for the K_{eff}, $\sigma^2_{\ln K}$ and the correlation scales (λ_x, λ_y, and λ_z), discharge rate (Q) from each pumping test, and steady state heads at constant boundaries, as well as available point (small-scale, i.e., core, slug, and single-hole tests) measurements of K. In all previous studies of laboratory sandbox evaluations of hydraulic tomography using a different sandbox aquifer [*Illman et al.*, 2007, 2008; *Liu et al.*, 2007; *Yin and Illman*, 2009], we excluded small-scale data for the

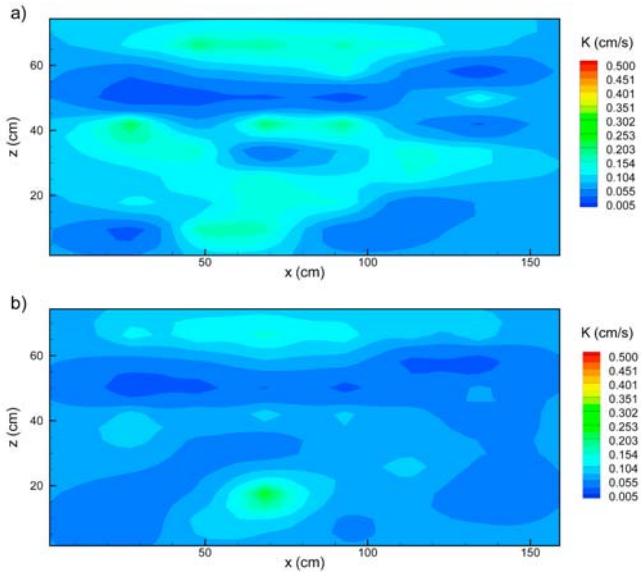

Figure 4. K distributions obtained via kriging of (a) core K estimates and (b) single-hole K estimates.

inverse modeling effort to test the algorithms without conditioning. For this study, we also exclude core and single-hole K estimates and conduct the steady state hydraulic tomography analysis without conditioning.

[34] We obtained the initial estimate of K_{eff} by taking the geometric mean of all K_{eq} estimates obtained through the analysis of data from observation ports during the cross-hole pumping tests by treating the heterogeneous medium to be homogeneous. This is a reasonable estimate of K_{eff} considering that in practice, a value of K_{eff} is not readily available as a large number of small-scale K estimate needs to be collected and a suitable stochastic theory applied to compute the K_{eff}. The traditional analysis of cross-hole pumping tests by treating the medium to be homogeneous (if they are all averaged) should yield a representative value of K for a large portion of the flow and simulation domain. However, our results from Table 3 show that cross-hole pumping tests conducted at different sections of the aquifer by pumping at a different port could yield different mean K estimates rendering the estimation of K_{eff} by relying solely on averaging K_{eq} values to be unreliable. Therefore care must be taken in estimating the K_{eff}.

[35] Estimation of $\sigma^2_{\ln K}$ always involves uncertainty. A previous numerical study conducted by *Yeh and Liu* [2000], however, has demonstrated that $\sigma^2_{\ln K}$ has negligible effects on the estimated K using the inverse model. Therefore we use an initial value of 3.0, which is larger than that obtained from the available small-scale data (0.87) and set this as our input $\sigma^2_{\ln K}$ in the inverse model.

[36] Correlation scales represent the average size of heterogeneity that is critical for analyzing the average behavior of aquifers. Correlation scales of any geological formation are difficult to determine. The effects of uncertainty in correlation scales on the estimate based on the tomography are negligible because the tomography produces a large number of head measurements, reflecting the detailed site-specific heterogeneity [*Yeh and Liu*, 2000]. Therefore the correlation scales were approximated based only on the average thickness and length of the discontinuous sand bodies, which are set as $\lambda_x = 50$ cm, $\lambda_y = 10.2$ cm, and $\lambda_z = 10.0$ cm.

[37] For the steady state hydraulic tomography analysis, we selected eight pumping tests at ports 2, 5, 14, 17, 32, 35, 44, and 47 and the corresponding steady state head observations at the rest of 47 ports during each test as data sets. We elected to not use the head data from the pumped port from each test because the pumped port could be affected by skin effects [*Illman et al.*, 2007]. The remaining 16 tests with pumping taking place at ports (8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42) were reserved for the validation of the K tomogram.

[38] Prior to the computation of the K tomogram with the SSLE algorithm, we preprocessed the hydraulic head data. This is because *Illman et al.* [2008] found from the analysis of cross-hole pumping test data obtained in a laboratory sandbox aquifer that the signal-to-noise ratio can be critical in inverse modeling of cross-hole pumping test data. Despite calibrating the pressure transducer prior to the start of each pumping test, there is a minute level of drift always present in each of the pressure transducers. Therefore we removed this drift by shifting the head value to a common one for all transducers in the sandbox prior to each test. The collected data were additionally processed by taking the average of the steady state head values collected over a period of time in a given monitoring port.

[39] *Illman et al.* [2008] previously found that the order in which the pumping test data is included into the SSLE algorithm is critical to the final result. This is because each cross-hole pumping test data set contains different levels of noise. Our experiments in the laboratory were conducted as uniformly as possible, but there are noises that we could not control from one test to the next. Our findings include the importance of examining test data carefully and using the cleanest data (with the highest signal-to-noise ratio) first and progressively including noisier data (with lower signal-to-noise ratio) into the SSLE algorithm. This is because the SSLE algorithm in its current form [*Yeh and Liu*, 2000; *Zhu and Yeh*, 2005, 2006] is more sensitive to noise during the beginning stages of K tomogram computation. This sensitivity is due to the uniform convergence criteria used for sequentially inverting all pumping test data. Careful examination of head data showed that the data most devoid of noise were found near the bottom of the sandbox, while the noisiest data were usually located near the top of the sandbox. This is due to the fact that a higher water column sits on the pressure transducers near the bottom of the sandbox, which causes the pressure transducers to be less affected by noise. Also, the signal is magnified near the bottom no-flow boundary due to the superposition principle. Therefore we included the steady state head data from cross-hole tests with pumping taking place in the order of ports 47, 44, 35, 32, 17, 14, 5, and 2 into the SSLE algorithm.

5.3. Computation of K Tomograms and its Visual Assessment

[40] All computations for steady state hydraulic tomography analyses were executed using eight of 16 processors on a PC-cluster consisting (of one master and 15 slaves each with Pentium IV 3.2 GHz with 1 GB of RAM) at the University of Arizona. The total computational time was about 1 min indicating the efficiency of the SSLE algorithm. Figures 5a–5d are the K tomograms obtained by inverting the steady state head data from two, four, six, and eight pumping tests, respectively. Figure 5a shows that with only two pumping tests, a coarse picture of the heterogeneity pattern emerges and as more tests are included into the SSLE algorithm, more detail to the heterogeneity structure emerges. The final K tomogram obtained (Figure 5d) using eight pumping tests revealed considerable detail to the heterogeneity structure.

[41] Table 5 summarizes the K_G, $\sigma^2_{\ln K}$, and correlation lengths of the resulting K tomograms. The estimated K_G of the K tomogram after including data from eight tests was 1.0×10^{-1} cm/s while the estimated $\sigma^2_{\ln K}$ was 1.12. The K_G was in the midrange of the values from other characterization approaches, while $\sigma^2_{\ln K}$ was highest among all approaches (Table 3). It is of interest to note that there is little change in the K_G and the correlation lengths of the K distribution as more tests were included. On the other hand, the $\sigma^2_{\ln K}$ increased as more tests were included into the SSLE algorithm. These results imply that with as few as two pumping tests, one could reliably estimate the K_G and the correlation lengths of the K distribution in the synthetic aquifer; however, the accurate estimation of $\sigma^2_{\ln K}$ requires more cross-hole pumping tests.

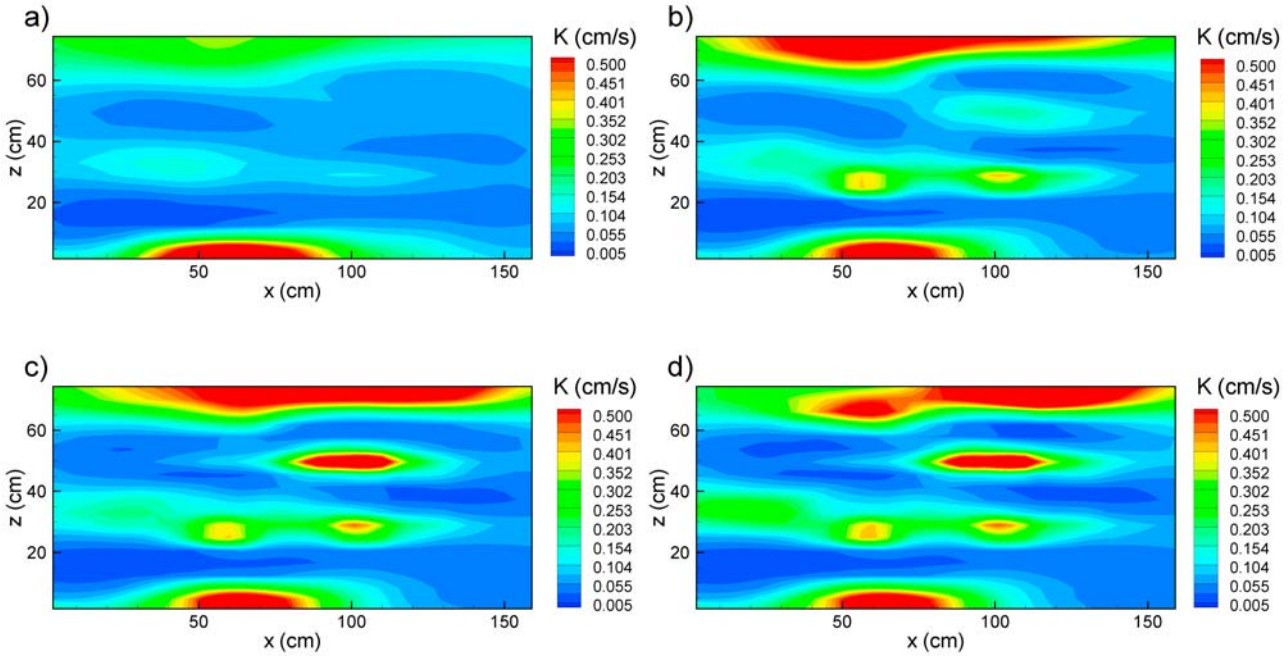

Figure 5. *K* tomograms computed using (a) two pumping tests (ports 47 and 44); (b) four pumping tests (ports 47, 44, 35, and 32); (c) six pumping tests (ports 47, 44, 35, 32, 17, and 14); and (d) eight pumping tests (ports 47, 44, 35, 32, 17, 14, 5, and 2).

[42] To compare the *K* tomogram (Figure 5d) to the location of the sand deposits or lithofacies (Figure 3), we overlay Figure 5d over Figure 3 using Adobe Photoshop. Examination of Figure 6 shows a good correspondence between the high *K* zones of *K* tomogram to layer 1 (20/30), 7 (#12), 9 (20/40), 11 (#12), 15 (20/40), and 18 (20/30). Despite the good correspondence of the general locations of these features, it is also apparent from Figure 6 that the boundaries of the lithofacies from the deposits to the boundaries of high *K* features from the tomogram are not in close agreement. We also note that layers 4 (20/40), 12 (16/30), and 13 (20/30), all expected to be high *K* features, are not shown on the *K* tomogram.

[43] We also compare the *K* tomogram (Figure 5d) to the kriged *K* estimates of small-scale data (Figures 4a and 4b). A visual comparison of the estimated *K* tomogram to the kriged *K* fields from the small-scale estimates of *K* reveals a weak correspondence implying that steady state hydraulic tomography yields results that are significantly different from the standard geostatistical analysis of small scale *K* data.

6. Forward Simulations of Independently Conducted Cross-Hole Pumping and Flow-Through Tests

[44] A number of approaches have been proposed in the literature to validate forward and inverse groundwater flow models. For example, *Foglia et al.* [2007] tested alternative forward groundwater models using the corrected Akaike information criterion, Bayesian information criterion, and generalized cross-validation. *Poeter and Anderson* [2005] utilized a Kullback-Leibler information criterion that selects parsimonious groundwater models that provided a more realistic measure of precision than evaluation of any one model or evaluation based on other commonly referenced model selection criteria. In terms of the validation of inverse modeling results, *Painter et al.* [2007] compared three separate transmissivity estimation approaches and compared the resulting heads from forward simulation and actual head data from the Edwards aquifer in Texas. *Illman et al.* [2007] and *Liu et al.* [2007] used various data (core, slug, and single-hole test results) and statistical moments of these small-scale data, as well as independent cross-hole pumping test data to validate their *K* and S_s tomograms. In both studies, *Illman et al.* [2007] and *Liu et al.* [2007] both concluded that the best approach to validate a *K* or S_s tomogram was to simulate an independently conducted test not used in the construction of the tomograms. We consider the cross-hole tests to be the most appropriate testing approach for validation purposes because the tests can be conducted at different locations within the aquifer with a large number of observation points as implemented in our sandbox study. Pumping at different locations will induce flow throughout the aquifer yielding drawdowns along the layers as well as across them. Other methods could potentially involve

Table 5. Statistical Properties of the Estimated *K* Tomograms

Case	K_G (cm/s)	$\sigma^2_{\ln K}$	λ_x (cm)	λ_z (cm)
Hydraulic tomography (Sequential – 2 tests)	0.090	0.542	16	7
Hydraulic tomography (Sequential – 4 tests)	0.101	0.890	18	9
Hydraulic tomography (Sequential – 6 tests)	0.099	1.036	14	6
Hydraulic tomography (Sequential – 8 tests)	0.100	1.121	18	7

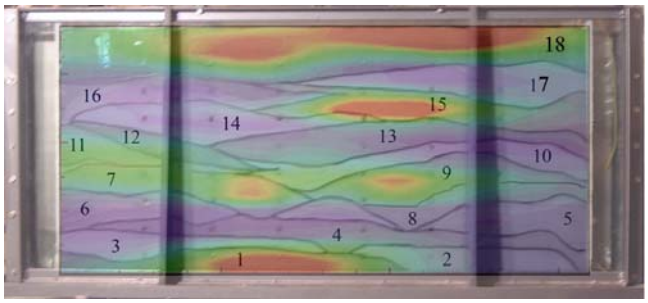

Figure 6. Overlay of K tomogram computed using eight pumping tests over the picture of the heterogeneous sandbox aquifer using Adobe Photoshop.

the discrete collection of aquifer samples, small-scale hydraulic tests, and flux measurements all of which yield localized information that may not be entirely suitable for large-scale validation of groundwater flow models.

[45] Another method to test the performance of the various K fields is to simulate the flow-through experiments using the various homogeneous and heterogeneous K fields. If the estimated K field is accurate, it should, in principle, yield accurate estimates of total flow rate on the outflow end of the sandbox resulting from the flow-through tests.

[46] Here, to quantitatively assess the validity of the various homogeneous and heterogeneous K estimates, we simulate 16 additional cross-hole pumping and six bidirectional flow-though tests using the forward groundwater model MMOC3 developed by *Yeh et al.* [1993]. For the validation using cross-hole tests, the simulated and measured drawdown values for each case are plotted on separate scatterplots and a linear model is fitted to each case without forcing the intercept to zero. The linear model fit and the coefficient of determination (R^2) provide indications of scatter and bias. The R^2 is a statistic that provides a quantitative measure of similarity between the simulated and measured drawdown values. A high R^2 value means that the simulated and measured drawdown values are linearly correlated, even though the mean values could be different. Other measures of correspondence between the simulated and observed drawdown values are the mean absolute error (L_1) and the mean square error (L_2) norms. The L_1 and L_2 norms are computed as:

$$L_1 = \frac{1}{n}\sum_{i=1}^{n} |\chi_i - \hat{\chi}_i| \tag{3}$$

$$L_2 = \frac{1}{n}\sum_{i=1}^{n} (\chi_i - \hat{\chi}_i)^2 \tag{4}$$

where n is the total number of drawdown data, i indicates the data number, and χ_i and $\hat{\chi}_i$ represent the estimates from the simulated and measured drawdowns, respectively. The slope and intercept of the linear model fit, the R^2 values, and the L_1 and L_2 norms are summarized in Table 6. These statistics collectively provide quantitative measures of the performance of each steady state forward groundwater model with different K fields in simulating the 16 cross-hole pumping tests. The use of more than one cross-hole pumping test conducted at different parts of the synthetic heterogeneous aquifer ensures a more credible validation of each of these models.

[47] For the validation of the various K fields using the flow-through tests, we simply compare the total flow rates obtained through the forward simulation of the flow-through tests using MMOC3 to those obtained from the actual experimental data.

6.1. Use of Independent Cross-Hole Tests to Validate Homogeneous and Heterogeneous *K* Fields

[48] Figure 7 shows a series of scatterplots that compare the simulated drawdown values from various K_G fields from Table 3 against the measured values from 16 separate cross-hole pumping tests. Figure 7 includes a 45° line which indicates a perfect correlation. In particular, Figure 7a shows that the K_G from the permeameter analysis of cores yields reasonable predictions of 16 separate cross-hole tests over a large range of drawdown values. Our assessment is based on the near unit slope (1.02) of the linear model fit to the data. However, the intercept of the linear model is 0.20 suggests that the predictions are biased and a R^2 value of 0.83 reveals that there is some scatter. Figure 7b shows that the geometric mean of the K_G from single-hole tests yields biased predictions of 16 separate cross-hole tests. The simulated drawdowns are in general higher than the measured ones, which is a consequence of the lower K_G estimated using the single-hole tests conducted in this aquifer.

[49] The K_G values from the two cross-hole tests yielded mixed results. For the K_G value estimated from the cross-hole pumping test with pumping taking place at port 17, the prediction of the 16 cross-hole pumping tests are reasonable (Figure 7c) and is in similar quality as the core case (Figure 7a). However, the prediction based on the K_G from cross-hole pumping test with pumping taking place at port 21 is biased (Figure 7d). For this case, the measured drawdowns are in general higher than the simulated ones, which is a consequence of the higher K_G estimated through cross-hole pumping tests. The flow-through test estimate of K_G, likewise yielded biased predictions (Figures 7e), similar to cross-hole pumping test

Table 6. Slope and Intercept of Fitted Linear Model, Coefficient of Determination, and Standard Errors of Various Homogeneous and Heterogeneous Cases[a]

Case	Slope	Intercept	R^2	L_1	L_2
Homogeneous K Field					
K_G (core)	1.02	0.20	0.83	0.33	0.22
K_G (single-hole)	1.32	0.26	0.83	0.60	0.70
K_G (cross-hole – port 17)	1.03	0.21	0.83	0.34	0.23
K_G (cross-hole – port 21)	0.71	0.15	0.83	0.31	0.17
K_G (flow through)	0.58	0.13	0.83	0.39	0.28
Heterogeneous K Field					
Kriged core K	1.10	0.14	0.89	0.30	0.18
Kriged single-hole K	1.45	0.13	0.89	0.58	0.67
Hydraulic tomography (Sequential – 2 tests)	1.07	−0.02	0.93	0.18	0.07
Hydraulic tomography (Sequential – 4 tests)	1.05	0.00	0.96	0.13	0.04
Hydraulic tomography (Sequential – 6 tests)	1.08	0.05	0.97	0.15	0.05
Hydraulic tomography (Sequential – 8 tests)	1.07	0.02	0.97	0.13	0.04

[a] R^2 is the coefficient of determination and L_1 and L_2 statistics are standard errors.

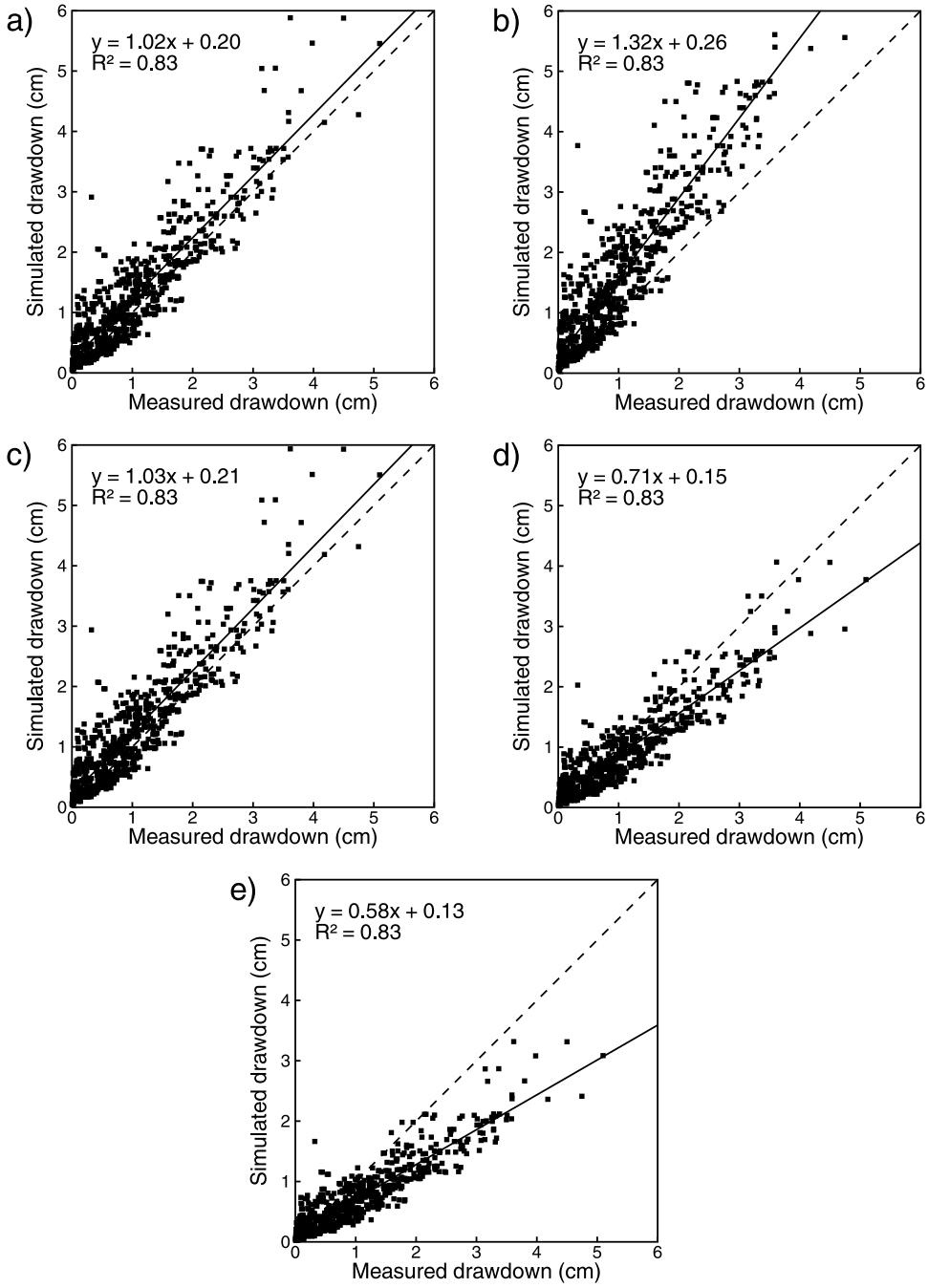

Figure 7. Arithmetic plots of drawdown predictions of 16 separate cross-hole pumping tests using the (a) core K_G estimates; (b) single-hole K_G estimates; (c) geometric mean of cross-hole test K_{eq} estimates (port 17); (d) geometric mean of cross-hole test K_{eq} estimates (port 21); and (e) K_{eff} from flow-through tests.

case where pumping took place at port 21. This suggests that K_G estimated by flow-through tests is higher than the true mean K of this sandbox aquifer. The latter makes sense considering that flow during the flow-through tests takes place horizontally primarily through the high K layers. This should result in a mean K value that should be closer to the arithmetic mean K, which is higher than the geometric mean.

[50] Figure 8 shows a series of scatterplots which compare the simulated drawdown values from various heterogeneous K fields against the measured values from 16 separate cross-hole pumping tests. Figure 8a shows that the kriged K field based on permeameter tests on cores yields slightly improved predictions of 16 separate cross-hole pumping tests over a large range of drawdown values in comparison to the K_G case. This is supported by the slight improvement of the R^2 value as well as the decrease in the L_1 and L_2 norms (Table 6) when the kriged case is compared to the K_G case. Likewise, Figure 8b illustrates that the kriged K field from single-hole pumping test data shows little improvement from its homogeneous counterpart based on

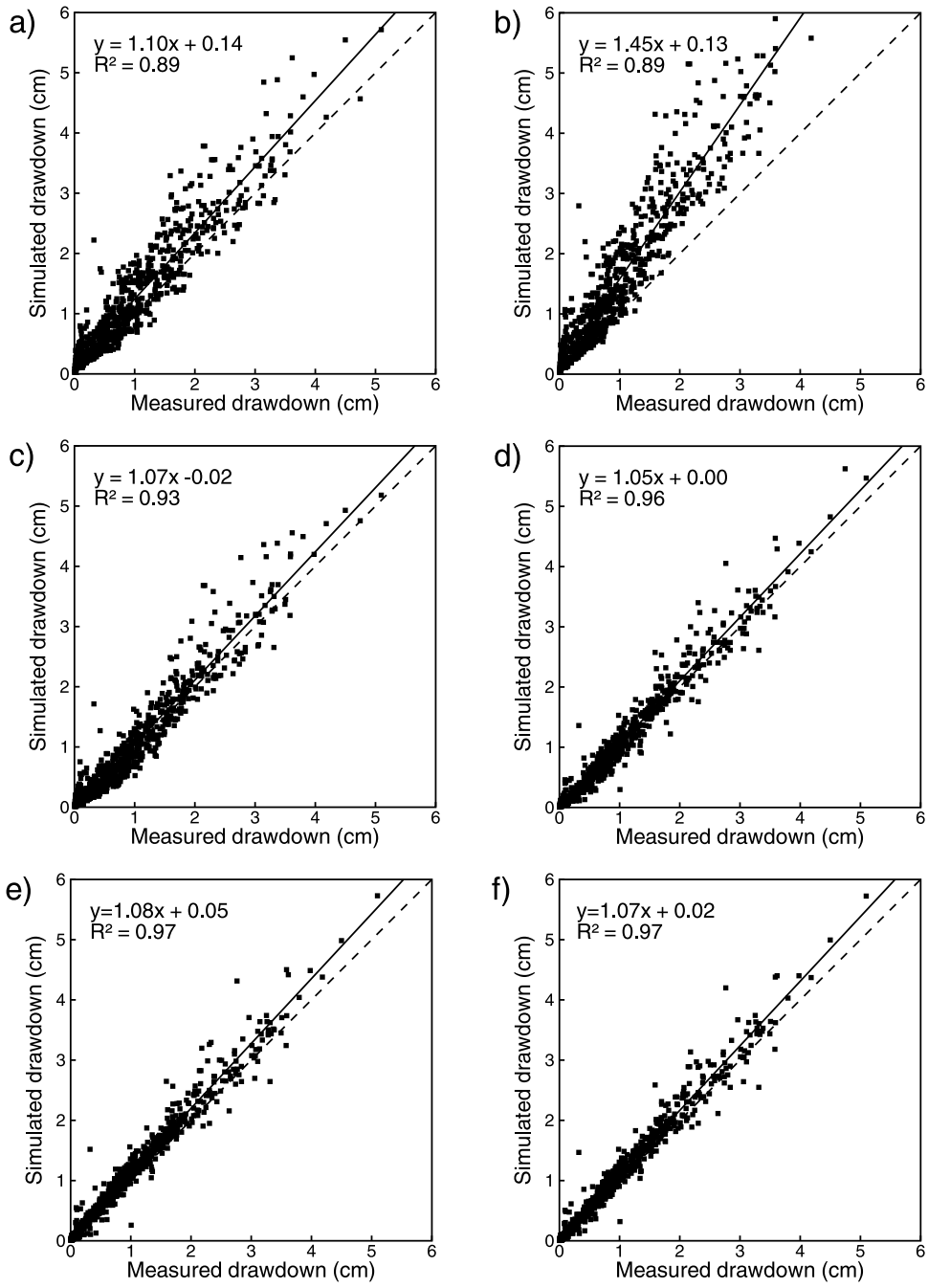

Figure 8. Arithmetic plots of drawdown predictions of 16 separate cross-hole pumping tests using (a) kriged core K estimates; (b) kriged single-hole K estimates; (c) K tomogram from SSHT (two tests, ports 47 and 44); (d) K tomogram from SSHT (four tests, ports 47, 44, 35, and 32); (e) K tomogram from SSHT (six tests, ports 47, 44, 35, 32, 17, and 14); and (f) K tomogram from SSHT (eight tests, ports 47, 44, 35, 32, 17, 14, 5, and 2).

the linear model fit, the R^2 value as well as the lower L_1 and L_2 norms.

[51] In contrast, the use of the K tomograms from hydraulic tomography yields a drastically improved prediction of measured drawdowns using 16 independent cross-hole pumping tests not used in the construction of the tomogram. Figures 8c–8f show these comparisons for K tomograms constructed using two, four, six, and eight cross-hole pumping tests, respectively. It is evident from Figures 8c–8f that the predictability of independently conducted cross-hole pumping tests improves with a larger number of tests included into the SSLE algorithm.

6.2. Use of Bidirectional Flow-Through Tests to Validate Various Homogeneous and Heterogeneous K Fields

[52] The various homogeneous and heterogeneous K fields were additionally validated by simulating the actual flow-through tests using MMOC3. This was done by prescribing hydraulic head values at the left and right bound-

Table 7. Relative Error Computed for the Simulated Flow Rates With Respect to the Measured Values[a]

	Homogeneous K Fields					Heterogeneous K Fields		
Test	FT	Core	SH	CH (17)	CH (21)	Kriged (core)	Kriged (SH)	K Tomogram
LR1	3.2	−41.6	−55.2	−42.3	−15.8	−37.9	−52.6	11.1
LR2	1.6	−42.5	−55.9	−43.2	−17.1	−38.9	−53.4	9.3
LR3	0.1	−43.3	−56.6	−44.1	−18.3	−39.8	−54.1	7.7
RL4	−4.7	−46.0	−58.6	−46.7	−22.2	−42.7	−56.2	2.6
RL5	−2.0	−44.5	−57.5	−45.2	−20.0	−41.0	−55.0	5.5
RL6	−2.3	−44.7	−57.6	−45.4	−20.3	−41.2	−55.1	5.2

[a]Relative error is measured in %. LR1 stands for flow-through test number one with flow from left to right. FT is flow-through test; SH is single-hole tests, and CH (17) is cross-hole test with pumping at port 17.

aries of the sandbox aquifer and conducting forward simulations. The flow rates from the outflow boundary were computed by first calculating flow for every element along the outflow column and second summing the flows from these elements to obtain the total flow rate. In the first step, the flow rate for each outflow element was calculated using Darcy's law. The computed total flow rate is then compared against the measured total flow rate from the actual flow-through tests.

[53] Table 7 summarizes the relative error of the simulated flow rate with respect to the measured flow rates from six separate, bidirectional flow-through tests. In general, we see that the homogeneous K fields from the core, single-hole, and cross-hole tests underestimate the flow rates of the various flow-through tests resulting in large relative errors (i.e., $-15.8 \sim -58.6\%$). This implies that the total outflow through the synthetic heterogeneous aquifer cannot be accurately predicted using the various homogeneous K fields. However, this is not the case when we use the K_G determined through the flow-through experiments. This is an expected result because the simulation of the flow-through tests was done with K_G determined from the flow-through tests. We speculate that the slight difference in flow rates results from experimental errors and the differences in how total flow is measured for the simulation and the actual experiments.

[54] The simulated flow rates from the kriged K fields from cores and single-hole tests also yielded inaccurate estimates of total flow as shown by the high relative error ($-37.9 \sim -56.2\%$). The relative error was slightly less for kriged core and single-hole K fields in comparison to their homogeneous counterparts, which suggests that accounting for the heterogeneous nature of the strata can lead to improved estimates of flow.

[55] In contrast, the relatively error is significantly reduced when we use the K tomogram from steady state hydraulic tomography to predict total outflow. In particular, the relative error ranges from 2.6 to 11.1%, which is significantly less than the previous cases examined. The main reason for the slight inaccuracy is due to the uncertain nature of K values away from the observation ports. Figure 9 is a contour map of ln K variance for the case in which eight cross-hole tests are included into the SSLE algorithm (Figure 8d). Figure 9 shows that the variance is highest outside the interior of the synthetic heterogeneous aquifer, where observation ports are sparse. Given the uncertain nature of the K estimates in these regions, it is not surprising to us that there is a slight discrepancy in the simulated and measured outflow rates. The prediction of total outflow rate will likely improve with additional pressure transducers placed in the high ln K variance areas. On the basis of these results, it leads us to believe that the SSLE algorithm yields the most accurate K field among all the characterization approaches validated here.

7. Discussion

7.1. On Model Validation

[56] The validation of groundwater models has become an important issue [e.g., *Hassanizadeh and Carrera*, 1992; *Konikow and Bredehoeft*, 1992; *Oreskes et al.*, 1994; *Refsgaard and Knudsen*, 1996; *Refsgaard*, 1997; *Refsgaard and Henriksen*, 2004] over the last several decades. Model validation involves the establishment of greater confidence in the model by conducting simulations of the system under conditions in which the data has not been used for calibration purposes. For example, one can calibrate a groundwater flow model using one set of pumping test data. If the calibrated model from this first pumping test can predict system response accurately in a second pumping test (e.g., conducted using another well), one can have greater confidence in the calibrated model. On the other hand, if the parameters need to be adjusted to match the response of the second pumping test, the process becomes a second calibration and additional data sets are needed to continue with the validation exercise.

[57] We emphasize that successful model validation implied in this paper involves the establishment of greater confidence in the model by conducting simulations of the system under conditions in which the data has not been used for calibration purposes and obtaining excellent predictions of not only one, but multiple independent tests. It does not mean that the model, per se, is invincible and predicts the truth at all occasions. It is also important to recognize that "absolute" validation of a model is likely never possible. In

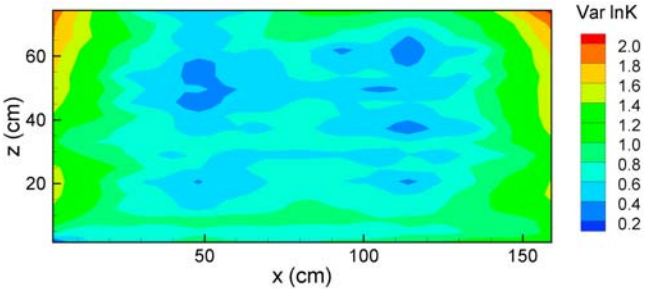

Figure 9. Contour plot of ln K variance with eight cross-hole tests included into SSLE algorithm.

this sense, we agree with *Refsgaard and Henriksen*'s [2004] definition of model validation: it is the "substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model."

7.2. On the Accuracy of *K* Estimates and its Impact on Model Validation

[58] The laboratory experimental results clearly showed that the K_G estimates from various established characterization methods yielded biased predictions of independently conducted cross-hole pumping tests. This, in turn resulted in erroneous predictions of total outflow through our synthetic heterogeneous aquifer. Therefore steady state groundwater models constructed using the K_G estimates obtained via traditional characterization approaches are considered to be invalid based on the definition of model validation given in the previous section. In fact, our results illustrate that small differences in the K_G values among various characterization methods result in large differences in steady state drawdown and flow rate predictions of independent tests. Inaccurate predictions of drawdowns and flow rates will undoubtedly and negatively impact contaminant transport predictions.

[59] A surprising finding from this study is that heterogeneous *K* fields from kriging did not cure this problem and yielded biased predictions. However, the *K* tomogram from steady state hydraulic tomography was able to predict drawdowns of all independently conducted cross-hole hydraulic tests with considerable accuracy in comparison to the steady state groundwater model based on other *K* estimates. The *K* tomogram was also able to provide better predictions of flow rates in comparison to other approaches. These results lead us to conclude that the steady state groundwater flow models based on the *K* tomogram can be considered to be validated sufficiently based on the definition of model validation given before.

7.3. Why Does Hydraulic Tomography Yield Improved *K* Fields Over Traditional Methods?

[60] One reason why hydraulic tomography yields improved *K* fields in comparison to other traditional methods is because the sequential or simultaneous inclusion of multiple pumping test data into the SSLE algorithm amounts to a repeated calibration and validation of the estimated *K* field with multiple data sets that are analyzed by the inverse algorithm. Another reason is that hydraulic tomography relies on multiple pumping tests conducted at different locations in the tested medium. The signals generated through multiple pumping events at different points in the aquifer are captured with pressure transducers placed in neighboring boreholes and synthesized using a suitable inverse approach. The large amount of data generated using hydraulic tomography surveys leads to improved estimates of *K* in the aquifer. Of course, the accuracy of the *K* tomogram computed will depend on the inverse modeling strategy used and the number of pumping tests included into the inverse algorithm and the pumped locations as well as the number of observation points available for the inverse analysis.

[61] Potential reasons why established approaches may not perform as well as hydraulic tomography include (1) experimental errors arising from packing of core samples and anisotropy in *K* resulting from the 1-D nature of permeameter measurements, (2) small sampling volume, (3) near-well effects that could affect single-hole or other small-scale tests that involve injection or withdrawal of fluids, and (4) lack of adequate sampling between boreholes that could affect the estimation of connectivity of *K* heterogeneity. Furthermore, it appears that the homogeneous *K* value obtained by averaging of small-scale *K* data or taking the geometric mean of the K_{eq} from cross-hole tests is simply not adequate enough to predict drawdowns of independent cross-hole pumping tests in a heterogeneous aquifer because the computed mean represents a volume smaller than the representative elementary volume (REV) [e.g., *Jankovic et al.*, 2003]. However, as pointed out by *Neuman* [1987] and *Tartakovsky and Neuman* [1998], there is generally no guarantee that a REV can be defined for a given geologic medium. Even if such an REV can be defined, it will be so large that estimation of *K* using available experimental method becomes impractical. In the real world, estimates of *K* or their distributions are often obtained at a scale smaller than the REV or what we refer to as the sub-REV scale. This is partially due to the lack of guidelines in estimating the size of a REV suitable for a given site leading to greater uncertainty in *K* estimation using traditional characterization approaches. This is another reason why we consider hydraulic tomography to be a robust approach in delineating the subsurface *K* heterogeneity as it does not rely on any REV assumptions for accurate estimation of the *K* field.

[62] Our laboratory experiments have shown the powerful capability of steady state hydraulic tomography to obtain an accurate *K* tomogram. However, as robust as hydraulic tomography may be, we acknowledge that it is by no means a panacea technology. This is because the computed *K* tomogram is nonunique as there are an infinite number of solutions to the steady state inverse problem for a heterogeneous *K* field, even when all of the forcing functions are fully specified. Only when data are available at all estimated locations will the inverse problem be well-posed and ultimately lead to a unique solution [e.g., *Yeh et al.*, 1996; *Yeh and Liu*, 2000; *Yeh and Simunek*, 2002]. This, obviously, is not the case here. For example, we noted earlier that some of the layers in the synthetic heterogeneous aquifer were not identified by steady state hydraulic tomography. We suspect that some of these features were not identified because of the limited number of pressure transducers installed along each column of wells (Figure 1). A larger number of pressure transducers or conditioning the *K* tomogram with higher-resolution measurements of *K* along the columns would likely improve the resolution of the *K* tomogram. In any case, it is important to note that the misidentification of some of these layers could have large impacts on predicting solute transport; thus further improvement in hydraulic tomography technology is desirable. This improvement will require technological advancements that will lower the cost of downhole pressure measurements. If monitoring costs can be substantially lowered, a denser pressure measurement network becomes more feasible increasing the ability of hydraulic tomography to capture finer-scale details of geologic heterogeneity.

[63] Despite some shortcomings, it is important to recognize that we have obtained a solution to the inverse problem (i.e., the *K* tomogram) that is consistent with the heterogeneity patterns that yield accurate predictions of steady state drawdowns from multiple, independent cross-

hole pumping tests not used in the construction of the K tomogram. The robustness of our results is further supported by the accurate prediction of flow rates through the synthetic heterogeneous aquifer. Results presented here show that accurate steady state groundwater models can be developed and that the degree of its validation will in general improve when there is more accurate information on subsurface heterogeneity and forcing functions. Therefore we contend that more emphasis should be placed on accurate site characterization to delineate subsurface heterogeneity patterns and their connectivity, developing improved methods for defining source/sinks (other than the pumping/injection rates which are typically known), and developing improved methods for increasing the accuracy in specifying boundary conditions. If this is done adequately, we predict that steady state groundwater model will become more accurate and its predictability of future scenarios will improve. Of course, this prediction is based on the caveat that no fundamental changes occur to the aquifer (e.g., deformation) after the calibration process has been completed.

8. Summary, Findings, and Conclusions

[64] Groundwater modeling has become a vital and indispensable component to water supply and contaminant transport investigations throughout the world in various applications. An important component of groundwater modeling under steady state conditions is selecting a representative hydraulic conductivity (K) estimate or set of estimates which defines the K field of the studied region. Currently, there are a number of characterization approaches to obtain K at various scales and in varying degrees of detail, but there is a paucity of information in terms of which approach best predicts drawdowns caused by external forcings and/or flow through aquifers. The main objective of this paper was to assess K estimates obtained by various approaches by predicting independent cross-hole pumping tests and results from flow-through tests in a synthetic sandbox aquifer. To accomplish this, we built a synthetic heterogeneous aquifer in the laboratory through cyclic flux of sediment-laden water in a laboratory sandbox. We then characterized the synthetic aquifer through various established techniques (permeameter analyses of core samples, single-hole, cross-hole, and flow-through testing). The test data then were subjected to traditional methods of analysis by treating the medium to be homogeneous. We then obtained mean K estimates from the data collected using traditional approaches. Heterogeneous K fields were also obtained through kriging of small-scale samples and steady state hydraulic tomography analysis of eight cross-hole pumping test data. To validate the various K estimates, we conducted forward simulations of 16 pumping tests using these homogeneous and heterogeneous K fields and comparing them to observation drawdown data. Simulation of the 16 independent tests showed that the steady state head distribution can be simulated very accurately with the K distribution or "K tomogram," but not as well with the geometric mean of hydraulic conductivity (K_G) values from established characterization approaches, and through kriging of the small-scale K data. We also simulated the flow-through tests through the synthetic heterogeneous aquifer using the various homogeneous and heterogeneous K fields, calculated total flow from the outflow boundary and compared to those from the actual experiments. These results again showed that the K tomogram provides the best estimates of total flow through the synthetic aquifer, which adds further support for the robustness of the K estimates. Overall, our results provide encouraging, positive evidence that groundwater models could be validated if the geologic medium is characterized sufficiently through hydraulic tomography and if the forcing functions are specified correctly.

[65] This study leads to the following major findings and conclusions:

[66] 1. The traditional characterization by means of permeameter analysis of cores, single-hole, and cross-hole tests of this laboratory heterogeneous aquifer yielded a range of geometric mean of hydraulic conductivity (K_G) and variance of ln K ($\sigma^2_{\ln K}$). The K_G ranged from 0.059 to 0.136 cm/s, while the $\sigma^2_{\ln K}$ ranged from 0.001 to 0.868. The implication of these results is that no one method could be relied upon to characterize the aquifer accurately to obtain a reliable K_G or $\sigma^2_{\ln K}$ because of differences in scales in which the data are collected as well as experimental and interpretive errors that arise from various methods.

[67] 2. Cross-hole pumping tests conducted at different locations within the aquifer (with different pumping locations) yielded different K_G values of approximately a factor of three difference. We emphasize that this small difference in the K_G could be interpreted to be small or lie within a margin of error given the uncertainties associated with each pumping test. In this sandbox aquifer, the factor of three difference in K caused a large variation in terms of model validation. A similar finding that the mean K from cross-hole tests differed from various cross-hole tests was reported by *Vesselinov et al.* [2001a, 2001b] and *Illman* [2006] in the context of cross-hole pneumatic injection tests conducted in unsaturated fractured tuff at the Apache Leap Research Site (ALRS) and by *Illman and Tartakovsky* [2006] in the context of cross-hole hydraulic tests in fractured granite at the Grimsel site.

[68] 3. Steady state hydraulic tomography of cross-hole pumping tests on the synthetic heterogeneous aquifer yielded the highest $\sigma^2_{\ln K}$ among the various characterization approaches. A similar finding was reported by *Vesselinov et al.* [2001b] who analyzed three cross-hole pneumatic injection tests conducted in unsaturated fractured tuffaceous rock at the ALRS in a tomographic manner. The cause of the high $\sigma^2_{\ln K}$ estimates from hydraulic and pneumatic tomography result from multiple factors, including but not limited to the fact that the approach is more suitable in capturing the heterogeneity patterns, but can also be affected by conceptual, numerical, and estimation errors.

[69] 4. Steady state hydraulic tomography analysis of cross-hole pumping tests yielded a K tomogram that was qualitatively similar looking to the photograph of the synthetic aquifer deposits. In particular, a number of high K features and their locations estimated using the steady state hydraulic tomography approach corresponded quite well with certain deposits known to have high K values in the aquifer. However, this correspondence was far from perfect with a few strata expected to have high K layers did not appear in the K tomogram. Despite the good correspondence of the general locations of high K features, the comparison showed that the boundaries of the lithofacies from the deposits to the boundaries of high K features from the tomo-

gram are not in close agreement. The similarity or difference between the lithofacies from geological investigations and hydrofacies determined from hydraulic tomography and other methods should be examined with greater scrutiny in future studies.

[70] 5. The geometric mean K value determined from small-scale data and homogeneous interpretation of cross-hole tests were used in a forward model to simulate 16 cross-hole pumping tests conducted in the synthetic aquifer. The forward simulations revealed that the predictions of 16 pumping tests using geometric and effective K estimates from various methods showed biased results in terms of predicting drawdowns from independent cross-hole tests. In particular, when the estimated geometric mean value was higher, the measured drawdown values were found to be higher in comparison to the simulated values. The converse was also found to be true.

[71] 6. Kriging of K estimates from small-scale tests (core and single-hole) was conducted. Forward simulations of the 16 pumping tests showed biased results in terms of predicting drawdowns from independent cross-hole tests. This suggests that the traditional geostatistical analysis of small-scale K data may not yield accurate K distributions for the prediction of heads from large-scale cross-hole pumping tests.

[72] 7. Steady state hydraulic tomography analysis of eight cross-hole tests yielded a more accurate heterogeneous distribution of K in comparison to the kriged fields. Our observation is based on the forward simulation of 16 cross-hole tests not used in the construction of the tomogram which revealed an excellent correspondence between the measured and simulated drawdowns.

[73] 8. The various homogeneous and heterogeneous K estimates were then further tested by simulating six bidirectional flow-through tests. The homogeneous K estimates by taking the geometric mean of core, single-hole, and cross-hole estimates of K significantly underestimated total flow from the outflow boundary. Likewise, forward simulations using heterogeneous K distributions obtained by kriging of core and single-hole K also underestimated total flow, although the predictions were slightly better than their homogeneous counterpart. In contrast, total flow estimates obtained using the K tomogram from steady state hydraulic tomography analysis of eight cross-hole tests was considerably more accurate providing further support for the robustness of hydraulic tomography in accurately characterizing the subsurface.

[74] 9. On the basis of our experimental work and modeling efforts, we conclude that steady state groundwater models can be validated at least in this laboratory sandbox aquifer. Of course, absolute validation will likely never be possible. This conclusion comes with the caveat that groundwater model validation is possible only when the subsurface heterogeneity patterns and forcing functions are captured to sufficient accuracy. Finally, this study suggests that more effort should go into the accurate characterization of the subsurface if the objective is to better predict subsurface head distributions, which then should yield better predictions of the fate of subsurface contaminants. Our results lend strong support to *Yeh and Lee*'s [2007] opinion that it is time to change the way we collect and analyze data for subsurface characterization.

[75] **Acknowledgments.** This research was supported by the Strategic Environmental Research and Development Program (SERDP) under grant ER-1365, ER-1610, by the National Science Foundation (NSF) under grants EAR-0229713, IIS-0431069, and EAR-0450336, and by the Natural Sciences and Engineering Research Council (NSERC) of Canada through the Discovery grant to WAI.

References

Barth, G., M. Hill, T. Illangasekare, and H. Rajaram (2001), Predictive modeling of flow and transport in a two-dimensional intermediate-scale, heterogeneous porous medium, *Water Resour. Res.*, *37*(10), 2503–2512, doi:10.1029/2001WR000242.

Boggs, J. M., S. C. Young, L. M. Beard, L. W. Gelhar, K. R. Rehfeldt, and E. E. Adams (1992), Field study of dispersion in a heterogeneous aquifer: 1. Overview and site description, *Water Resour. Res.*, *28*(12), 3281–3291, doi:10.1029/92WR01756.

Bohling, G. C., X. Zhan, J. J. Butler Jr., and L. Zheng (2002), Steady shape analysis of tomographic pumping tests for characterization of aquifer heterogeneities, *Water Resour. Res.*, *38*(12), 1324, doi:10.1029/2001WR001176.

Bohling, G. C., J. J. Butler Jr., X. Zhan, and M. D. Knoll (2007), A field assessment of the value of steady shape hydraulic tomography for characterization of aquifer heterogeneities, *Water Resour. Res.*, *43*, W05430, doi:10.1029/2006WR004932.

Boman, G. K., F. J. Molz, and K. D. Boone (1997), Borehole flowmeter application in fluvial sediments: Methodology, results, and assessment, *Ground Water*, *35*(3), 443–450, doi:10.1111/j.1745-6584.1997.tb00104.x.

Brauchler, R., R. Liedl, and P. Dietrich (2003), A travel time based hydraulic tomographic approach, *Water Resour. Res.*, *39*(12), 1370, doi:10.1029/2003WR002262.

Butler, J. J., Jr. (2005), Hydrogeological methods for estimation of spatial variations in hydraulic conductivity, in *Hydrogeophysics*, *Water Sci. and Technol. Libr.*, vol. 50, edited by Y. Rubin and S. S. Hubbard, pp. 23–58, Springer, New York.

Cardiff, M., W. Barrash, P. K. Kitanidis, B. Malama, A. Revil, S. Straface, and E. Rizzo (2009), A potential-based inversion of unconfined steady-state hydraulic tomography, *Ground Water*, *47*(2), 259–270, doi:10.1111/j.1745-6584.2008.00541.x.

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

Chao, H. C., H. Rajaram, and T. Illangasekare (2000), Intermediate-scale experiments and numerical simulations of transport under radial flow in a two-dimensional heterogeneous porous medium, *Water Resour. Res.*, *36*(10), 2869–2884, doi:10.1029/2000WR900096.

Craig, A. J. (2005), Measurement of hydraulic parameters at multiple scales in two synthetic heterogeneous aquifers constructed in the laboratory, M.S. thesis, Dept. of Civ. and Environ. Eng., Univ. of Iowa, Iowa City.

Danquigny, C., P. Ackerer, and J. P. Carlier (2004), Laboratory tracer tests on three-dimensional reconstructed heterogeneous porous media, *J. Hydrol. Amsterdam*, *294*(1–3), 196–212, doi:10.1016/j.jhydrol.2004.02.008.

Fernàndez-Garcia, D., H. Rajaram, and T. H. Illangasekare (2005), Assessment of the predictive capabilities of stochastic theories in a three-dimensional laboratory test aquifer: Effective hydraulic conductivity and temporal moments of breakthrough curves, *Water Resour. Res.*, *41*, W04002, doi:10.1029/2004WR003523.

Fienen, M. N., T. Clemo, and P. K. Kitanidis (2008), An interactive Bayesian geostatistical inverse protocol for hydraulic tomography, *Water Resour. Res.*, *44*, W00B01, doi:10.1029/2007WR006730.

Foglia, L., S. W. Mehl, M. C. Hill, P. Perona, and P. Burlando (2007), Testing alternative ground water models using cross-validation and other methods, *Ground Water*, *45*(5), 627–641, doi:10.1111/j.1745-6584.2007.00341.x.

Gelhar, L. W., and C. L. Axness (1983), Three-dimensional stochastic analysis of macrodispersion in aquifers, *Water Resour. Res.*, *19*(1), 161–180, doi:10.1029/WR019i001p00161.

Gottlieb, J., and P. Dietrich (1995), Identification of the permeability distribution in soil by hydraulic tomography, *Inverse Probl.*, *11*(2), 353–360, doi:10.1088/0266-5611/11/2/005.

Hao, Y., T.-C. J. Yeh, W. A. Illman, K. Ando, and K.-C. Hsu (2008), Hydraulic tomography for detecting fracture connectivity, *Ground Water*, *46*(2), 183–192, doi:10.1111/j.1745-6584.2007.00388.x.

Hassanizadeh, S. M., and J. Carrera (1992), Editorial, *Adv. Water Resour*, *15*(1), 1–3, doi:10.1016/0309-1708(92)90027-Y.

Hess, K. M., S. H. Wolf, and M. A. 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*(8), 2011–2027, doi:10.1029/92WR00668.

Hubbard, S. S., and Y. Rubin (2000), Hydrogeological parameter estimation using geophysical data: A review of selected techniques, *J. Contam. Hydrol.*, *45*, 3–34, doi:10.1016/S0169-7722(00)00117-0.

Hufschmeid, P. (1986), Estimation of three-dimensional statistically anisotropic hydraulic conductivity field by means of single well pumping tests combined with flowmeter measurements, *Hydrogeologie*, *2*, 163–174.

Hyndman, D. W., and S. M. Gorelick (1996), Estimating lithological transport properties in three dimensions using seismic and tracer data: The Kesterton Aquifer, *Water Resour. Res.*, *32*(9), 2659–2670, doi:10.1029/96WR01269.

Illangasekare, T. H., E. J. Armbruster, and D. N. Yates (1995), Non-aqueous phase fluids in heterogeneous aquifers–Experimental study, *J. Environ. Eng.*, *121*(8), 571–579, doi:10.1061/(ASCE)0733-9372(1995)121:8(571).

Illman, W. A. (2006), Strong field evidence of directional permeability scale effect in fractured rock, *J. Hydrol. Amsterdam*, *319*(1–4), 227–236, doi:10.1016/j.jhydrol.2005.06.032.

Illman, W. A., and S. P. Neuman (2001), Type-curve interpretation of a cross-hole pneumatic test in unsaturated fractured tuff, *Water Resour. Res.*, *37*(3), 583–604, doi:10.1029/2000WR900273.

Illman, W. A., and S. P. Neuman (2003), Steady-state analyses of cross-hole pneumatic injection tests in unsaturated fractured tuff, *J. Hydrol. Amsterdam*, *281*(1–2), 36–54, doi:10.1016/S0022-1694(03)00199-9.

Illman, W. A., and D. M. Tartakovsky (2006), Asymptotic analysis of cross-hole hydraulic tests in fractured granite, *Ground Water*, *44*(4), 555–563, doi:10.1111/j.1745-6584.2006.00201.x.

Illman, W. A., X. Liu, and A. Craig (2007), Steady-state hydraulic tomography in a laboratory aquifer with deterministic heterogeneity: Multi-method and multiscale validation of hydraulic conductivity tomograms, *J. Hydrol. Amsterdam*, *341*(3–4), 222–234, doi:10.1016/j.jhydrol.2007.05.011.

Illman, W. A., A. J. Craig, and X. Liu (2008), Practical issues in imaging hydraulic conductivity through hydraulic tomography, *Ground Water*, *46*(1), 120–132.

Illman, W. A., X. Liu, S. Takeuchi, T. 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.

Jankovic, I., A. Fiori, and G. Dagan (2003), Effective conductivity of an isotropic heterogeneous medium of lognormal conductivity distribution, *Multiscale Model. Simul.*, *1*(1), 40–56, doi:10.1137/S1540345902409633.

Jose, S. C., M. A. Rahman, and O. A. Cirpka (2004), Large-scale sandbox experiment on longitudinal effective dispersion in heterogeneous porous media, *Water Resour. Res.*, *40*, W12415, doi:10.1029/2004WR003363.

Kabala, Z. J. (1993), The dipole flow test: A new single-borehole test for aquifer characterization, *Water Resour. Res.*, *29*(1), 99–107, doi:10.1029/92WR01820.

Klute, A., and C. Dirksen (1986), Hydraulic conductivity and diffusivity: Laboratory methods, in *Methods of Soil Analysis, Part 1, Physical and Mineralogical Methods*, 2nd ed., edited by A. Klute, chap. 28, 687–734, Am. Soc. of Agron., Madison, Wisc.

Konikow, L. F., and J. D. Bredehoeft (1992), Ground-water models cannot be validated, *Adv. Water Resour.*, *15*(1), 75–83, doi:10.1016/0309-1708(92)90033-X.

Lee, D. M., W. D. Reynolds, and D. E. Elrick (1985), A comparison of 3 field methods for measuring saturated hydraulic conductivity, *Can. J. Soil Sci.*, *65*(3), 563–573.

Li, W., W. Nowak, and O. A. Cirpka (2005), Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown, *Water Resour. Res.*, *41*, W08403, doi:10.1029/2004WR003874.

Li, W., A. Englert, O. A. Cirpka, and H. Vereecken (2008), Three-dimensional geostatistical inversion of flowmeter and pumping test data, *Ground Water*, *46*(2), 193–201, doi:10.1111/j.1745-6584.2007.00419.x.

Liu, S., T.-C. J. Yeh, and R. Gardiner (2002), Effectiveness of hydraulic tomography: Sandbox experiments, *Water Resour. Res.*, *38*(4), 1034, doi:10.1029/2001WR000338.

Liu, X., W. A. Illman, A. J. Craig, J. Zhu, and T.-C. J. Yeh (2007), Laboratory sandbox validation of transient hydraulic tomography, *Water Resour. Res.*, *43*, W05404, doi:10.1029/2006WR005144.

Molz, F. J., R. H. Morin, A. E. Hess, J. G. Melville, and O. Guven (1989), The impeller meter for measuring aquifer permeability variations: Evaluation and comparison with other tests, *Water Resour. Res.*, *25*(7), 1677–1683, doi:10.1029/WR025i007p01677.

Neuman, S. P. (1987), Stochastic continuum representation of fractured rock permeability as an alternative to the REV and fracture network concepts, in *Rock Mechanics: Proceedings of the 28th U.S. Symposium, Tucson, Arizona*, edited by I.W. Farmer et al., pp. 533–561, A. A. Balkema, Rotterdam.

Neuman, S. P. (2005), Trends, prospects and challenges in quantifying flow and transport through fractured rocks, *Hydrogeol. J.*, *13*, 124–147, doi:10.1007/s10040-004-0397-2.

O'Connor, D. R. (2002), *Report of the Walkerton Inquiry*, Ont. Minist. of the Attorney Gen., Toronto, Ont., Canada.

Oreskes, N., K. Shraderfrechette, and K. Belitz (1994), Verification, Validation, and confirmation of numerical-models in the earth-sciences, *Science*, *263*(5147), 641–646, doi:10.1126/science.263.5147.641.

Painter, S. L., A. L. Woodbury, and Y. Jiang (2007), Transmissivity estimation for highly heterogeneous aquifers: Comparison of three methods applied to the Edwards Aquifer, Texas, USA, *Hydrogeol. J.*, *15*, 315–331, doi:10.1007/s10040-006-0071-y.

Poeter, E. P., and D. R. Anderson (2005), Multi-model ranking and inference in ground water modeling, *Ground Water*, *43*(4), 597–605, doi:10.1111/j.1745-6584.2005.0061.x.

Refsgaard, J. C. (1997), Parameterisation, calibration and validation of distributed hydrological models, *J. Hydrol. Amsterdam*, *198*(1–4), 69–97, doi:10.1016/S0022-1694(96)03329-X.

Refsgaard, J. C., and H. J. Henriksen (2004), Modelling guidelines-terminology and guiding principles, *Adv. Water Resour.*, *27*(1), 71–82, doi:10.1016/j.advwatres.2003.08.006.

Refsgaard, J. C., and J. Knudsen (1996), Operational validation and intercomparison of different types of hydrological models, *Water Resour. Res.*, *32*(7), 2189–2202, doi:10.1029/96WR00896.

Rehfeldt, K. R., J. M. Boggs, and L. W. Gelhar (1992), Field study of dispersion in a heterogeneous aquifer: 3. Geostatistical analysis of hydraulic conductivity, *Water Resour. Res.*, *28*(12), 3309–3324, doi:10.1029/92WR01758.

Renard, P., and G. de Marsily (1997), Calculating equivalent permeability: A review, *Adv. Water Resour.*, *20*(5–6), 253–278, doi:10.1016/S0309-1708(96)00050-4.

Schincariol, R., and F. Schwartz (1990), An experimental investigation of variable density flow and mixing in homogeneous and heterogeneous media, *Water Resour. Res.*, *26*(10), 2317–2329.

Shepherd, R. G. (1989), Correlations of permeability and grain size, *Ground Water*, *27*(5), 633–638, doi:10.1111/j.1745-6584.1989.tb00476.x.

Silliman, S. E. (2001), Laboratory study of chemical transport to wells within heterogeneous porous media, *Water Resour. Res.*, *37*(7), 1883–1892, doi:10.1029/2001WR900005.

Silliman, S. E., and E. S. Simpson (1987), Laboratory evidence of the scale effect in dispersion of solutes in porous media, *Water Resour. Res.*, *23*(8), 1667–1673, doi:10.1029/WR023i008p01667.

Silliman, S. E., L. Zheng, and P. Conwell (1998), The use of laboratory experiments for the study of conservative solute transport in heterogeneous porous media, *Hydrogeol. J.*, *6*(1), 166–177, doi:10.1007/s100400050142.

Straface, S., T.-C. J. Yeh, J. Zhu, S. Troisi, and C. H. Lee (2007), Sequential aquifer tests at a well field, Montalto Uffugo Scalo, Italy, *Water Resour. Res.*, *43*, W07432, doi:10.1029/2006WR005287.

Sudicky, E. A. (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*(13), 2069–2082, doi:10.1029/WR022i013p02069.

Sudicky, E. A., W. A. Illman, R. G. McLaren, I. K. Goltz, and J. J. Adams (2010), Heterogeneity in hydraulic conductivity and its role on the macroscale transport of solute plume from a landfill: From measurements to a practical application of stochastic flow and transport theory, *Water Resour. Res.*, *46*, W01508, doi:10.1029/2008WR007558.

Tartakovsky, D. M., and S. P. Neuman (1998), Transient effective hydraulic conductivities under slowly and rapidly varying mean gradients in bounded three-dimensional random media, *Water Resour. Res.*, *34*(1), 21–32, doi:10.1029/97WR01965.

Vesselinov, V. V., S. P. Neuman, and W. A. Illman (2001a), Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff: 1. Methodology and borehole effects, *Water Resour. Res.*, *37*(12), 3001–3018, doi:10.1029/2000WR000133.

Vesselinov, V. V., S. P. Neuman, and W. A. Illman (2001b), Three-dimensional numerical inversion of pneumatic cross-hole tests in unsat-

urated fractured tuff: 2. Equivalent parameters, high-resolution stochastic imaging and scale effects, *Water Resour. Res.*, *37*(12), 3019–3042, doi:10.1029/2000WR000135.

Welty, C., and M. M. Elsner (1997), Constructing correlated random fields in the laboratory for observations of fluid flow and mass transport, *J. Hydrol. Amsterdam*, *202*(1–4), 192–211, doi:10.1016/S0022-1694(97)00066-8.

Wolf, S. H., M. A. Celia, and K. M. Hess (1992), Evaluation of hydraulic conductivities calculated from multiport-permeameter measurements, *Ground Water*, *29*(4), 516–525, doi:10.1111/j.1745-6584.1991.tb00543.x.

Wu, C.-M., T.-C. J. Yeh, J. Zhu, T. H. Lee, N.-S. Hsu, C.-H. Chen, and A. F. Sancho (2005), Traditional analysis of aquifer tests: Comparing apples to oranges?, *Water Resour. Res.*, *41*, W09402, doi:10.1029/2004WR003717.

Xiang, J., T.-C. J. Yeh, C.-H. Lee, K.-C. Hsu, and J.-C. Wen (2009), A simultaneous successive linear estimator and a guide for hydraulic tomography analysis, *Water Resour. Res.*, *45*, W02432, doi:10.1029/2008WR007180.

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*(2), 116–118, doi:10.1111/j.1745-6584.2006.00292.x.

Yeh, T.-C. J., and S. Liu (2000), Hydraulic tomography: Development of a new aquifer test method, *Water Resour. Res.*, *36*(8), 2095–2105, doi:10.1029/2000WR900114.

Yeh, T.-C. J., and J. Simunek (2002), Stochastic fusion of information for characterizing and monitoring the Vadose Zone, *Vadose Zone J.*, *1*, 207–221, doi:10.2113/1.2.207.

Yeh, T.-C. J., R. Srivastava, A. Guzman, and T. Harter (1993), A numerical model for water flow and chemical transport in variably saturated porous media, *Ground Water*, *31*(4), 634–644, doi:10.1111/j.1745-6584.1993.tb00597.x.

Yeh, T.-C., M. Jin, and S. Hanna (1996), An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields, *Water Resour. Res.*, *32*(1), 85–92, doi:10.1029/95WR02869.

Yin, D., and W. A. Illman (2009), Hydraulic tomography using temporal moments of drawdown recovery data: A laboratory sandbox study, *Water Resour. Res.*, *45*, W01502, doi:10.1029/2007WR006623.

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

Zhu, J., and T.-C. J. Yeh (2006), Analysis of hydraulic tomography using temporal moments of drawdown recovery data, *Water Resour. Res.*, *42*, W02403, doi:10.1029/2005WR004309.

Zlotnik, V. A., and B. R. Zurbuchen (2003), Field study of hydraulic conductivity in a heterogeneous aquifer: Comparison of single-borehole measurements using different instruments, *Water Resour. Res.*, *39*(4), 1101, doi:10.1029/2002WR001415.

Zlotnik, V. A., B. R. Zurbuchen, and T. Ptak (2001), The steady-state dipole-flow test for characterization of hydraulic conductivity statistics in a highly permeable aquifer: Horkheimer Insel site, Germany, *Ground Water*, *39*(4), 504–516, doi:10.1111/j.1745-6584.2001.tb02339.x.

A. J. Craig, IIHR-Hydroscience and Engineering, University of Iowa, Iowa City, IA 52242, USA.

W. A. Illman, Waterloo Institute for Groundwater Research, Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, ON N2L 3G1, Canada. (willman@uwaterloo.ca)

D. Yin, 26037 Stone CYN, San Antonio, TX 78260, USA.

J. Zhu, Kentucky Geological Survey, University of Kentucky, Lexington, KY 40506, USA.