Journal of Hydrology (2007) 341, 222–234

available at www.sciencedirect.com

journal homepage: www.elsevier.com/locate/jhydrol

Steady-state hydraulic tomography in a laboratory aquifer with deterministic heterogeneity: Multi-method and multiscale validation of hydraulic conductivity tomograms

Walter A. Illman [a,b,c,d,*], Xiaoyi Liu [a,c], Andrew Craig [a]

[a] *IIHR-Hydroscience and Engineering, The University of Iowa, Iowa City, IA 52242, USA*
[b] *Department of Civil and Environmental Engineering, University of Iowa, Iowa City, IA 52242, USA*
[c] *Department of Geoscience, University of Iowa, Iowa City, IA 52242, USA*
[d] *Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ont., Canada N2L 3G1*

Received 30 September 2006; received in revised form 2 March 2007; accepted 7 May 2007

KEYWORDS
Hydraulic tomography;
Pneumatic tomography;
Inverse modeling;
Model validation;
Delineation of subsurface heterogeneity;
Scale effect

Summary Hydraulic tomography potentially is a viable technology that facilitates subsurface imaging of hydraulic heterogeneity. To date, a comprehensive validation of hydraulic tomography has not been done either at the laboratory or field scales. The main objective of this paper is to examine the accuracy of hydraulic conductivity (*K*) tomograms obtained from the steady-state hydraulic tomography algorithm of [Yeh, T.-C. J., Liu, S., 2000. Hydraulic tomography: development of a new aquifer test method. Water Resources Research 36, 2095–2105]. We first obtain a reference *K* tomogram through the inversion of synthetic cross-hole test data generated through numerical simulations. The purpose of reference *K* tomogram generation is to examine the ability of the algorithm to image the heterogeneity pattern under optimal conditions without experimental errors and with full control of forcing functions (initial and boundary conditions as well as source/sink terms). Parallel to the generation of synthetic data, we conduct hydraulic tests at multiple scales in a laboratory aquifer with deterministic heterogeneity to generate data that are used to validate *K* tomograms from hydraulic tomography. Measurements include multiple *K* estimates from core, slug, single-hole and cross-hole tests as well as several unidirectional, flow-through experiments conducted on the sandbox under steady-state conditions. Validation of *K* tomograms involved a multi-method and multiscale approach proposed herein which include: (1) visual comparisons of *K* tomograms to the true sand distributions and the reference *K* tomogram; (2) testing the ability of

* Corresponding author. Address: Department of Geoscience, University of Iowa, Iowa City, IA 52242, USA.
E-mail address: walter-illman@uiowa.edu (W.A. Illman).

0022-1694/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.jhydrol.2007.05.011

the *K* tomogram to predict the hydraulic head distribution of an independent cross-hole test not used in the computation of the *K* tomogram; (3) comparison of the conditional mean and variance of local *K* from the *K* tomograms to the sample mean and variance of results from other measurements; (4) comparison of local *K* values from *K* tomograms to those from the reference *K* tomogram; and (5) comparison of local *K* values from *K* tomograms to those obtained from cores and single-hole tests. The multi-method and multiscale validation approach proposed herein further illustrates the robustness of steady-state hydraulic tomography in subsurface heterogeneity delineation.
Published by Elsevier B.V.

Introduction

The subsurface is heterogeneous at multiple scales and is the rule rather than the exception. The knowledge of detailed three-dimensional distributions of hydraulic conductivity (*K*) is critical in prediction of contaminant transport, delineation of well catchment zones, and quantification of groundwater fluxes including surface-water–groundwater exchange.

Information about the spatial variability of flow parameters is most commonly obtained through inference from small-scale measurements of cores, slug/bail tests, flowmeter tests and single-hole pressure tests. This requires the drilling of numerous boreholes and conducting of multiple measurements at various depth intervals in each of them using sophisticated equipment. The approach is expensive and time consuming. As a consequence, little detailed site characterization has been implemented in practice.

Most recently, tomographic surveys have shown that a detailed heterogeneity distribution can be obtained from multiple pumping tests using a limited number of bore holes. For example, Illman et al. (1998); (see also Illman (1999) and Illman and Neuman (2001, 2003)) conducted a new type of pneumatic cross-hole test in unsaturated fractured tuff. The first unique aspect of these tests was that they were conducted sequentially in a tomographic manner at multiple injection points with a large number of monitoring intervals. The second unique aspect of Illman's (1999) work was the use of multiple injection and observation interval lengths amounting to cross-hole tests conducted at multiple scales. These cross-hole pneumatic tests were then analyzed by Vesselinov et al. (2001b) with a three-dimensional numerical inverse model, which amounted to a pneumatic tomography, yielding spatial distributions of permeability and porosity. They compared kriged estimates obtained via simultaneous inversion of three cross-hole tests with single-hole permeabilities determined along four boreholes by Guzman et al. (1996). The correlations were good in one borehole but not as good in three others. Vesselinov et al. (2001b) suggested that this may be due to the number of pilot points (de Marsily, 1978) used in their inversion of the cross-hole test data rather than the number of measurements available from single-hole tests.

Yeh and Liu (2000) developed a sequential geostatistical inverse method which can be applied to hydraulic tomography for the interpretation of cross-hole hydraulic tests under steady-state conditions. The main advantage of sequentially including pumping tests is its computational efficiency. The method is based on the sequential successive linear estimator (SSLE) and these authors conducted synthetic simulations for 2- and 3-dimensional cases to test their approach. Validation of the steady-state hydraulic tomography was limited to error-free cases of synthetic simulations.

Liu et al. (2002) conducted a laboratory sandbox study to evaluate the performance of hydraulic tomography in characterizing aquifer heterogeneity. This was the first validation study of hydraulic tomography, but the *K* tomograms were only visually compared to the distribution of sand types and to results from synthetic simulations. The *K* tomograms were not compared to small scale estimates of *K* directly and the authors explicitly state that the true *K* distributions were not available for either one of the sandboxes used in the study. The authors mentioned that errors and biases have an effect on their *K* tomograms, but they did not examine the role of errors and biases directly by isolating their causes.

Other researchers (Gottlieb and Dietrich, 1995; Butler et al., 1999; Bohling et al., 2002; Brauchler et al., 2003; McDermott et al., 2003; Zhu and Yeh, 2005, 2006) have developed methods for interpreting hydraulic and pneumatic tomography but none of them conducted a detailed validation of the *K* tomograms. Therefore, the main objective of this paper is to examine the accuracy of the *K* tomograms obtained from the steady-state hydraulic tomography algorithm developed by Yeh and Liu (2000). We first obtain a reference *K* tomogram through the inversion of synthetic cross-hole test data generated through numerical simulations. The purpose of reference *K* tomogram generation is to examine the ability of the algorithm to image the heterogeneity pattern under optimal conditions without experimental errors and with full control of forcing functions (initial and boundary conditions as well as source/sink terms). Parallel to the generation of synthetic data, we conduct hydraulic tests at multiple scales in a laboratory aquifer with deterministic heterogeneity to generate data that are used to validate the *K* tomogram from hydraulic tomography. Measurements include multiple *K* estimates from core, slug, single-hole and cross-hole tests as well as several unidirectional, flow-through experiments conducted upon the sandbox under steady-state flow conditions. Validation of *K* tomograms involved a multi-method and multiscale approach which included: (1) visual comparisons of experimental *K* tomograms (from now on *K* tomogram) to the true sand distributions and the reference *K* tomogram generated using synthetic pumping test data via numerical simulations (from now on reference *K* tomogram); (2) testing the ability of the *K* tomogram to predict the hydraulic head distribution of an independent cross-hole test not used in the computation of the *K* tomogram; (3) comparison of

the conditional mean and variance of local K from the K tomograms to the sample mean and variance of results from other measurements; (4) comparison of local K values from K tomograms to those from the reference K tomogram; and (5) comparison of local K values from K tomograms to those obtained from cores and single-hole tests. We also examine the influence of errors and biases on inversion results using forward and inverse simulations of cross-hole tests. Errors and biases arising from the conduct of experiments are usually nonexistent in synthetic simulations as implemented by Yeh and Liu (2000) and Zhu and Yeh (2005, 2006), so one cannot directly examine their influence. On the other hand, in the field, biases and errors are generally unknown, so their influence on the K tomograms cannot be quantified. Therefore, laboratory sandbox studies can be very important in quantifying the role of errors and biases on the K tomograms because all the forcing functions (initial/boundary conditions; sources/sink terms) can be controlled.

We discuss the sandbox used in the study, provide descriptions of various hydraulic tests conducted in the sandbox for characterization, and discuss methods used to obtain data that will be later used to validate the K tomograms. We then discuss the forward and inverse analyses of cross-hole tests, including descriptions of data diagnostic tools, inverse modeling results with and without experimental bias and a multi-method/multiscale approach in validating the K tomograms.

Sandbox description

The synthetic heterogeneous aquifer constructed in the sandbox was designed to validate various fluid flow and solute transport algorithms and in particular, the hydraulic tomography algorithm. 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. As the wells have a diameter of 1.3 cm, wall friction effects are considered negligible (at an order of 0.0009 cm) on hydraulic tests conducted in the sandbox. 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 port and used as a water sampling port. Fig. 1 is a computer aided design (CAD) drawing of the sandbox frontal view, showing the 48 port and pressure transducer locations as well as water reservoirs for controlling hydraulic head.

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 compared to the situation if we had to maintain an external reservoir alone. The boundary head levels can be easily adjusted to be equal or to create a desired hydraulic gradient. We also utilized an intermediate overflow device to maintain equal constant heads on both boundaries in the experiments. The device consisted of a reservoir with an overflow pipe and tubing connecting it to the inlets at the bottom of the constant head reservoirs. The developed system is also 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.

Four different commercially sieved sands [20/30 and 4030, *U.S. Silica*; F-75 and F-85, *Unimin Corporation*] were used to pack the sandbox by hand. A special packing tool was developed to achieve uniform compaction. The sand was wetted from the bottom and the water levels were increased while packing. It was of paramount importance to pack the sandbox with a known K distribution in order to validate the computed tomogram. The chosen heterogeneity structure is complex enough so that it provides adequate testing of various numerical codes. Fig. 2 is a photograph

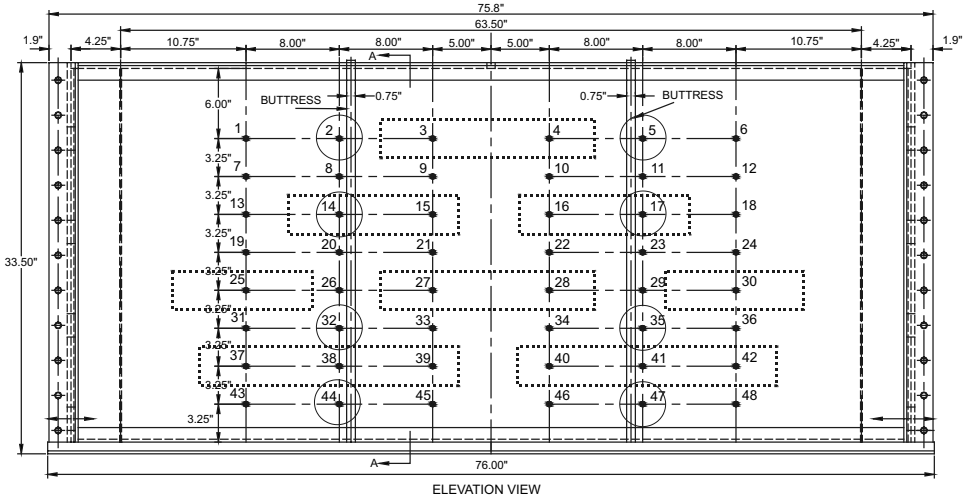

Figure 1 Computer aided design (CAD) drawing of sandbox used for the validation of hydraulic tomography. Open circles indicate pumped port locations.

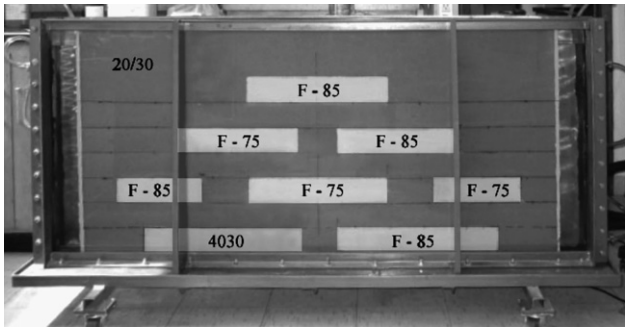

Figure 2 Photograph of the sandbox with a deterministic heterogeneous aquifer.

of the sandbox with the synthetic heterogeneous aquifer showing the different sand types and their distributions.

Discontinuous sand bodies were packed to simulate a heterogeneous aquifer to test the ability of the inverse algorithm to delineate sharp contrasts in K. It is also important to note that the sand distribution was designed to test the effectiveness of hydraulic tomography to characterize multiple sand grain sizes. Liu et al. (2002) used only two grain sizes to test the effectiveness of the code so the inclusion of four distinct grain sizes would be a stronger test.

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—1 psi and a proof pressure of 2 psi, 48 of which measured hydraulic head (h) 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 flow cell. 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 assembled at the IIHR-Hydroscience & Engineering of the University of Iowa. The third component was a dedicated PC with *National Instruments LabVIEW* software for automated data acquisition. Further details of the sandbox construction and experimental results can be found in Craig (2005).

Sandbox aquifer characterization

Different hydraulic tests were performed in the sandbox to characterize the hydrogeologic parameters at multiple scales. We first determined the K of the four types of sands from the extracted core using a constant head permeameter (Klute and Dirksen, 1986). Because the extracted core was very small, a special permeameter was built for testing purposes. We also conducted slug tests at each of the 48 ports. Due to the small size and configuration of the ports on the sandbox, an external well was attached to the ports instead of boring vertical wells into the sandbox.

After completing the slug tests, cross-hole tests were conducted at each of the 46 ports. Two ports were damaged and so we did not conduct cross-hole tests in these ports. The cross-hole tests were conducted by raising the h in the constant head reservoirs high enough so that water was ponded above the sand. The intermediate overflow device constantly supplied water so that constant h conditions were maintained at the top and the two side boundaries, while the bottom boundary remained a no-flow boundary.

We also conducted nine unidirectional flow-through experiments through the entire sandbox to obtain the effective hydraulic conductivity (K_{eff}) of the entire sandbox under steady-state flow conditions. These experiments were conducted by establishing a constant hydraulic gradient and measuring the discharge through the flow cell.

Description of the forward and inverse models

Forward model

Forward modeling of various hydraulic tests in the sandbox were done using the two-dimensional water flow and solute transport code MMOC2 as well as its three-dimensional version MMOC3 (Yeh et al., 1993). The forward model is capable of simulating water flow and chemical transport through variably saturated porous media. The flow equation is solved using the Galerkin finite element technique with either the Picard or Newton—Raphson iteration scheme.

We also used VSAFT2, a Graphical User Interface (GUI) program based on MMOC2, to obtain equivalent estimates of hydraulic parameters through manual calibration from the slug, single-hole, and cross-hole tests.

Inverse model

Inverse modeling of cross-hole tests in the sandbox were 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 sequential successive linear estimator (SSLE) is employed for each cross-hole test. The estimator begins by cokriging the initial K value determined and observed h 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. We select an initial K obtained from the traditional analysis of pumping test treating the medium to be homogeneous. Cokriging does not take full advantage of the observed h values because it assumes a linear relationship (Yeh and Liu, 2000) between h and K, while the true relationship is nonlinear. To circumvent this problem, a linear estimator based on the differences between the simulated and observed h values is successively employed to improve the estimate.

The second step of Yeh and Liu's (2000) approach is to use the h datasets sequentially instead of simultaneously including them in the inverse model. In essence, the sequential approach uses the estimated K field and covariances, conditioned on previous sets of h 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 the present 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 in Zhu and Yeh (2005).

Estimates of hydraulic conductivity for validation of *K* tomograms

We first determined the *K* of the four types of sands from the horizontal cores obtained during the completion of wells and port placement. The extracted cores had dimensions of 1.28 cm in diameter and 10.16 cm in length. These cores were then attached to a custom-made constant head permeameter (Klute and Dirksen, 1986) for determination of *K*. Details to the core extraction method and the design of the constant head permeameter is provided in Craig (2005). The *K* values from cores are calculated using Darcy's law.

We also conducted slug tests at each of the 48 ports. Due to the small size and configuration of the ports on the sandbox, an external well was attached to the ports instead of boring vertical wells into the sandbox. A slug was introduced to perturb the water level in the horizontal well connected to the port and the corresponding recovery was monitored using a pressure transducer. Because existing analytical solutions cannot be used to interpret the slug tests with our current setting, we analyzed the data by manually calibrating VSAFT2 (Yeh et al., 1993), available at www.hwr.arizona.edu/yeh, by treating the model domain to be a two-dimensional, homogeneous medium (Craig, 2005). A fine numerical grid (1.64 cm by 1.64 cm) was developed for the slug test analysis. The numerical simulations were conducted by raising the initial head at the elements corresponding to the slugged port and monitoring the corresponding decay in the head profile. VSAFT2 was chosen to analyze the test data for consistency because the code contains the forward model used later for hydraulic tomography. We report the geometric mean of 40 values that we deem to match the observed data well in Table 1. Results obtained revealed that the *K* values are several orders of magnitude smaller than the core values. We suspected that the data are affected by skin effects and wellbore storage. In fact, we investigated the issue further by conducting additional experiments to examine the effects of the number of cuts on the head response to slug tests. In particular, slug tests were conducted in a separate flow cell with tubes consisting of different number of cuts (2–8). This effort revealed that the head response stabilizes after six cuts were made on the well. Therefore, all wells in the sandbox discussed in this paper were made by making six cuts. Despite these efforts, the *K* values determined using slug tests are very low, thus we question their reliability and do not use them for validation purposes.

Table 1 Summary of hydraulic properties determined from core, slug, single-hole, cross-hole pumping test data and flow-through experiments

Test type	*N*	$\ln - K$	
		$\overline{\ln - K}$(K cm s^{-1})	$\sigma^2_{\ln - K}$
Core	48	-2.166 (1.146×10^{-1})[a]	1.498[a]
Slug	40	-10.692 (2.272×10^{-5})	0.431
Single-hole	9	-3.174 (4.182×10^{-2})	0.570
Cross-hole	96	-1.757 (1.726×10^{-1})	0.074
Flow-through	9	-1.757 (1.725×10^{-1})	0.002

[a] The volume-weighted mean and its corresponding variance are -1.920 (1.467×10^{-1} cm s^{-1}) and 1.560, respectively.

We then conducted pumping tests at each of the 46 out of available 48 ports. Ports 36 and 38 have been damaged so we do not pump from these ports. During the pumping tests, the top and two sides of the aquifer served as constant head boundaries, as described earlier, while the bottom remained a no-flow boundary. Pumping rates ranged from 150 to 190 mL/min in most cases. For each test, data collection started before the pump was activated to obtain the initial hydraulic head in the sandbox and the data were collected from all ports every 0.75 s set to be constant throughout the duration of each experiment. A data collection interval of 0.75 s was selected to allow for the expected rapid transient change in hydraulic head at the monitoring ports. A peristaltic pump was then activated at the pumping port and allowed to run until the development of steady-state flow conditions. The pump was then shut off to collect recovery data until the hydraulic head recovered fully. During each pumping test, pressure heads were collected at all 48 ports.

The datasets were analyzed in several ways. First, we analyzed the 48 drawdown-time datasets induced by pumping at port 22 and those caused by pumping at port 28 by manually calibrating VSAFT 2 and assuming the aquifer is homogeneous. The numerical setup for the manual calibration is identical to the slug test analysis. For the pumping test at port 28 (located in 20/30 sand), all 47 cross-hole intervals were matched and 1 single-hole match was made which yielded a total of 48 estimates for that pumping test. The pumping test at port 22 (located in F-75 sand) yielded also 47 cross-hole matches for observed and simulated drawdown. The single-hole match in this case was unattainable due to the very large drawdown, which VSAFT2 could not simulate. Analysis of the two pumping tests thus yielded 95 estimates of *K* and S_s for the equivalent homogeneous medium. These two tests will be denoted as cross-hole tests hereafter.

Out of the 46 pumping tests, we analyzed the drawdown-time data at selected 9 pumping ports (2, 5, 14, 17, 32, 35, 44, 46, and 47) using VSAFT2 to yield local or single-hole estimates of *K*. These results are denoted as the single-hole results.

We also conducted nine flow-through experiments through the entire sandbox to obtain the effective hydraulic conductivity (K_{eff}) of the entire sandbox under steady-state unidirectional flow conditions. Specifically, each of these nine experiments was conducted by changing the height of the 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. The nine pairs of gradient and discharge were computed using Darcy's law to obtain the K_{eff}.

Table 1 summarizes the results from all these tests. The mean estimates were obtained by computing the arithmetic

mean of the natural logarithm transformed data. The variance was likewise computed using the natural logarithm transformed data set. We also calculated a volume-weighted mean and variance of the core values which are also listed in Table 1. The purpose of computing the volume-weighted mean and variance of the core K values was so that these values are upscaled to the size of the finite element grid used for the inversion so that we can compare them later.

In Table 1, we see that, in general, the mean values of the cross-hole and flow-through values coincide in this sandbox, however, core, slug, and single-hole test values are noticeably smaller suggesting a scale effect. As mentioned earlier, the slug test values are considerably smaller, so we conclude that near well effects and/or borehole storage dominate the response, causing K estimates to be less reliable in comparison to other measurements. However, K estimates from cross-hole tests in the observation well are very close to the overall K value derived from the flow-through experiments suggesting that these estimates are less affected by near well effects. Therefore, we conclude that the cross-hole observation well data are reliable and we retain them in our analysis.

Examination of Table 1 also shows that $\sigma^2_{\ln K}$ varies from one type of test to the next with variance decreasing with the increasing scale. This is because the support volume of each estimate increases from the core, slug, single-hole, cross-hole and flow-through experiments. As the sample volume increases, K is averaged over the investigated volume. We note that the calculated K values when the medium is treated to be homogeneous are useful, but provide a very limited resolution of the spatial variability in K. In addition, estimation of equivalent parameters treating the medium to be homogeneous may yield biased values. This is one important reason why we conduct hydraulic tomography to determine how the K values vary spatially.

Data diagnostics

Examination of hydraulic head records and their animations

Prior to inverse modeling of cross-hole hydraulic tests, we conducted a detailed diagnostic study of the data. Such diagnostic tests of data used in forward and inverse models are rarely discussed in the literature, but we found that it should be an integral component to all phases of numerical forward and inverse modeling of cross-hole tests as the use of data corrupted by noise can have a profound effect on model results.

We first plotted the transient head records in all 48 ports including the two pressure transducers placed in the constant head reservoirs to examine the propagation of a pressure pulse throughout the aquifer. Plotting of h records in this manner also allowed us to examine whether the pressure transducers were functioning properly as well as to identify the magnitude of noise during a given cross-hole test. The initial diagnosis of the data revealed that pressure transducers can be subject to different noise sources including electrical interference, barometric effects, and minute uncontrollable variations of the water supply. In general, the noise can be removed through signal conditioning and de-trending procedures applied to raw data.

We also contoured the initial head distribution within the sandbox to identify whether there was any water flow prior to the cross-hole tests. This also allowed us to study the presence of any drift in pressure transducers. Because the transient head record at a given monitoring port provides only limited information about the evolution of the pressure pulse through the aquifer during a given pumping test, we also made animations of head contours using all head records from all 48 ports during a given cross-hole test by plotting successive frames of head distributions over the duration of each test. This process ensured that each test was conducted correctly and gave a pictorial representation of the pressure propagation throughout the sandbox during a given test.

Diagnostic forward modeling

We next conducted forward modeling of cross-hole tests to further diagnose the available data. We assumed that the pumping rate (Q) was deterministic and was accurately measured, the pressure transducers were properly calibrated and the drift was removed, the boundary conditions remained stable throughout the experiments and there was no noise to affect the experimental data. The K values used in the forward model are those obtained from taking the mean value from the core for each type of sand. With the forward model, we then simulate each cross-hole test and compare the results from the synthetic to the real data through a scatter plot.

The forward modeling of the cross-hole tests showed that the simulated hydraulic head (h_s) values are in general higher than the measured hydraulic head (h_m) values near the top and bottom of the sandbox. The head value at the pumped port also differs considerably from the simulation results, consistently throughout the sandbox. As we are not aware of the cause(s) of these biases, we conducted a detailed study to determine their cause. A bias can be introduced due to the collection of inaccurate data or through the misapplication of the forward model.

We first examined all head records carefully. This showed that the initial heads are inconsistent indicating the presence of drift in pressure transducers, so we modified the test data to reduce this bias. In particular we accounted for pressure transducer drift by calculating the drawdown (s_i) at port i, in the following manner:

$$s_i = h_{i,0} - h_i \tag{1}$$

where $h_{i,0}$ was the initial head at port i during a given cross-hole test and h_i is head at port i at time t. We then averaged the starting h_i by taking the arithmetic mean

$$\overline{h_0} = \frac{1}{n}\sum_{i=1}^{n} h_{i,0} \tag{2}$$

and used this value as the initial head for all ports. We then subtracted the s_i to this starting head to get the modified head for each of the pumping tests.

We also see that the h_s are considerably higher than the h_m values at the pumping port. The extra head drop can be due to inertial effects due to a high pumping rate and/or the development of a low K region at the pumped port (i.e., skin effect). As discussed earlier, a series of diagnostic tests not

Table 2 L_2 norms for various forward run scenarios of cross-hole tests used in steady-state hydraulic tomography and its validation

Pumping port number	L2	
	Original data	After data modification
2	6.23E−04	1.31E−04
5	2.94E−04	1.49E−04
14	5.15E−04	1.47E−03
17	7.94E−04	6.15E−04
32	1.26E−02	1.99E−02
35	2.95E−02	7.87E−03
44	3.90E−03	8.96E−03
46	5.49E−02	1.53E−02
47	7.95E−03	1.10E−02

shown here showed that the number of cuts made in the brass tube can affect the K measured at the pumped port and can be responsible for the increased drawdown. Therefore, we tried to account for the skin effect by locally changing the K of the element to the value determined from the slug tests, which was several orders of magnitude lower than the values determined through other methods. Results of the simulations accounting for the skin effect at the pumped port showed that the h_s values are closer to the h_m values at the pumping port. We also modified the K of all observation ports to account for this skin effect by artificially lowering the K to the value determined from the slug tests. However, the bias could not be removed completely so we made a decision not to use the data from the pumping port, but use all original data from the observation intervals without the adjustment of local K in the inverse model.

To evaluate the goodness-of-fit between the simulated and measured hydraulic head responses, we calculate the mean squared error norm (L_2):

$$L_2 = \frac{1}{n}\sum_{i=1}^{n}(h_{s,i} - h_{m,i})^2 \quad (3)$$

where $h_{s,i}$ is the simulated value of hydraulic head at port i and $h_{m,i}$ is the measured value of hydraulic head at port i. The smaller the L_2 norm, the better the estimate is expected to be. Table 2 summarizes the L_2 norm for each cross-hole test analyzed showing that forward simulations using modified data (modified only for initial head) reveal a large reduction in the L_2 norm in most of the tests. In summary, diagnostic plots, animations of head contours, and forward modeling runs have helped us identify and remove errors/biases in our experiments.

Inputs to the inverse model

To obtain a K tomogram from multiple cross-hole tests, we solve a 3D inverse problem for steady flow conditions. The sandbox was discretized into 741 elements and 1600 nodes with element dimensions of 4.1 cm × 10.2 cm × 4.1 cm. Both sides and the top boundary were set to the same constant head boundary condition, while the bottom boundary of the sandbox was a no-flow boundary. We solve the inverse problem using a consistent grid for both the synthetic and real cases. Here, the synthetic case means that we generate a set of pumping test data by running a series of steady-state forward simulations using MMOC3 with the grid blocks representing the discontinuous blocks of sands in the sandbox. We then use these head and discharge records at the pumping point and observation points in the steady-state hydraulic tomography code of Yeh and Liu (2000). For the real case, we mean the inverse modeling of data collected from the actual cross-hole tests conducted in the sandbox.

Inputs to the inverse model include the initial estimate of K_{eff}, the variance $(\sigma^2_{\ln K})$ and the correlation scales of hydraulic conductivity $(\lambda_x, \lambda_y, \lambda_z)$, h_i, volumetric discharge (Q_n), where n is the test number, and available point (small-scale) measurements of hydraulic conductivity (K_c, K_{SH}). Here, we do not use the available point scale measurements of K to test the ability of the algorithm to delineate the heterogeneity patterns.

We obtained the initial estimate of K_{eff} separately for the synthetic and real cases. For the synthetic case, we simply take the geometric mean of the small scale or local estimates of the material from the core experiments at the sampling ports. For the inversion of the real cross-hole hydraulic test data, a number of approaches can be used to obtain the initial estimate of K_{eff} for the medium. One is to calculate the geometric mean of the available small scale data (i.e., core, slug, and single-hole data). The second approach is to do the same for the K_{eq} estimates obtained through the analysis of data from monitoring ports during cross-hole tests by treating the heterogeneous medium to be homogeneous. We also have the results from the flow-through experiments to obtain the K_{eff}. We select the K_{eq} obtained through the traditional analysis of cross-hole tests by treating the medium to be homogeneous because in practice, this is the most logical approach in obtaining a large scale K that should be representative of a large portion of the flow and simulation domain.

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 obtain $\sigma^2_{\ln K}$ estimates from the available small scale data and use this as our input $\sigma^2_{\ln K}$ in the inverse model for the real data set. For the inversion of the synthetic data set, we calculate the $\sigma^2_{\ln K}$ from the local values input into the forward model.

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. Table 3 lists values of these statistical parameters ($\sigma^2_{\ln K}$, $\lambda_x, \lambda_y, \lambda_z$) used in our inverse analysis of cross-hole hydraulic tests for the inversion of synthetic and real cross-hole tests.

Table 3 Input data for inverse modeling of eight pumping tests in sandbox

Test type	K_{eff} (cm s^{-1})	σ_f^2	λ_x (cm)	λ_y (cm)	λ_z (cm)	Covariance model for f	Q (cm^3 s^{-1})
Synthetic	0.19	2.0	30	10	10	Exponential	2.9–3.17
Real	0.17	2.0	30	10	10	Exponential	2.92–3.17

Results

Inverse modeling of synthetic cross-hole tests in sandbox

We begin the inversion of cross-hole tests by first generating synthetic hydraulic tomography data and inverting them to generate a reference K tomogram of the sandbox so that we can compare the results of the inversion of the real data later. This case illustrates the ability of the steady-state hydraulic tomography algorithm of Yeh and Liu (2000) to obtain a K tomogram under idealized conditions where model and measurement errors are excluded and when forcing functions are fully controlled.

All 48 ports (see Fig. 1 for locations) were used for steady-state hydraulic tomography. The steady-state head data were collected at these wells. Fig. 3a is the true synthetic K field used to generate the cross-hole test data employed for hydraulic tomography. Fig. 3b–i (cases 1–8) then shows the results of the estimation of the K tomogram from the successive inclusion of test data from pumping tests at ports 47, 44, 35, 32, 17, 14, 5, and 2, conducted in that order. These results clearly show that the inversion algorithm is capable of capturing the low K blocks, and other details of aquifer heterogeneity such as windows in low K strata that could provide continuous pathways for contaminant transport. It is of interest to note that the synthetic aquifer has a K distribution that is non-Gaussian and nonstationary. Because Yeh and Liu's (2000) algorithm assumes a Gaussian and a stationary field, one would not expect this approach to be applicable to this K distribution. However, the statistical assumptions inherent in the algorithm become less important as we include large number of observation well datasets in the hydraulic tomography algorithm. This fact is evident from our results. Therefore, Yeh and Liu's (2000) SSLE is not limited to Gaussian and stationary random K fields.

The results were also quantitatively evaluated using the L_2 norm, defined earlier, but in this case for the estimated and true parameters for the entire computational domain.

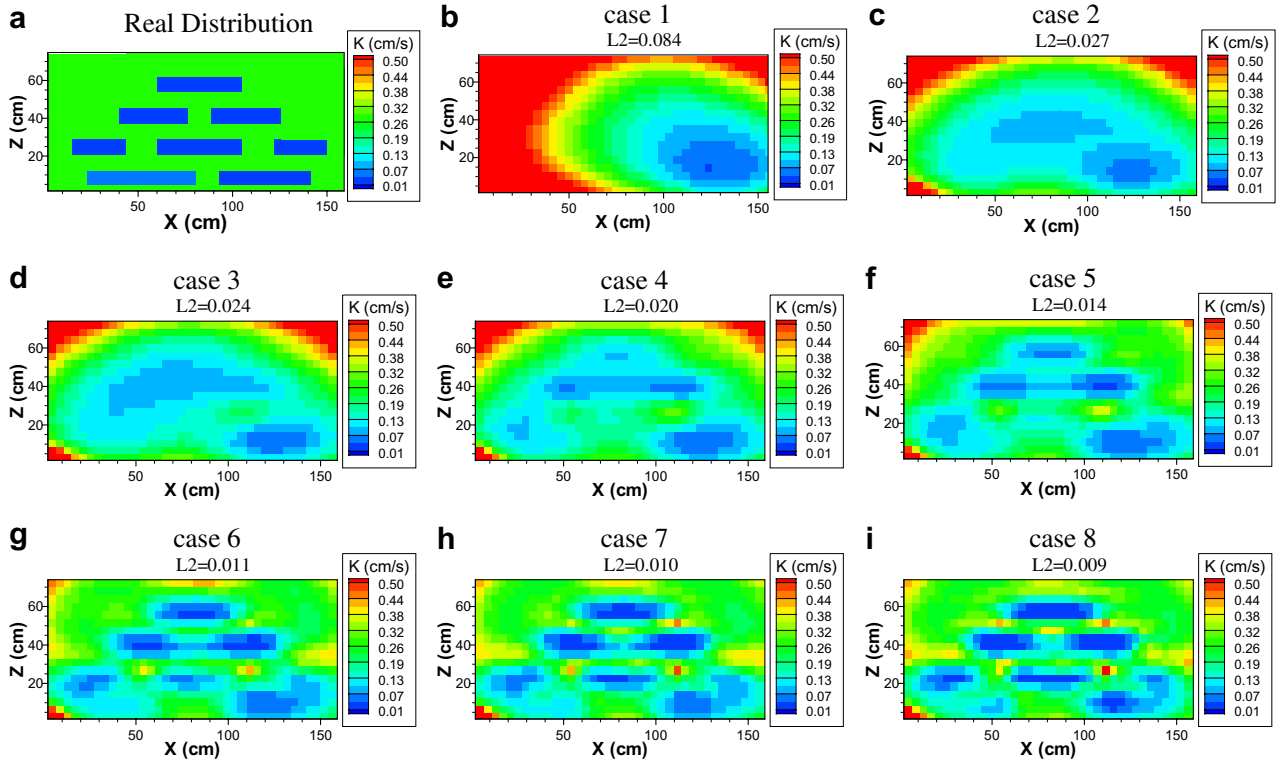

Figure 3 (a) The synthetic true K distribution used to generate synthetic cross-hole test data; (b) (47 – case 1) (c) (47, 44 – case 2); (d) (47, 44, 35 – case 3); (e) (47, 44, 35, 32 – case 4); (f) (47, 44, 35, 32, 17 – case 5); (g) (47, 44, 35, 32, 17, 14 – case 6); (h) (47, 44, 35, 32, 17, 14, 5 – case 7); and (i) (47, 44, 35, 32, 17, 14, 5, 2 – case 8) show the resulting K tomograms by sequentially inverting the synthetic cross-hole test data. Numbers in parentheses from (b) to (i) indicate the port numbers (Fig. 1) used as the pumped well for each cross-hole test.

Fig. 3b–i shows that the L_2 norm decreases as more pumping tests are added, but the rate of reduction diminishes and stabilizes through cases 6–8.

We use the best synthetic result (case 8) as a reference K tomogram in which we later compare our tomograms from the inversion of real cross-hole test data.

Inverse modeling of real cross-hole tests in sandbox

We next examine the results from the inversion of real cross-hole test data obtained in identical fashion to the synthetic case. In reality we do not know the true K distribution as in the synthetic case, so we do not show a true K distribution as in Fig. 3a. Instead, we examine how the computed K tomogram improves after the identified errors and biases are sequentially removed. Also, the L_2 norm is computed only at the ports in which the estimated K is compared to the true K from available cores. Fig. 4a–d shows the sequential improvement of the computed tomograms as we remove the errors and biases as discussed in the previous section. Specifically, Fig. 4a is the tomogram of the original data set from pumping taking place at 8 ports 47, 44, 35, 32, 17, 14, 5, and 2 without removing the errors and biases (case 9). Clearly, these results do not compare favorably with the reference K tomogram (case 8).

Fig. 4b shows the tomogram after removal of pumped well data affected by skin effect (case 10) and Fig. 4c shows the tomogram after the h measurements are corrected for variations in the drift or offset in the pressure transducers (case 11). We see a progressive improvement in the computed tomograms as the low K blocks begin to emerge. Therefore removal of the pumped well data from the inverse model improves the quality of the computed tomogram. In contrast, the skin effect at the observation wells does not have a significant impact on the results of the steady-state hydraulic tomography. The effect of initial head variations due to the drift in pressure transducers also does not have a large effect on the inversion results for this particular case.

The largest effect on the inversion results seems to come from the slight variations in the boundary condition (~1–2 mm differences in water level in the constant head reservoirs) from one pumping test to the next. When we account for the varying boundary condition for each test, the results improve dramatically (Fig. 4d, case 12). This suggests that steady-state hydraulic tomography is very sensitive to the boundary conditions that are input into the model and that it needs to be studied carefully in the field.

Multi-method and multiscale validation of K tomograms

Model validation in the traditional sense means that the accuracy and predictive capability [of a model] can be shown to lie within acceptable limits of error by test data that are independent of the calibration data (Konikow, 1978). In other words, model validation is done through the comparison of simulated and measured heads for both forward and inverse modeling efforts. The comparison of head/K data alone may not be a robust approach in validating the inverse model, so we instead utilize a multi-method and multiscale approach developed herein. The multi-method

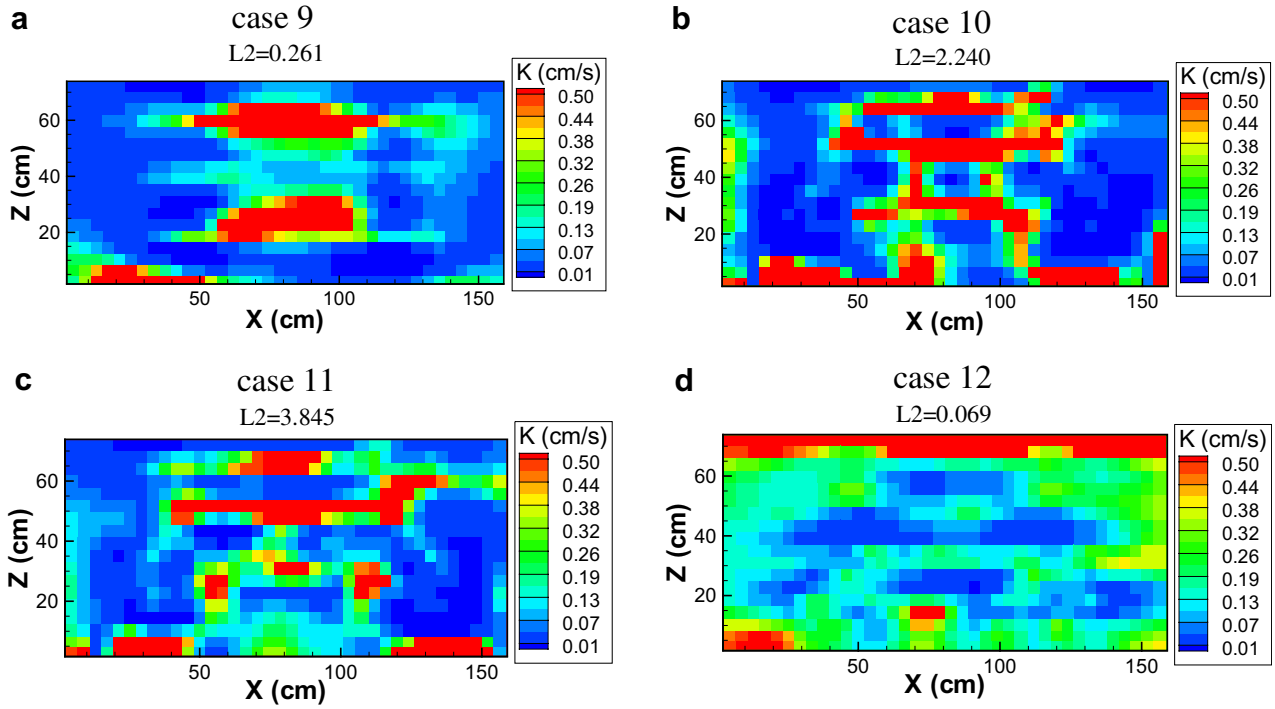

Figure 4 Resulting K tomogram with: (a) original data set without removing bias/errors (case 9); (b) after removal of pumped well data affected by skin effect (case 10); (c) hydraulic head measurements corrected for variations in the offset (case 11); (d) accounting for varying boundary conditions for each pumping test (case 12). All results shown here are after including eight pumping tests in the inversion as in case 8 described in Fig. 3i.

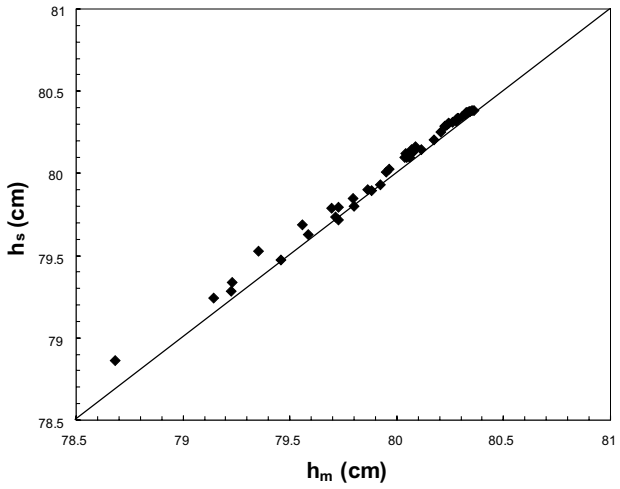

Figure 5 Scatter plot of simulated hydraulic head (h_s) versus measured hydraulic head (h_m). The h_s values were obtained by simulating the cross-hole test conducted at port 46 using the *K* tomogram (Fig. 4d). The h_m values are from an actual cross-hole test conducted at port 46.

Table 4 Conditional mean log-transformed hydraulic conductivity ($\overline{\ln K}$) and conditional variance of the log-transformed hydraulic conductivity ($\sigma^2_{\ln K}$) estimates for the real inversions with number of cross-hole tests used in the analysis

Number of cross-hole tests included in analysis (test number)	$\overline{\ln K}$(K cm s^{-1})	$\sigma^2_{\ln K}$
1 (47)	−1.71 (0.18)	0.34
2 (47 + 44)	−1.81 (0.16)	0.44
3 (47 + 44 + 35)	−1.79 (0.17)	0.60
4 (47 + 44 + 35 + 32)	−1.78 (0.17)	0.80
5 (47 + 44 + 35 + 32 + 17)	−1.76 (0.17)	0.96
6 (47 + 44 + 35 + 32 + 17 + 14)	−1.74 (0.18)	1.16
7 (47 + 44 + 35 + 32 + 17 + 14 + 5)	−1.72 (0.18)	1.35
8 (47 + 44 + 35 + 32 + 17 + 14 + 5 + 2)	−1.73 (0.18)	1.37

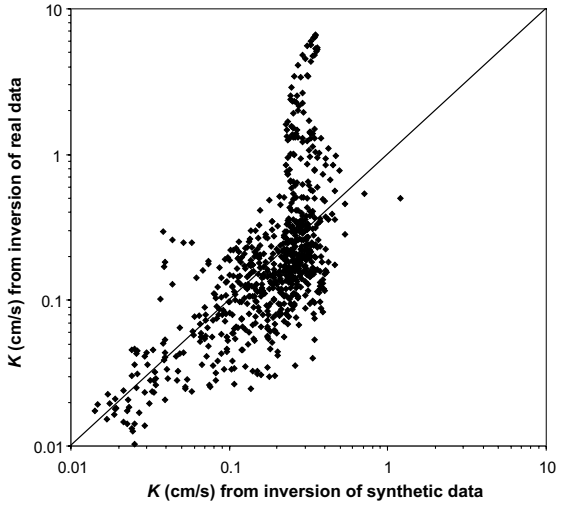

Figure 6 Scatter plot of *K* values from the real (case 12) and the reference (case 8) *K* tomograms.

validation approach consists of validation through different, yet complementary approaches in validating the resulting *K* tomogram. One approach is to visually examine whether the *K* tomogram resembles the true *K* distribution from the sandbox. Another validation approach is to simulate an independent cross-hole test not used in the construction of the *K* tomogram. The most direct validation is achieved through the comparison of the local *K* values obtained from the *K* tomograms to those obtained by independent experimental results such as *K* values from cores and single-hole tests. The cross-hole and flow-through experiments provide equivalent (K_{eq}) and effective (K_{eff}) values, which can also be compared by averaging the local *K* values from the tomograms. Because the independent experimental data are collected at different spatial scales, we consider the comparison of local and averaged *K* values from the *K* tomograms to independent datasets to be validation conducted at different spatial scales. We consider the *K* tomogram to be validated when it can be validated by employing all of these methods utilizing various datasets collected at multiple spatial scales.

Multi-method validation approach 1: visual comparison

A visual comparison of the *K* tomogram (Fig. 4d) obtained by sequentially including eight tests into the inversion algorithm and the photograph of the sand bodies (Fig. 2) shows that the steady-state hydraulic tomography is able to delineate the major low and high *K* features that comprise the aquifer heterogeneity. This includes the correct delineation of the morphology and positions of low *K* blocks/layers and stratigraphic windows which can allow contaminant migration. We also visually compare the *K* tomogram resulting from the real datasets (Fig. 4d) to the synthetic *K* tomogram (Fig. 3i). This comparison shows that the *K* near the top of the aquifer for Fig. 4d is higher than that in Fig. 3i. This is likely due to the lack of sand compaction in the sandbox, artificially inflating the *K* values in Fig. 4d. In general, it is evident from the comparisons that the approach is able to correctly depict major heterogeneity features using the hydraulic tomography technology.

Multi-method validation approach 2: validation of *K* tomogram with additional cross-hole hydraulic tests not used in the inversion

Another method of validation is to simulate an independent cross-hole test that has not been previously used to calculate the *K* tomogram and examine whether the hydraulic head of this independent test can be predicted accurately. For this, we utilize the *K* tomogram obtained from the inversion of eight cross-hole tests (Fig. 4d) and simulate a cross-hole test with pumping taking place at port 46. Fig. 5 shows the results of comparing the h_s measurements obtained from a synthetic cross-hole test conducted at port 46 and the h_m from these tests. Results show that the h_s is slightly higher than the h_m. However, the comparison is very good considering that this is an independent cross-hole test not used in the construction of the *K* tomogram.

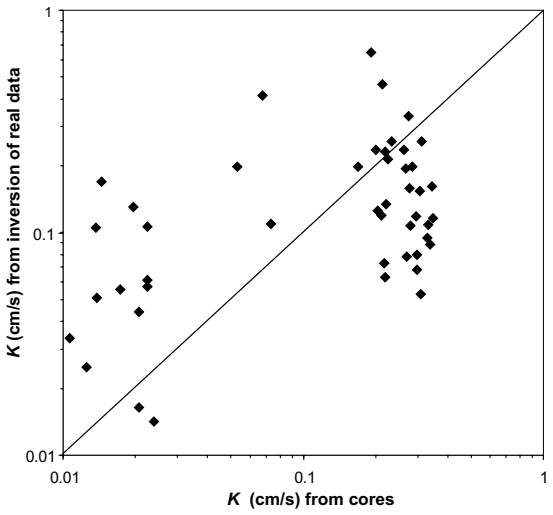

Figure 7 Scatter plot of *K* values from the *K* tomogram (case 12) at the observation point and core *K* estimates.

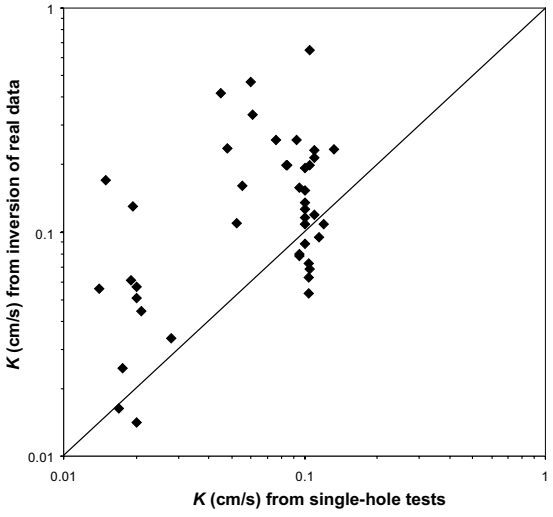

Figure 8 Scatter plot of *K* values from the *K* tomogram (case 12) at the observation point and single-hole *K* estimates.

Multi-method validation approach 3: comparison of statistical moments

We next validate the *K* tomogram by comparing the conditional mean log-transformed hydraulic conductivity $\overline{\ln K}$ and conditional log-transformed hydraulic conductivity variance $(\sigma^2_{\ln K})$ obtained from the entire *K* field in the tomogram to the corresponding statistical moments obtained from other validation datasets. We first compare the conditional $\overline{\ln K} = -1.73$ obtained from our best *K* tomogram (Fig. 4d, Table 4) to the $\ln K_{\text{eff}}$ (−1.76) obtained from the flow-through experiment (Table 1). This comparison is excellent showing that the inverse model is able to accurately and uniquely obtain the conditional mean value after including eight pumping tests in the inversion. We note that the difference between the conditional mean value and the actual mean value decreases quickly from the addition of one cross-hole test to the next (Table 4). We find that this result holds even when the initial value of *K* used in the SLE is varied.

The same cannot be said about the conditional $\sigma^2_{\ln K}$. It takes a larger number of pumping tests for a reasonable value to be estimated. To compare the conditional variance obtained from the inversion, we use the estimates of sample variance obtained from the available core *K* data (1.50). The latter may be slightly higher than the actual value for the entire population (1.37) because we only use 48 values to compute the sample variance as opposed to the conditional $\sigma^2_{\ln K}$ from the *K* tomogram where we use all values in the computational domain. This comparison shows that the inverse approach is able to estimate the conditional variance quite well after the sequential analysis of eight cross-hole tests. We consider this comparison to be quite good based on the fact that we do not have the variance of the true *K* field and that the conditional variance approaches the sample variance as more cross-hole test data are included (see Table 4). Other estimates of variances listed in Table 1 are representative of the medium at the larger scale, thus do not provide a fair comparison. Therefore, we restrict the comparison to the variance estimate from core samples.

Multi-method validation approach 4: comparison of local *K* values from the *K* tomogram to the reference *K* tomogram

We next compare the *K* tomogram to those obtained from the synthetic tests that we consider to be the reference *K* tomogram (Fig. 3i, case 8). As mentioned earlier, the reference *K* tomogram is computed from synthetic cross-hole tests on the computer. The reference *K* tomogram is our best *K* tomogram that can be obtained from the steady-state hydraulic tomography approach of Yeh and Liu (2000) under optimal conditions without experimental errors and with full control of forcing functions (initial and boundary conditions as well as source/sink terms). Therefore, one approach to validate the *K* tomogram is to compare it against the reference *K* tomogram. Fig. 6 shows a scatter plot of the local *K* estimates from the two cases. We see that the comparison between the two is very good and the differences between the two arise from some errors in the experiments as well as the imperfect packing of the sandbox.

Multi-method validation approach 5: comparison of local *K* values from the *K* tomogram to *K* from core and single-hole tests

To examine the performance of the algorithm in greater detail, we next compare local *K* values from the *K* tomogram to the *K* estimates from the cores (Fig. 7) and single-hole tests (Fig. 8). Both figures show that there is some scatter as well as bias in results, with the latter that is especially evident in Fig. 8. Robust as it is, neither the hydraulic tomography nor the SSLE is a perfect method. The more

head observations are collected, the higher the resolution of the estimates will be (i.e., there is no optimum).

Likewise, *K* estimates from core and single-hole tests are not devoid of errors contributing to the scatter in Figs. 7 and 8. In addition, inaccurate head observations and hydraulic property measurements (i.e., noise) during HT can unequivocally lead to an inaccurate estimate or cause the estimates to become unstable. The SSLE can overcome the impacts of noise through loosening of convergence criteria, but this causes the estimates to become smoother, which effectively results in a loss of information gained from the hydraulic head records.

Discussion

While the validation of hydraulic and pneumatic tomography under field conditions is our ultimate goal, the validation of hydraulic tomography and other tomography technologies in the laboratory is very important and a necessary step. In the laboratory, we are better able to control the forcing functions fully and quantify the errors. We can also pack a heterogeneous structure that is almost fully prescribed. However, the true *K* field remains unknown because of packing variations. In addition, there are only a finite number of small scale samples that can be used for validation purposes. Therefore, we emphasize that even in the laboratory setting, a direct and complete validation of results is generally difficult.

In this study, we have shown that the *K* tomogram can be validated using multiple methods. The tomograms can also be validated at multiple scales from the smaller scale, in which the local *K* values are compared to core and single-hole *K* estimates to the larger-scale *K* estimates from other cross-hole and flow-through experiments. Such a multi-faceted approach in validation adds more confidence on the ability of the algorithm to tackle field scale problems.

One form of model validation involves the establishment of greater confidence in a given model by conducting simulations of datasets that have not been used for calibration purposes. For example, one can calibrate a 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 2nd pumping test, the process becomes a second calibration and additional datasets are needed to continue with the validation exercise. Model validation is complete when the validation datasets are matched against simulated values resulting from the previously calibrated parameter values.

This is precisely the essence of hydraulic tomography conducted with sequential inclusion of data. Therefore, it amounts to a repeated validation of the estimated *K* field with new datasets that are sequentially added. The method is robust, but it is not the panacea technology. This is because the computed *K* tomogram is non-unique 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; Liu et al., 2002 and Yeh and Simunek, 2002). This is not the case here. However, it is important to recognize that we have obtained a solution to the inverse problem that is consistent with the heterogeneity patterns that we can visualize and directly compare against the experimentally packed sand distributions (Fig. 2). In addition, we were able to validate the resulting *K* tomogram using multiple methods and at multiple scales so our approach provides more confidence in the solution of the inverse problem.

Earlier, we saw that errors and biases can be very important in the result of the *K* tomogram and the blind addition of new data does not mean that it will automatically generate better results. Therefore, more effort should be expended on collecting accurate data and additional research should be conducted on improving hydraulic tomography technology both in sandbox experiments and in the field. This study also emphasizes the importance of reducing experimental errors and biases during validation of traditional groundwater flow and contaminant transport models.

Summary

Hydraulic tomography is a technology that facilitates the imaging of subsurface heterogeneity in hydraulic parameters. To date, a comprehensive validation of the hydraulic conductivity (*K*) tomogram has not been done either at the laboratory or field scales. Previous laboratory investigations assumed that packing was perfect and in general, small scale data were not available for a direct comparison. This study provides the first such examination using small-scale *K* data obtained from cores and single-hole tests as well as large-scale *K* estimates obtained from flow-through experiments in a sandbox with deterministic heterogeneity in hydraulic parameters.

Prior to inverse modeling of data, we conducted a detailed diagnostic study to investigate the magnitude and cause of errors and biases in data through scatter plots, contour plots and data animations. Such diagnostic tests of data used in forward and inverse models are rarely discussed in the literature, but we found that it should be an integral component of all phases of numerical forward and inverse modeling of cross-hole tests as the use of data corrupted by noise can have a profound effect on both forward and inverse model results.

Validation of the *K* tomograms involved a multi-method and multiscale approach which included: (1) visual comparisons of *K* tomograms to the true sand distributions as well as to the reference *K* tomogram; (2) testing the ability of the *K* tomogram to predict the hydraulic head distribution of an independent cross-hole test not used in the computation of the *K* tomogram; (3) comparison of the conditional mean and variance of local *K* from the *K* tomograms to the sample mean and variance of results from other measurements; (4) comparison of local *K* values in *K* tomograms to those from the reference tomogram; and (5) comparison of local *K* values in *K* tomograms to those obtained from cores and single-hole tests. The multi-method and multiscale validation approach proposed herein further illus-

trates the robustness of hydraulic tomography in subsurface heterogeneity delineation.

Previously, the effects of errors and biases on the *K* tomograms have not been investigated in detail. The steady-state inversion of cross-hole tests in a synthetic laboratory aquifer showed that the approach is sensitive to errors and biases. Data diagnostics combined with forward modeling provided valuable insight into identifying the cause of such errors and biases. Specifically, the errors identified include drift in pressure transducer readings, a skin effect influencing hydraulic head at the pumped well, and inaccurate treatment of boundary conditions, among others. We found that accurate modeling of boundary conditions is essential in conducting steady-state hydraulic tomography and obtaining accurate *K* tomograms. In real field situations, the boundary conditions of the field site need to be studied carefully through forward modeling and better site characterization. Further research is clearly needed to improve hydraulic tomography technology both in the laboratory and under field conditions.

Acknowledgements

This research was supported by the Strategic Environmental Research & Development Program (SERDP) as well as by funding from the National Science Foundation (NSF) through Grants EAR-0229713, IIS-0431069, and EAR-0450336. The inversion code used in this analysis is developed by our collaborator at the University of Arizona, Tian-Chyi J. Yeh and his students, under a collaborative project funded by NSF and SERDP. We thank Junfeng Zhu and Tian-Chyi J. Yeh for their assistance in using the steady-state hydraulic tomography code. We also thank Mike Kundert of IIHR-Hydroscience & Engineering for drafting Fig. 1. Finally, we greatly appreciate the constructive comments made by the anonymous reviewers which improved the manuscript.

References

Bohling, G.C., Zhan, X., Butler Jr., J.J., Zheng, L., 2002. Steady shape analysis of tomographic pumping tests for characterization of aquifer heterogeneities. Water Resources Research 38, 1324. doi:10.1029/2001WR001176.

Brauchler, R., Liedl, R., Dietrich, P., 2003. A travel time based hydraulic tomographic approach. Water Resources Research 39, 1370. doi:10.1029/2003WR002262.

Butler, J.J., McElwee, C.D., Bohling, G.C., 1999. Pumping tests in networks of multilevel sampling wells: Motivation and methodology. Water Resources Research 35, 3553—3560.

Craig, A.J., 2005. Measurement of hydraulic parameters at multiple scales in two synthetic heterogeneous aquifers constructed in the laboratory, M.S. thesis, Department of Civil and Environmental Engineering, The University of Iowa.

de Marsily, G., 1978. De l'Identification des systemes en hydrogeologiques, these de docteur es sciences. Univ. Pierre et Marie Curie-Paris VI, Paris.

Gottlieb, J., Dietrich, P., 1995. Identification of the permeability distribution in soil by hydraulic tomography. Inverse Problems 11, 353—360.

Guzman, A.G., Geddis, A.M., Henrich, M.J., Lohrstorfer, C.F., Neuman, S.P., 1996. Summary of Air Permeability Data From Single-Hole Injection Tests in Unsaturated Fractured Tuffs at the Apache Leap Research Site: Results of Steady-State Test Interpretation, NUREG/CR-6360, U.S. Nucl. Regul. Comm., Washington, DC.

Illman, W.A., 1999. Single- and cross-hole pneumatic injection tests in unsaturated fractured tuffs at the Apache Leap Research Site near Superior, Arizona, Ph.D. dissertation, Department of Hydrol. and Water Resour., University of Arizona, Tucson.

Illman, W.A., Neuman, S.P., 2001. Type-curve interpretation of a cross-hole pneumatic test in unsaturated fractured tuff. Water Resources Research 37, 583—604.

Illman, W.A., Neuman, S.P., 2003. Steady-state analyses of cross-hole pneumatic injection tests in unsaturated fractured tuff. Journal of Hydrology 281, 36—54.

Illman, W.A., Thompson, D.L., Vesselinov, V.V., Chen, G., Neuman, S.P., 1998. Single- and Cross-Hole Pneumatic Tests in Unsaturated Fractured Tuffs at the Apache Leap Research Site: Phenomenology, Spatial Variability, Connectivity and Scale, NUREG/CR-5559, U.S. Nucl. Regul. Comm., Washington D.C.

Klute, A., Dirksen, C., 1986. Hydraulic conductivity and diffusivity: laboratory methods. In: Klute, A. (Ed.), Methods of Soil Analysis Part 1 — Physical and Mineralogical Methods, second ed. American Society of Agronomy, Inc. (Chapter 28).

Konikow, L., 1978. Calibration of ground-water models. In: Verification of Mathematical and Physical Models in Hydraulic Engineering. American Society of Civil Engineers, New York, pp. 87—93.

Liu, S., Yeh, T.-C.J., Gardiner, R., 2002. Effectiveness of hydraulic tomography: sandbox experiments. Water Resources Research 38. doi:10.1029/2001WR000338.

McDermott, C.I., Sauter, M., Liedl, R., 2003. New experimental techniques for pneumatic tomographical determination of the flow and transport parameters of highly fractured porous rock samples. Journal of Hydrology 278, 51—63.

Vesselinov, V.V., Neuman, S.P., Illman, W.A., 2001b. Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff: 2. Equivalent parameters, high-resolution stochastic imaging and scale effects. Water Resources Research 37, 3019—3042.

Yeh, T.-C.J., Srivastava, R., Guzman, A., Harter, T., 1993. A numerical model for water flow and chemical transport in variably saturated porous media. Ground Water 31, 634—644.

Yeh, T.-C.J., Jin, M., Hanna, S., 1996. An iterative stochastic inverse approach: conditional effective transmissivity and head fields. Water Resources Research. 32, 85—92.

Yeh, T.-C.J., Liu, S., 2000. Hydraulic tomography: development of a new aquifer test method. Water Resources Research 36, 2095—2105.

Yeh, T.-C.J., Simunek, J., 2002. Stochastic fusion of information for characterizing and monitoring the vadose zone. Vadose Zone Journal 1, 207—221.

Zhu, J., Yeh, T.-C.J., 2005. Characterization of aquifer heterogeneity using transient hydraulic tomography. Water Resources Research 41, W07028. doi:10.1029/2004WR003790.

Zhu, J., Yeh, T.-C.J., 2006. Analysis of hydraulic tomography using temporal moments of drawdown recovery data. Water Resources Research 42, W02403. doi:10.1029/2005WR004309.