Practical Issues in Imaging Hydraulic Conductivity through Hydraulic Tomography

by Walter A. Illman1,2,3, Andrew J. Craig1, and Xiaoyi Liu1

Abstract

Hydraulic tomography has been developed as an alternative to traditional geostatistical methods to delineate heterogeneity patterns in parameters such as hydraulic conductivity ($K$) and specific storage ($S_s$). During hydraulic tomography surveys, a large number of hydraulic head data are collected from a series of cross-hole tests in the subsurface. These head data are then used to interpret the spatial distribution of $K$ and $S_s$ using inverse modeling. Here, we use the Sequential Successive Linear Estimator (SSLE) of Yeh and Liu (2000) to interpret synthetic pumping test data created through numerical simulations and real data generated in a laboratory sandbox aquifer to obtain the $K$ tomograms. Here, we define “$K$ tomogram” as an image of $K$ distribution of the subsurface (or the inverse results) obtained via hydraulic tomography. We examine the influence of signal-to-noise ratio and biases on results using inverse modeling of synthetic and real cross-hole pumping test data. To accomplish this, we first show that the pumping rate, which affects the signal-to-noise ratio, and the order of data included into the SSLE algorithm both have large impacts on the quality of the $K$ tomograms. We then examine the role of conditioning on the $K$ tomogram and find that conditioning can improve the quality of the $K$ tomogram, but can also impair it, if the data are of poor quality and conditioning data have a larger support volume than the numerical grid used to conduct the inversion. Overall, these results show that the quality of the $K$ tomogram depends on the design of pumping tests, their conduct, the order in which they are included in the inverse code, and the quality as well as the support volume of additional data that are used in its computation.

Introduction

Ground water investigations have relied on the determination of aquifer parameters such as hydraulic conductivity ($K$) and specific storage ($S_s$). In the past, these were determined by conducting slug and/or pumping tests while using an analytical solution that treats the geological medium to be homogeneous. These solutions and the estimated parameters have been used in a variety of applications despite the fact that the subsurface is heterogeneous at multiple scales, which is the rule rather than the exception. In particular, the knowledge of detailed three-dimensional (3D) distributions of $K$ is critical in prediction of contaminant transport, delineation of well catchment zones, and quantification of ground water fluxes, including surface water/ground water exchange.

However, characterization of subsurface heterogeneity of hydraulic parameters is fraught with difficulties (e.g., Wu et al. 2005). Information about the spatial variability of flow parameters is most commonly obtained through inference from small-scale measurements of cores, slug/bail tests, and single-hole pressure tests by geostatistical methods, which require numerous measurements. This requires the drilling of numerous boreholes, collection of a large number of samples, and the conduct of multiple measurements within various depth intervals in each of them using sophisticated equipment. The approach is expensive and time consuming, and thus has not been adopted widely in practice.

In addition, the physical meaning of the flow parameter estimates from either traditional pumping or slug tests...
is considered to be questionable (e.g., Wu et al. 2005). Furthermore, it is not clear that geostatistical analysis of data collected on relatively small support scales is necessarily indicative of medium properties that impact flow and transport on scales that are much larger. One alternative to traditional geostatistical analyses is hydraulic/pneumatic tomography [e.g., Gottlieb and Dietrich (1995); Butler et al. (1999); Yeh and Liu (2000); Vesselinov et al. (2001); Liu et al. (2002, 2007); Bohling et al. (2002); Brauchler et al. (2003); McDermott et al. (2003); Zhu and Yeh (2005, 2006); and Illman et al. (2007)]. Hydraulic tomography is a cost-effective technique for characterizing subsurface heterogeneity of hydraulic parameters. During hydraulic tomography surveys, drawdowns induced by sequential pumping or injection tests at different locations of an aquifer are collected at a large number of subsurface locations. These hydraulic head data are then used to interpret the spatial distribution of hydraulic parameters of the aquifer using an inverse model. Pneumatic tomography is similar in concept to hydraulic tomography, but the well tests are conducted with air in the unsaturated zone (Illman and Neuman 2001, 2003; Vesselinov et al. 2001).

In particular, Yeh and Liu (2000) developed a sequential geostatistical inverse method, which they refer to as the Sequential Successive Linear Estimator (SSLE). This approach can be applied to hydraulic tomography for the interpretation of cross-hole pumping tests under steady-state conditions. Their approach combines the traditional geostatistical approach and governing flow principles to interpolate and extrapolate at locations where samples are not available. As a consequence, the SSLE as implemented in hydraulic tomography yields more realistic \( K \) estimates than kriging, which does not consider principles of flow, and deterministic-zone-based or stochastic inverse modeling approaches that use one pumping or injection data set only. The main advantage of sequentially including pumping tests is its computational efficiency. Yeh and Liu (2000) showed that accurate \( K \) distributions can be obtained through data sets from numerically simulated pumping tests in synthetic heterogeneous aquifers. Validation of the steady-state hydraulic tomography in Yeh and Liu (2000) was limited to error-free cases of synthetic simulations.

Liu et al. (2002) then 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 visually compared only 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 effects on their \( K \) tomograms, but they did not examine the role of errors and biases directly by isolating their causes. Here, we define “\( K \) tomogram” as an image of \( K \) distribution of the subsurface (or the inverse results) obtained via hydraulic tomography.

Illman et al. (2007) further examined the accuracy of the \( K \) tomograms obtained from the steady-state hydraulic tomography algorithm developed by Yeh and Liu (2000). They obtained 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. Illman et al. (2007) also examined the influence of errors and biases on inversion results using forward and inverse simulations of cross-hole tests. They found out that the pressure transducer offsets, skin effect at the pumped well, among other sources of errors can have a large impact on the quality of the inverse modeling results. Likewise Liu et al. (2007) conducted a laboratory validation of the transient hydraulic tomography to estimate both \( K \) and \( S \) tomograms simultaneously, but both Illman et al. (2007) and Liu et al. (2007) have not examined practical factors such as the effects of signal-to-noise ratio and the role of conditioning on the quality of the inversion results.

There are several issues that need to be further examined. One important issue in applying hydraulic tomography in the field is how much noise a given data set can contain in obtaining an accurate \( K \) tomogram without having to apply smoothing and/or signal processing techniques, which may result in the loss of information on the parameters contained in the data set. Another issue is that SSLE incorporates pumping test data sequentially, but no studies have been published to date that examines the role of varying the order of pumping test data included in the SSLE algorithm. Finally, during site characterization, test data other than cross-hole hydraulic test data such as from direct push technologies, core, slug, geophysical, and geochemical data may be collected. In some cases, one could use some of these data to condition the inverse modeling results. However, there are no studies to our knowledge in which the role of conditioning on the estimated \( K \) tomogram by means of hydraulic/pneumatic tomography was systematically studied.

To address these practical yet important issues, we continue our study on hydraulic tomography using synthetic data generated through numerical simulations and real data collected in the laboratory. The main objectives of this paper are (1) to further study the validity of steady-state hydraulic tomography through synthetic pumping test data obtained through forward numerical simulations and real data obtained using a laboratory sandbox aquifer; (2) to investigate the effect of varying the pumping rate, which affects the signal-to-noise ratio, and its impact on \( K \) tomograms; (3) to investigate the effect of varying the order of test data included into the SSLE inversion algorithm; and (4) to investigate the effect of conditioning on \( K \) tomograms.

We first describe the experimental setup for the synthetic and real cases used to conduct our study. Next, we discuss the numerical simulation approach that underlies hydraulic tomography and approaches used to generate synthetic hydraulic test data and corresponding laboratory data sets. We then briefly discuss the reference \( K \) tomograms from synthetic and real data generated through laboratory sandbox experiments by Illman et al. (2007). These reference \( K \) tomograms will be used to compare against the new results presented in this paper. In all cases, we first compare our results visually to the
reference $K$ tomograms obtained as a benchmark result for both the synthetic and real cases. We also assess the validity of the synthetic and real $K$ tomograms by simulating an independently conducted pumping test and comparing the head values obtained from the simulated and observed cases.

We emphasize that the synthetic experiments conducted on the computer are necessary to test the SSLE algorithm under optimal conditions in which the experimental errors are neglected and the forcing functions (boundary condition and source/sink terms) are fully controlled. In addition, through numerical simulations, hydraulic tomography and the $K$ tomograms obtained can be tested rigorously. The laboratory experiments described subsequently are also required to test the SSLE algorithm under controlled conditions, which is a necessary step toward its field applications.

Setup for Synthetic and Real Experiments

All $K$ tomograms, whether synthetic or real, are generated based on the experimental setup that we describe subsequently. Here, a synthetic $K$ tomogram is one that is generated by inverting pumping test data generated on the computer. A real $K$ tomogram is obtained by inverting actual pumping test data obtained through laboratory experiments. The synthetic aquifer was built in a sandbox that is 193.0 cm in length, 82.6 cm in height, and has a width of 10.2 cm. 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. The core/well assemblies were inserted into the completed sandbox and the cores were removed from the wells. No sand was lost from the cores upon completion, and we are confident that there was minimal impact on the surrounding materials within the sandbox. The horizontal wells fully penetrated the thickness of the laboratory aquifer. This allowed the hydraulic head to be monitored by a pressure transducer. Because the well diameter is very small in comparison to the sandbox, the wells should have a negligible impact on the flow field. The ports can additionally be used as pumping ports and to obtain water samples. Figure 1a is a drawing of the sandbox frontal view, showing the 48 port and pressure transducer locations as well as water reservoirs for controlling hydraulic head. We additionally indicate on Figure 1a the ports that have been used for pumping tests for hydraulic tomography through open circles drawn around the port numbers. Figure 1b is a photograph of the sandbox showing the heterogeneous aquifer built in the laboratory with different sand types and their distributions. Four different commercially sieved sands [20/30 and 4030, U.S. Silica (Ottawa, Illinois); F-75 and F-85, Unimin Corp. (New Canaan, Connecticut), shown on Figure 1b] were used to pack the sandbox by hand. The sand was wetted from the bottom and the water levels were increased while packing the aquifer as uniformly as possible. It was of paramount importance to pack the sandbox with a known heterogeneous $K$ distribution in order to validate the computed $K$ tomograms. This included limiting the development of anisotropy in $K$.

The flow system for the sandbox is driven by two constant-head reservoirs, one at each end. We also used an intermediate overflow device to maintain equal constant heads on both boundaries during all experiments. The device consisted of a reservoir with an overflow pipe and tube connecting it to the inlets at the bottom of the constant-head reservoirs. The adjoining reservoirs and overflow device are capable of supplying water throughout the sandbox length and thickness. 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 constant-head reservoirs. The top boundary has the same head as the two side boundaries as they are directly connected. The main reason for maintaining three constant-head boundaries was to avoid unsaturated flow conditions at the top of the aquifer, which we found to cause numerical difficulties during inverse modeling.

Pressure measurements were made with 50 Setra model 209-gauge 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 flow cell. Further details of the laboratory aquifer, its specifications, and its capabilities are described in Craig (2005), Illman et al. (2007), and Liu et al. (2007).

Numerical Simulation Methods

Description of Inverse Modeling Approach

Inverse modeling of all synthetic and real pumping tests was conducted using a sequential geostatistical inverse approach developed by Yeh and Liu (2000), in which all details to the algorithm are provided. Here, we provide a brief description of the numerical inversion approach. The inverse model assumes a steady flow field and the natural logarithm of $K$ (ln $K$), which is treated as a stationary stochastic process. The model additionally assumes that the mean and correlation structure of the $K$ field are known a priori. The algorithm essentially is composed of two parts. First, the Successive Linear Estimator is employed for each cross-hole test. The estimator begins by cokriging the initial $K$ value determined and observed heads collected in one pumping test during the tomographic sequence to create a cokriged, mean-removed ln $K$ (i.e., perturbation of ln $K$) map. 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.

The second step of the Yeh and Liu (2000) approach is to use the head data sets 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 head measurements as prior information for the next estimation based on a new set of pumping data. The estimated $K$ field and
covariances are updated sequentially as new pumping test data are included into the inverse algorithm. This process continues until all the data sets are fully used.

**Inputs to the Inverse Model**

To obtain a $K$ tomogram from multiple cross-hole pumping 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 $4.1 \times 10.2 \times 4.1$ cm. Both sides and the top boundary were set to the same constant-head boundary condition, while the bottom boundary of the sandbox was set to be 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 a finite-element flow model MMOC3 (Yeh et al. 1993). 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 real cross-hole tests conducted in the sandbox.

Inputs to the inverse model include the initial estimate of effective hydraulic conductivity ($K_{eff}$), variance ($\sigma_{lnK}^2$), correlation scales of hydraulic conductivity ($\lambda_x$, $\lambda_y$, $\lambda_z$), volumetric discharge ($Q_n$ where $n$ is the test number), and available point (small-scale) measurements of $K$. Results that do not use available point-scale measurements of $K$ to test the ability of the algorithm to delineate the heterogeneity patterns are described in Illman et al. (2007) as well as in this paper. Later in this paper, we examine the effect of using available point-scale $K$ data (i.e., conditioning) on the computation of $K$ tomograms.

We obtained the initial estimate of $K_{eff}$ by averaging the $K$ 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 (described later) to obtain the $K_{eff}$. However, we select the $K_{eff}$ obtained through the traditional analysis of cross-hole tests by treating the medium to be homogeneous because in practice such values are most readily available through type curve, straight line, or asymptotic analysis (Illman and Tartakovsky 2006) of pumping test data.

Figure 1. (a) Computer-aided design (CAD) drawing of the sandbox used for laboratory aquifer tests with location of sand lenses shown and (b) photograph of the sandbox (modified after Liu et al. 2007; Illman et al. 2007).
Generation of Synthetic and Laboratory Hydraulic Test Data

Synthetic Hydraulic Test Data

We generated synthetic slug, single-hole, and cross-hole pumping test data on the computer using the MMOC3 forward model. There were two purposes to generate these synthetic data. One purpose was to generate synthetic cross-hole pumping test data, which are later used to generate the reference \( K \) tomogram. The reason for generating the slug and single-hole test data are so that we can interpret them and obtain small-scale \( K \) estimates, which we can in turn use to condition the \( K \) tomograms.

Figure 2 shows the \( K \) distribution of four sand types in the synthetic sandbox that is used to generate synthetic data on the computer. The \( K \) values for the four sand types are listed in Table 1 and are obtained through the analysis of core samples through a constant-head permeameter, which we describe later. We first conducted slug tests at the 48 port locations by raising the initial head and recording the corresponding decay in the head using MMOC3. We then conducted synthetic cross-hole tests by running steady-state forward simulations using MMOC3 at each of the 48 ports by setting a constant pumping rate and recording the hydraulic heads at the other ports. The pumping rate was set between 2.92 and 3.17 mL/s depending on the cross-hole pumping test. Boundary conditions for the simulations were equal constant-head conditions at the top and the two side boundaries, while the bottom boundary remained a no-flow boundary.

Laboratory Hydraulic Test Data

Parallel to the generation of synthetic hydraulic test data, we conducted different hydraulic tests in the sandbox to characterize the real aquifer and to generate conditioning data. We first determined the \( K \) of the four types of sands from the extracted core using a constant-head permeameter (Klute and Dirksen 1986). We also conducted slug tests at each of the 48 ports that used an external well connected to the port.

After completing the slug tests, cross-hole pumping tests were conducted at 46 ports. Ports 36 and 38 experienced minor well screen damage, so we do not pump from these ports. Rather, they are used only for head observations. For the cross-hole tests, pumping rates remained constant during the test duration. We used a pumping rate that is consistent with the synthetic case.

Table 2 provides the pumping rate and duration for the tests interpreted in this paper. Out of the 46 pumping tests, we selected 8 tests (pumping at ports 2, 5, 14, 17, 32, 35, 44, and 47) for hydraulic tomography that we describe later. One test conducted by pumping at port 46 was reserved for validation purposes. We selected the eight pumping test data for analysis in this paper as pumping took place in two vertical columns (Figure 1a), which represent vertical wells, a situation which could be readily replicated in the field.

We also conducted nine flow-through experiments through the entire sandbox to obtain the effective hydraulic conductivity (\( K_{\text{eff}} \)) 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 the both sides of the sandbox. After the flow reached steady state, we measured volumetric discharge from one side of the sandbox. The difference between the heights of the water column in the two constant-head reservoirs was measured to determine the hydraulic gradient. Further details of the hydraulic experiments conducted in the sandbox can be found in Craig (2005), Illman et al. (2007), and Liu et al. (2007).

<table>
<thead>
<tr>
<th>Sand Type</th>
<th>Manufacturer</th>
<th>( n )</th>
<th>( K ) (cm/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20/30 U.S. Silica</td>
<td>32</td>
<td>2.60( \times )10(^{-1} )</td>
<td></td>
</tr>
<tr>
<td>4030 U.S. Silica</td>
<td>3</td>
<td>6.42( \times )10(^{-2} )</td>
<td></td>
</tr>
<tr>
<td>F-75 Unimin Corp.</td>
<td>5</td>
<td>1.99( \times )10(^{-2} )</td>
<td></td>
</tr>
<tr>
<td>F-85 Unimin Corp.</td>
<td>8</td>
<td>1.61( \times )10(^{-2} )</td>
<td></td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Pumping Port No.</th>
<th>Pumping Rate (mL/s)</th>
<th>Pumping Duration (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.07</td>
<td>165</td>
</tr>
<tr>
<td>5</td>
<td>3.10</td>
<td>205</td>
</tr>
<tr>
<td>14</td>
<td>3.17</td>
<td>242</td>
</tr>
<tr>
<td>17</td>
<td>3.17</td>
<td>242</td>
</tr>
<tr>
<td>32</td>
<td>2.97</td>
<td>243</td>
</tr>
<tr>
<td>35</td>
<td>2.92</td>
<td>181</td>
</tr>
<tr>
<td>44</td>
<td>3.12</td>
<td>212</td>
</tr>
<tr>
<td>46</td>
<td>3.00</td>
<td>254</td>
</tr>
<tr>
<td>47</td>
<td>3.12</td>
<td>166</td>
</tr>
</tbody>
</table>

Note: Tests with pumping taking place at ports 2, 5, 14, 17, 32, 35, 44, and 47 were used in hydraulic tomography, while the test at port 46 was used for validation.
Estimation of $K$ from Synthetic and Laboratory Experimental Data Sets

Estimation of Synthetic $K$ Data

We estimate $K$ from synthetically generated core, slug, and single-hole data for conditioning the $K$ tomograms later. The core $K$ estimates were obtained by simply reading off the $K$ value assigned to the synthetic $K$ distribution (Figure 2) at the locations where the actual cores were extracted. This assumes that there is no disturbance to the cores and that core $K$ estimates are completely accurate (i.e., no experimental error).

We also conducted synthetic slug tests at each of the 48 ports on the computer using MMOC3 and analyzed the data by manually calibrating MMOC3 by treating the model domain to be a 3D, homogeneous medium. We also considered existing analytical solutions to interpret the data but decided against using them for consistency. That is, the synthetic slug test data were generated with the forward model MMOC3, so it would be best to interpret the data through manual calibration using MMOC3. The numerical grid used for the interpretation of synthetic data was identical to the one used for hydraulic tomography. Boundary conditions for the simulations involving manual calibrations were identical to the numerical simulations for synthetic data generation and to the real experiments. 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.

We then conducted synthetic pumping tests at each of the 48 ports on the computer using MMOC3. For each steady-state simulation, hydraulic head data were collected from all ports. We analyzed the steady-state head records at the pumping ports by manually calibrating MMOC3 and assuming that the aquifer is homogeneous. The numerical setup for the calibration is identical to the slug test analysis. The $K$ values obtained in this manner using MMOC3 yielded local or single-hole estimates of $K$. These results are denoted as the single-hole results.

Table 3 summarizes the results from all synthetic data sets. 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. In Table 3, we see that, in general, the mean values of the slug and single-hole test values are larger than that of the core values, which suggests a scale effect (e.g., Illman and Neuman 2001, 2003; Illman 2006). Examination of Table 3 also shows that the variance of $\ln K$ ($\sigma_{\ln K}^2$) 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, and single-hole tests. As the sample volume increases, $K$ is averaged over the investigated volume.

Table 4

<table>
<thead>
<tr>
<th>Test Type</th>
<th>$N$</th>
<th>$\ln K$ (K - cm$^{-1}$)</th>
<th>$\sigma_{\ln K}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>48</td>
<td>$-2.166 (1.146 \times 10^{-1})$</td>
<td>1.498</td>
</tr>
<tr>
<td>Slug (VSAFT2)</td>
<td>40</td>
<td>$-10.692 (2.273 \times 10^{-5})$</td>
<td>0.431</td>
</tr>
<tr>
<td>Slug (AQTESOLV)</td>
<td>46</td>
<td>$-1.910 (1.481 \times 10^{-1})$</td>
<td>0.521</td>
</tr>
<tr>
<td>Single-hole</td>
<td>48</td>
<td>$-2.835 (5.872 \times 10^{-2})$</td>
<td>0.589</td>
</tr>
<tr>
<td>Cross-hole</td>
<td>96</td>
<td>$-1.757 (1.726 \times 10^{-1})$</td>
<td>0.074</td>
</tr>
<tr>
<td>Flow-through</td>
<td>9</td>
<td>$-1.757 (1.726 \times 10^{-1})$</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Table 3

<table>
<thead>
<tr>
<th>Test Type</th>
<th>$N$</th>
<th>$\ln K$ (K - cm$^{-1}$)</th>
<th>$\sigma_{\ln K}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>48</td>
<td>$-2.166 (1.146 \times 10^{-1})$</td>
<td>1.498</td>
</tr>
<tr>
<td>Slug</td>
<td>48</td>
<td>$-1.906 (1.491 \times 10^{-1})$</td>
<td>0.216</td>
</tr>
<tr>
<td>Single-hole</td>
<td>48</td>
<td>$-1.953 (1.421 \times 10^{-1})$</td>
<td>0.083</td>
</tr>
</tbody>
</table>

Estimation of Laboratory $K$ Data

We then determined the $K$ of the four types of sands from the horizontal cores in the laboratory. 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 for determination of $K$. Details of the core extraction method and the design of the constant-head permeameter are provided in Craig (2005). The $K$ values from cores are calculated using Darcy’s law. We report the arithmetic mean of 48 natural logarithm transformed $K$ values in Table 4.

We also obtained $K$ estimates from slug tests conducted 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. Originally, Craig (2005) analyzed the slug test data in an identical fashion described for the synthetic slug test data, but the interpretation was done using VSAFT2, a graphical user interface (GUI) version of the MMOC3 code, available for free at http://tian.hwr.arizona.edu/yeh/. Results from these simulations are summarized in Table 4. Results obtained revealed that the $K$ values were 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 (two to eight). This effort revealed that the head response stabilizes after six cuts were made on the well. Therefore, all wells in the sandbox discussed
in our analysis. Examination of Table 4 also shows that \( \sigma_{\text{est}} \) varies from one type of test to the next with variance decreasing with the increasing scale. This behavior is similar to the results from the synthetic hydraulic test data (Table 3) and is attributed again to the increase in support volume as the scale of the test increases.

**Generation of Reference \( K \) Tomograms Using Synthetic and Real Pumping Test Data**

Here, we briefly describe the reference \( K \) tomograms computed using synthetic and real data by Illman et al. (2007). The purpose of constructing a reference \( K \) tomogram using synthetic data is to examine the ability of the SSLE 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). The reference \( K \) tomogram is then used to assess the effects of signal-to-noise ratio and conditioning. We also obtain a reference \( K \) tomogram using real data so that we can later compare these results to those generated with noisy data and those conditioned to available core, slug, and single-hole \( K \) estimates.

The reference \( K \) tomogram (Figure 3a) using synthetic data was computed through the inversion of eight synthetic cross-hole test data generated via numerical simulations. The computation was done sequentially by including pumping tests conducted at ports 47, 44, 35, 32, 17, 14, 5, and 2, in that order. All 48 ports (Figure 1a for locations) were used in steady-state hydraulic tomography. This result clearly shows that the SSLE algorithm is capable of capturing the correct position of the low \( K \) blocks, its morphology, dimensions, and other details of aquifer heterogeneity such as windows in low \( K \) strata that could provide continuous pathways for contaminant transport. However, the low \( K \) blocks and their dimensions at the bottom of the sandbox (Figure 3a) are not captured as clearly as the other blocks positioned higher in the sandbox when compared to Figure 2. This could perhaps be due to the fact that there are no observation ports beneath the low \( K \) blocks.

The reference \( K \) tomogram (Figure 3b) using real data was computed through the inversion of eight sets of cross-hole pumping test data collected in the laboratory sandbox aquifer. We used the pumping tests taking place at the same ports as in the synthetic case and in the same order (i.e., 47, 44, 35, 32, 17, 14, 5, and 2). It is important to note that no conditioning data were used to generate this reference \( K \) tomogram. Figure 3b shows the best \( K \) tomogram obtained by Illman et al. (2007) after various error and bias reduction schemes were applied to the raw data set collected in the sandbox. The main reasons for applying the error and bias reduction schemes were to remove outliers and excessively noisy data. Briefly, the error reduction scheme consisted of (1) removing pumped well data thought to be affected by a skin effect at the pumped port; (2) correcting for drift or offset in the pressure transducers; and (3) accounting for slight variations in boundary conditions from one pumping test to the next in the SSLE algorithm. Figure 3b shows that the
heterogeneity pattern consisting of low $K$ blocks are mostly captured, except for the bottom two blocks. There is also a thin and continuous high $K$ zone at the top of the image that is not visible on the reference $K$ tomogram (Figure 3a). This is likely due to the lack of compaction of sands near the top of the sandbox.

**Effects of Pumping Rate, Order of Test Data in SSLE, and Conditioning on Synthetic and Real $K$ Tomograms**

We next describe the effects of (1) varying the pumping rate, which affects the signal-to-noise ratio of head data collected during hydraulic tomography; (2) the order of test data included in the SSLE algorithm; and (3) conditioning on both synthetic and real $K$ tomograms. The reference $K$ tomograms from the synthetic (Figure 3a) and real (Figure 3b) cases will serve as our baseline results for purposes of comparison.

**The Effect of Pumping Rate and Signal-to-Noise Ratio**

We first examine the effect of varying the signal-to-noise ratio in the cross-hole pumping test data on the resulting $K$ tomograms. A larger signal-to-noise ratio data can be generated by increasing the pumping rate and/or decreasing the noise level through signal conditioning techniques. Here, we vary the signal-to-noise ratio by conducting pumping tests at two different rates. The high pumping rate case was already presented by Illman et al. (2007) and is shown here as Figures 3a and 3b. We present subsequently the low flow rate case. The flow rate used for the synthetic simulations was 1.6 mL/s, while for the real case, it varied for each test. It is evident from the synthetic case (Figure 4a) that the pumping rate and hence the signal-to-noise ratio have very little effect on

---

**Figure 3.** Reference $K$ tomograms generated by sequentially inverting eight (a) synthetic and (b) real cross-hole pumping test data (modified after Illman et al. 2007).

**Figure 4.** $K$ tomograms obtained from low flow rate pumping test data with lower signal-to-noise ratio obtained for the (a) synthetic and (b) real cases.

**Figure 5.** $K$ tomograms obtained by varying the order of pumping test data included in the SSLE algorithm for the (a) synthetic and (b) real cases.
the quality of the synthetic $K$ tomogram when compared to the reference $K$ tomogram (Figure 3a). However, for the real case (Figure 4b), it has a noticeable detrimental impact when compared to Figure 3b because of the larger noise level in the hydraulic head data associated with a lower pumping rate. We emphasize that the SSLE algorithm can overcome the effects of the noise level through loosening of the convergence criteria. However, loosening of the convergence criteria results in a smoother result that approaches the effective $K$ value, which means that the heterogeneities become less well defined.

The Effect of Order of Test Data Included Sequentially in the SSLE Algorithm

We next vary the order of test data included into the SSLE algorithm. For the synthetic case (Figure 5a), we included the tests with pumping taking place in the order of ports 2, 5, 14, 17, 32, 35, 44, and 47. Figure 5a shows that this has little effect on the computed $K$ tomogram when compared to Figure 3a. We tried various combinations and found that the order of test data included has very little effect on the $K$ tomograms.

In contrast, the order of test data included has a large impact on the computation of the real $K$ tomogram (Figure 5b) when compared to Figure 3b. This is because some head records from each pumping test are noisier than others. We found 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, a stronger signal is generated near the bottom of the sandbox because of the no-flow boundary and due to the superposition principle.

We tried various combinations of test data included in the SSLE algorithm and found that the order of test
The Effect of Conditioning Using Core $K$ Data

Conditioning is generally thought to improve results of conditional stochastic simulations and hydraulic tomography (e.g., Yeh and Liu 2000). We investigate this observation through the use of various conditioning data with the SSLE algorithm. For this investigation, 48 conditioning points consisting of various data types (core, slug, and single-hole $K$ estimates) are placed at the ports shown on Figures 1 and 2.

We first include core $K$ estimates into the computation of the $K$ tomogram. Figure 6a shows that conditioning of the $K$ tomogram with core $K$ data dramatically improves the quality of the synthetic tomogram when compared to Figure 3a. In particular, the low $K$ blocks throughout the tomogram appear clearer compared to the background aquifer material. We also see that the dimensions of the low $K$ blocks approach the true case (Figure 2) and more important, the low $K$ blocks at the bottom of the sandbox become clearer.

We next condition the real $K$ tomogram with real core $K$ data obtained in the laboratory at 48 port locations. Figure 6b again shows a dramatic improvement of the $K$ tomogram in comparison to the reference $K$ tomogram (Figure 3b). In particular, we see the low $K$ block appearing clearly at the bottom of the sandbox.

The Effect of Conditioning Using Slug $K$ Data

We next condition the $K$ tomogram with available synthetic and real slug $K$ data. For the synthetic case (Figure 7a), we see a slight deterioration in the $K$ tomogram. This is because the slug $K$ estimates represent a larger support volume than the core $K$ estimate. In fact, results from forward simulation of slug tests using VSAFT2 not shown here confirmed this finding. That is, the head distribution resulting from the synthetic slug test generated on the computer revealed a volume of influence that is larger than the numerical grid used in the computation of the $K$ tomogram. Therefore, inclusion of slug $K$ data as conditioning points causes smoothing of the $K$ tomogram.

Conditioning of the real $K$ tomogram (Figure 7b) using 48 slug test $K$ data leads to a slight deterioration in the tomogram when compared to Figure 3b. In particular, we see the smoothing of the low $K$ blocks throughout the aquifer as in the synthetic case. The Effect of Conditioning Using Single-Hole $K$ Data

We next condition the $K$ tomogram with available synthetic and real single-hole $K$ data. As in the case of conditioning with slug $K$ data, the synthetic result (Figure 8a) shows a slight deterioration and smoothing of the $K$ tomogram when compared to Figure 3a. This is because single-hole $K$ represents a much larger support volume. VSAFT2 simulations not shown here reveal that the cone of depression essentially reaches all boundaries at steady state. Therefore, single-hole $K$ estimates represent a considerably larger area around the pumping port. Therefore, inclusion of single-hole $K$ data as conditioning points causes smoothing of the $K$ tomogram. It is of interest to note that the $K$ tomograms from Figures 7a and 8a look similar.

Conditioning of the real $K$ tomogram (Figure 8b) using single-hole $K$ data leads to a slight deterioration in the tomogram when compared to Figure 3b. In particular, we see the smoothing of the low $K$ blocks throughout the aquifer. These results are similar to those obtained through conditioning with slug $K$ test data (Figure 7b).

How Good Are the $K$ Tomograms?

As shown in the previous section, one could assess the effects of varying signal-to-noise ratio and conditioning on the computed $K$ tomograms by comparing the results with the reference $K$ tomograms (Figures 3a and 3b). However, visual comparison of the tomograms is qualitative. A more rigorous and quantitative approach to validate the computed $K$ tomograms is necessary.

Illman et al. (2007) and Liu et al. (2007) suggested that an appropriate validation approach is to test the predictability of the hydraulic head estimates under different flow scenarios. In order to do this, we validate the $K$ tomograms by simulating an additional pumping test that was not used in the inversion. The forward simulation of this independent test will yield heads at the monitoring ports, which are then compared to the actual head data through a scatter plot. We adopt this approach to validate the synthetic and real $K$ tomograms. In particular, for the synthetic case, we simulate another pumping test using the $K$ tomograms (Figures 3a to 8a) and the true $K$ distribution (Figure 2). In all simulations, pumping takes place at port 46. A comparison of the steady-state head values through a scatter plot then should validate the $K$ tomograms.

Figures 9a to 9f show the results of this comparison for the synthetic case. It shows that the comparison is excellent for all cases for monitoring ports far away from the pumping port, where the drawdown is smaller. We notice that simulated head values are larger than the observed ones, when the head values are small for all cases. However, the $K$ tomogram obtained by conditioning using core $K$ data yields the smallest discrepancy between the simulated and observed head values near the pumping port. The comparison of the heads near the pumped port is worse for the other cases. This is consistent with the fact that uncertainty in the predicted drawdown grows with the mean gradient according to stochastic analysis and confirms the earlier finding by Liu et al. (2007).
Two criteria, the average absolute error norm ($L_1$) and the mean-squared error norm ($L_2$), were used to quantitatively evaluate the goodness of fit between the simulated and observed hydraulic head responses:

$$L_1 = \frac{1}{n} \sum_{i=1}^{n} |h_{s,i} - h_{m,i}| \quad (1)$$

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

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 $L_1$ and $L_2$ norms are, the better the estimate is. Specifically, when one compares the results (Figures 9b to 9f) to the unconditioned case (Figure 9a), (1) Figure 9b shows that the effect of low pumping rate has a negligible impact on results; (2) Figure 9c shows that changing the order of pumping tests slightly deteriorates the results; and (3) Figures 9d to 9f show that conditioning tends to improve the results for the synthetic cases.

We also use this validation approach on the real $K$ tomograms. For this, we use the various real $K$ tomograms (Figures 3b to 8b) and simulate an independent cross-hole pumping test with pumping taking place at port 46. Pumping test at port 46 was chosen for validation purposes because it was not used in the construction of the $K$ tomograms, and it is also a pumping test with the cleanest data. Figures 10a to 10f show the results of this comparison. According to this figure, the data pairs are scattered along the 45° line, indicating that predicted head distributions generally are statistically unbiased in comparison with the observed except for those generated with a low pumping rate (i.e., the results with low signal-to-noise ratio) (Figure 10b). This plot and the quantitative measures ($L_1$ and $L_2$ norms) show that the comparison is consistently very good, providing us with further confidence that SSLE can provide an unbiased estimation of the head distribution. Note that the simulated heads will not necessarily match the observed ones perfectly as they should, due to the fact that the tomograms are conditional effective $K$ fields and also due to noise in the observations.

Findings and Conclusions

The subsurface is inherently heterogeneous at multiple scales, but the delineation of subsurface heterogeneity using current technology such as with traditional geostatistical methods is time consuming and expensive. We presented here an application of hydraulic tomography using cross-hole pumping tests conducted in a laboratory aquifer to investigate practical issues that may be of concern to field hydrogeologists. The main objectives of this study were to examine the role of pumping rate, which has a large impact on signal-to-noise ratio, the order of test data included in the SSLE algorithm, and conditioning...
on the quality of \( K \) tomogram. We investigated these issues by comparing \( K \) tomograms obtained from various synthetic data generated through numerical simulations and real data collected in a laboratory sandbox aquifer with a prescribed heterogeneity pattern. This study leads to the following findings and conclusions.

1. We first investigated the mean values of the synthetic and real hydraulic tests conducted on the computer and in the laboratory sandbox aquifer. Results show that the mean \( K \) in general increases with the measurement scale for the synthetic simulations, when the mean \( K \) from cores is compared to the mean \( K \) from both slug and single-hole tests. There is also a scale effect in \( K \) for the real data, when mean \( K \) from cores is compared to mean \( K \) values from cross-hole and flow-through tests. The same holds true for the comparison between mean \( K \) from both slug and single-hole tests to the larger scale tests. The variance, however, decreases with the measurement scale as the larger scale measurements average the porous medium.

2. We investigated the effect of pumping rate, which affects the signal-to-noise ratio of observed test data on the quality of the computed \( K \) tomogram. We found that the signal-to-noise ratio is not as important for the synthetic case but very important for the real case. That is, the computed \( K \) tomogram was clearer (and closer to the reference \( K \) tomogram) when the pumping test was conducted at a higher rate, yielding higher signal-to-noise ratio data, when analyzing real data. This is because different cross-hole pumping test data sets contains different levels of noise, and its magnitude depends on a number of factors, including (1) quality of pressure transducers (i.e., accuracy and precision of pressure transducers); (2) calibration of pressure transducers and its repetition prior to pumping tests; (3) proximity of the monitoring ports to the pumping location; (4) the pumping rate which affects the signal-to-noise ratio of observed test data; (5) presence or absence of constant-head or no-flow boundaries; (6) external stresses; and (7) aquifer diffusivities and their heterogeneity. Therefore, we recommend that in practice, more attention should be placed in designing cross-hole pumping tests to maximize the information content of the data sets and to minimize noise through signal-processing and/or noise reduction techniques.

3. We also examined the order of pumping tests included in the inversion algorithm and found that the order has a minimal impact on the inversion results for synthetic data but greatly affects the \( K \) tomogram using real data. 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 SSLE in its current form is more sensitive to noise during the beginning

Figure 10. Scatter plots of simulated vs. observed head at steady state. The simulated head values were obtained by simulating a cross-hole pumping test with pumping taking place at port 46 on the computer using the real \( K \) tomograms [(a) unconditioned case—Figure 3b; (b) low pumping rate case—Figure 4b; (c) changing the order of pumping tests case—Figure 5b; (d) conditioning using core \( K \) data case—Figure 6b; (e) conditioning using slug \( K \) data case—Figure 7b; and (f) conditioning using single-hole \( K \) data case—Figure 8b]. The observed head values in this figure are from an actual cross-hole pumping test conducted at port 46 in the laboratory aquifer (Figures 1a and b).
stages of K tomogram computation. This sensitivity is due to the uniform convergence criteria used for sequentially inverting all pumping test data. One improvement that could be made to the SSLE algorithm is to include an option for variable convergence criteria to account for different pumping tests with different noise levels. Furthermore, when hydraulic tomography is conducted using field data with SSLE in its current form, we recommend that the cleanest data be included into the SSLE algorithm first and progressively including noisier data.

4. Conditioning has been thought to help constrain inverse modeling results. We found that this is certainly the case for synthetic test data when using the SSLE algorithm. However, this study shows that the conditioning data itself can be subject to errors and so conditioning may not necessarily help in obtaining an improved solution. In particular, conditioning with data that are corrupted with noise can actually worsen the quality of the K tomograms. Therefore, we recommend that more attention should be paid to collection of better conditioning data and minimization of its errors.

5. We also examined the type of data used to conduct the conditioning of K tomograms. In this paper, we used synthetic and real core, slug, and single-hole K estimates. Our study showed that core data (assuming that they can be accurately obtained) improved both synthetic and real K tomograms. Conditioning of the K tomogram with slug and single-hole K estimates smoothed the K tomograms for this sandbox due to the larger support volume associated with these tests in comparison to the numerical grid used to compute the K tomograms. Therefore, we recommend scrutinizing the data type used to do the conditioning as not all data are created equally. Furthermore, we recommend that one should consider the support volume of data used to condition the K tomogram as it can have an impact on the resolution of the K tomogram.

Acknowledgments

This research was supported by the Strategic Environmental Research & Development Program (SERDP) under project ER-1365 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 SERDP and NSF. We thank Tian-Chyi J. Yeh for his review of the initial version of the manuscript, and also thank the three anonymous reviewers for their constructive comments, which have improved the manuscript.

References


