

A field assessment of the value of steady shape hydraulic tomography for characterization of aquifer heterogeneities

Geoffrey C. Bohling,¹ James J. Butler, Jr.,¹ Xiaoyong Zhan,^{1,2} and Michael D. Knoll³

Received 27 January 2006; revised 23 January 2007; accepted 1 February 2007; published 22 May 2007.

[1] Hydraulic tomography is a promising approach for obtaining information on variations in hydraulic conductivity on the scale of relevance for contaminant transport investigations. This approach involves performing a series of pumping tests in a format similar to tomography. We present a field-scale assessment of hydraulic tomography in a porous aquifer, with an emphasis on the steady shape analysis methodology. The hydraulic conductivity (K) estimates from steady shape and transient analyses of the tomographic data compare well with those from a tracer test and direct-push permeameter tests, providing a field validation of the method. Zonations based on equal-thickness layers and cross-hole radar surveys are used to regularize the inverse problem. The results indicate that the radar surveys provide some useful information regarding the geometry of the K field. The steady shape analysis provides results similar to the transient analysis at a fraction of the computational burden. This study clearly demonstrates the advantages of hydraulic tomography over conventional pumping tests, which provide only large-scale averages, and small-scale hydraulic tests (e.g., slug tests), which cannot assess strata connectivity and may fail to sample the most important pathways or barriers to flow.

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

1. Introduction

[2] A large body of previous work has demonstrated that spatial variations in hydraulic conductivity (K) play an important role in determining how a conservative solute will move in a saturated flow system. Numerous studies have shown that information about K variations is required both for reliable prediction of contaminant movement and for effective design of remediation systems. The field characterization of K variations on the scale of relevance for these applications, however, has proven to be a difficult task [Butler, 2005]. Commonly utilized methods yield either large-scale averages (pumping tests) that are of limited utility for predicting contaminant transport, or essentially point measurements (slug tests) that are insensitive to strata connectivity and may, as a result of the sampling interval, fail to detect important pathways or barriers to flow.

[3] Figure 1 illustrates the above points using results from an induced gradient tracer test and associated characterization activities at a heavily studied coarse sand and gravel alluvial aquifer [Bohling, 1999; Butler, 2005; Zemansky and McElwee, 2005]. The solid curve shows the K distribution as determined from the results of the tracer test, while the dashed curve is the large-scale average K determined from a pumping test. Clearly, reliance on the pumping-test estimate

would lead to an underprediction of contaminant movement in certain zones by greater than a factor of 3. The curves with symbols represent K estimates determined with direct-push slug tests over 0.3-m vertical intervals at selected locations within the aquifer. Although the slug-test estimates do provide information about some of the important controls on solute movement at the site, the highest K zone identified from the tracer test was not detected, possibly due to the relatively coarse vertical spacing of the slug tests. The consistency between slug-test profiles separated in space indicates the possibility that a considerable degree of lateral continuity exists at the site, but that continuity cannot be confirmed with single-well slug tests. This example demonstrates the need for methods that provide reliable information about the detail and connectivity of the K field on the scale of relevance for transport applications. In this paper we discuss one particularly promising method in this regard, hydraulic tomography.

[4] Hydraulic tomography [Neuman, 1987; Tosaka et al., 1993; Bohling, 1993; Gottlieb and Dietrich, 1995; Butler et al., 1999b; Yeh and Liu, 2000; Vesselinov et al., 2001a, 2001b; Bohling et al., 2002; Liu et al., 2002; Brauchler et al., 2003; Zhu and Yeh, 2005] involves performing a series of pumping tests in which different vertical intervals in an aquifer are stressed sequentially in a tomographic format. Drawdown is measured at multiple observation points during each test. Simultaneous analysis of data from the full suite of tests allows characterization of the K distribution between wells at a higher resolution than is provided by more conventional aquifer testing methods [Butler, 2005]. Despite its considerable potential, there are still a number of questions concerning the practical viability of hydraulic tomography and the quality of the information that it can

¹Kansas Geological Survey, University of Kansas, Lawrence, Kansas, USA.

²Now at MACTEC Engineering and Consulting, Inc., Oakland, California, USA.

³Center for Geophysical Investigation of the Shallow Subsurface, Boise State University, Boise, Idaho, USA.

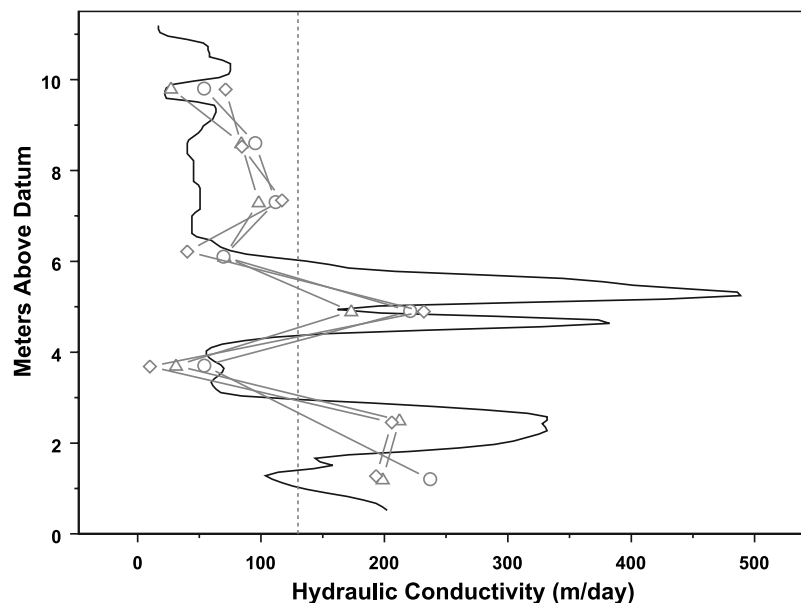


Figure 1. Hydraulic conductivity profiles developed from induced-gradient tracer test (GEMSTRAC1, solid curve) performed to the immediate northeast of wells Gems4N and Gems4S and from direct-push slug tests (HP1, triangles; HP8, circles; and DP808, diamonds). See Figure 2 for locations of wells and direct-push profiles. The vertical dashed line at 130 m/d represents the hydraulic conductivity estimate from a large-scale pumping test.

provide. The primary purpose of this paper is to investigate three critical issues that must be addressed if the considerable potential of this approach is to be realized.

[5] The first issue we address is that of nonuniqueness. Despite the high information density provided by a series of pumping tests performed in a tomographic format, the estimation of hydraulic conductivity from the observed drawdowns is still plagued by the nonuniqueness that typifies parameter estimation problems in the Earth sciences [Carrera and Neuman, 1986; Parker, 1994; Aster *et al.*, 2005]. Therefore an effective means of reducing the dimension of the parameter space (i.e., regularizing the inverse problem) is required to yield defensible K estimates. We address this issue here by reducing the number of unknown K values following the traditional approach of representing the K field as a relatively small number of constant-value zones, using both equal vertical spacing and cross-hole ground-penetrating radar surveys as the basis for the flow model zonation.

[6] The second issue we address is that of the computational efficiency of the methods used to analyze the suite of pumping tests performed in hydraulic tomography. Previous investigators have noted the large computational demands of a fully transient analysis of the drawdown data [e.g., Bohling *et al.*, 2002; Zhu and Yeh, 2005]. Hydraulic tomography data are analyzed here using the steady shape approach described by Bohling *et al.* [2002]. This method exploits the fact that within a given radial distance from a pumping well, the hydraulic gradients reach their ultimate steady state values before the heads themselves reach steady state. In many field situations, this steady shape head configuration is achieved long before actual steady state, if a true steady state response is reached at all [Kruseman and de Ridder, 1990]. In this case, the head configuration, and thus head differences within the region of investigation,

can be analyzed based on a steady state model, although the heads themselves cannot be. In addition, steady shape conditions, sometimes referred to as transient steady state [Kruseman and de Ridder, 1990], may be reached prior to the time when boundary conditions exert significant influence on the head response, meaning that a steady shape approach tends to reduce the influence of uncertain boundary conditions on the estimated K values. This is in contrast to a true steady state approach, as described by Yeh and Liu [2000] and Illman and Neuman [2003], since the steady state head configuration is always influenced by the boundary configuration.

[7] The third issue we address is that of the quality of the information provided by the hydraulic tomography procedure. There is a recognized need to assess the performance of this approach in well-controlled field settings [e.g., Liu *et al.*, 2002]. We address this need here by comparing the results of a series of hydraulic tomography experiments performed in an extensively studied coarse sand and gravel aquifer with estimates obtained using other methods (e.g., Figure 1) [Butler, 2005]. Previous assessments of hydraulic tomography in porous flow systems have been performed in idealized lab conditions with limited transferability to actual field settings (e.g., see discussion by Liu *et al.* [2002]), so this is the first field assessment of the approach in a saturated porous formation. Vesselinov *et al.* [2001a, 2001b] describe simultaneous three-dimensional inversion of a set of pneumatic cross-hole tests in unsaturated fractured tuff, essentially amounting to pneumatic tomography.

[8] This paper begins with an overview of the research site and the experimental procedures used in this work. Following this overview, the methods utilized for the steady shape and transient analysis of the drawdown data are reviewed. Zonation strategies and the drawdown analysis are then described for both the equal-thickness and radar-

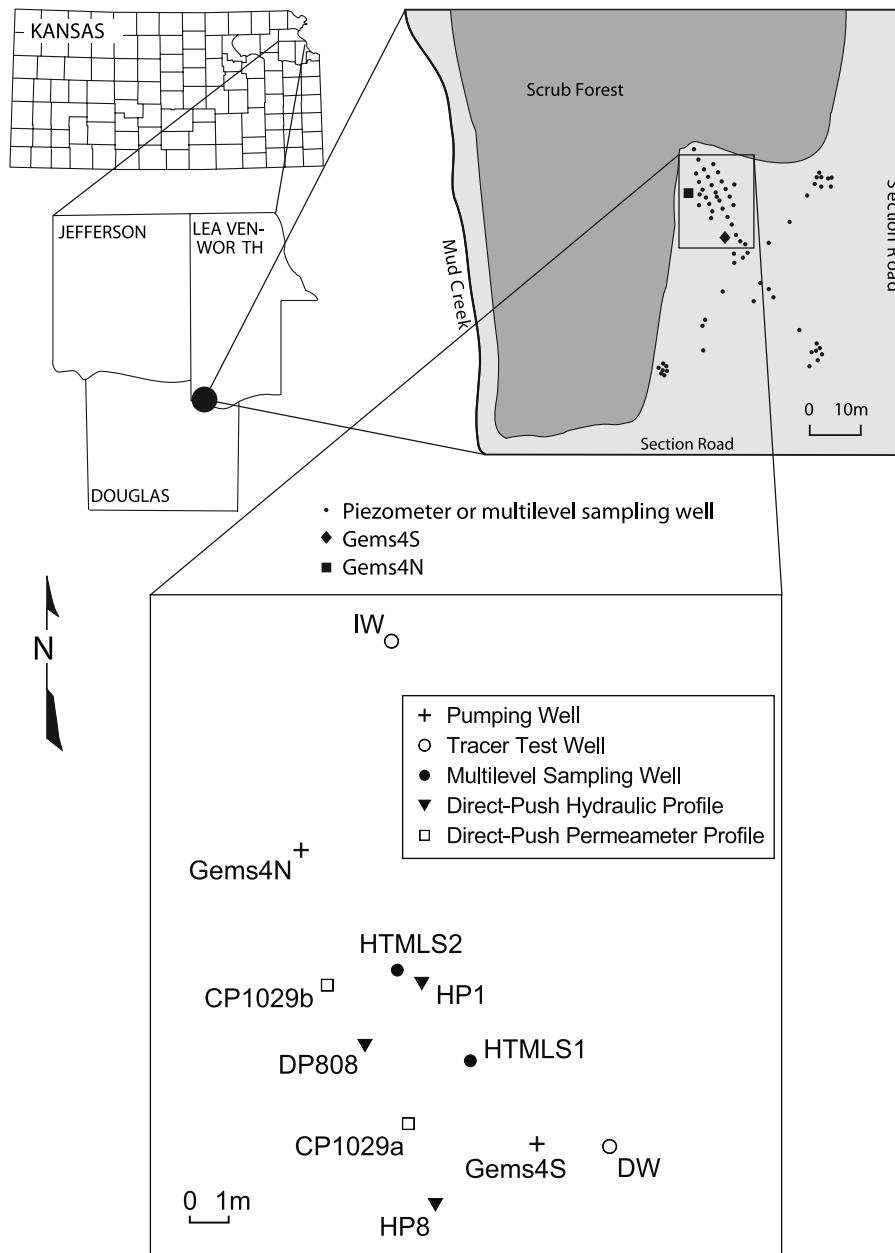


Figure 2. Location of the Geohydrologic Experimental and Monitoring Site (GEMS) and wells used in current study.

based zonations, with a particular emphasis on the seven-layer case. The results of the tomography analysis are then compared with hydraulic conductivity estimates obtained in a follow-up investigation using a new direct-push method. The paper concludes with a summary of the major findings of the investigation and some brief comments on the limitations of the study.

2. Experimental Setup

[9] The hydraulic tomography experiments were performed at the Geohydrologic Experimental and Monitoring Site (GEMS), a heavily studied site of the Kansas Geological Survey in the Kansas River valley northeast of Lawrence, Kansas (Figure 2). The alluvial aquifer at the site

consists of 11 m of sand and gravel overlain by 11 m of silt and clay. The shallow stratigraphy at the site is shown in Figure 3, together with an electrical conductivity profile obtained using a direct-push probe [Schulmeister *et al.*, 2003]. Butler [2005] and Zemansky and McElwee [2005] summarize much of the previous hydraulic characterization research at GEMS. Experience at the site has demonstrated that the alluvial aquifer behaves as an ideal, perfectly confined system over the pumping durations used for the hydraulic tomography experiments. All of the simulations performed for this study assume confined conditions.

[10] Gems4S and Gems4N, each 11 cm in diameter, were used as the pumping wells for the tomography experiments (Figure 4). Gems4N and Gems4S are both constructed of PVC and were installed inside hollow-stem augers with

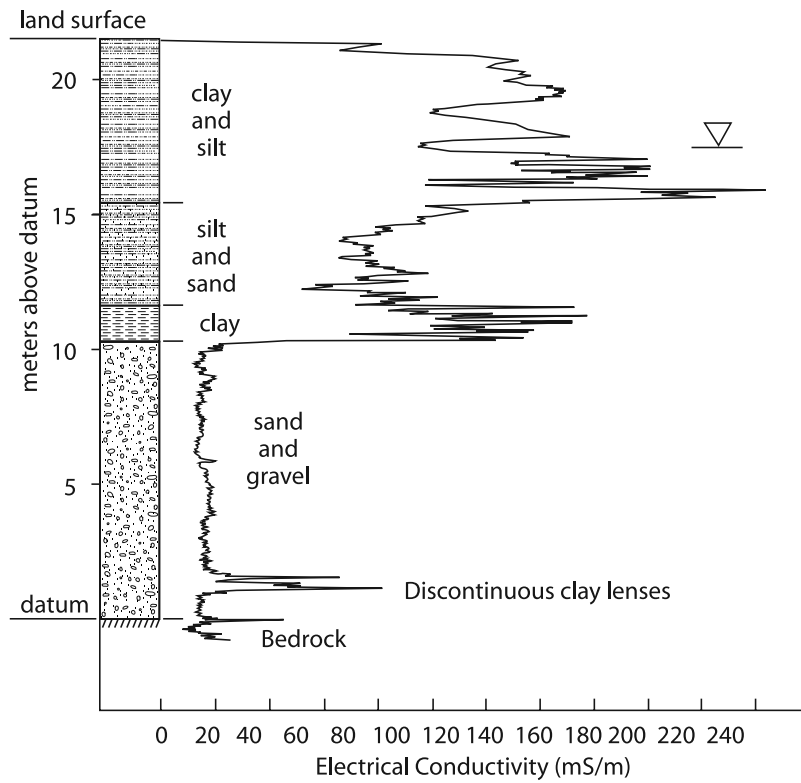


Figure 3. Shallow stratigraphy at GEMS together with electrical conductivity profile after *Butler et al.* [1999a].

28.6-cm outer diameter flights. The formation was allowed to collapse back with withdrawal of the auger flights. For each test, we utilized packers to isolate a 0.6-m interval in the well, and then pumped that interval at a roughly constant

rate of 1.3 L/s. Drawdowns were measured using pressure transducers installed in observation wells HTMLS1 and HTMLS2. Each of these observation wells is constructed from seven-chamber PVC pipe with a screened opening in

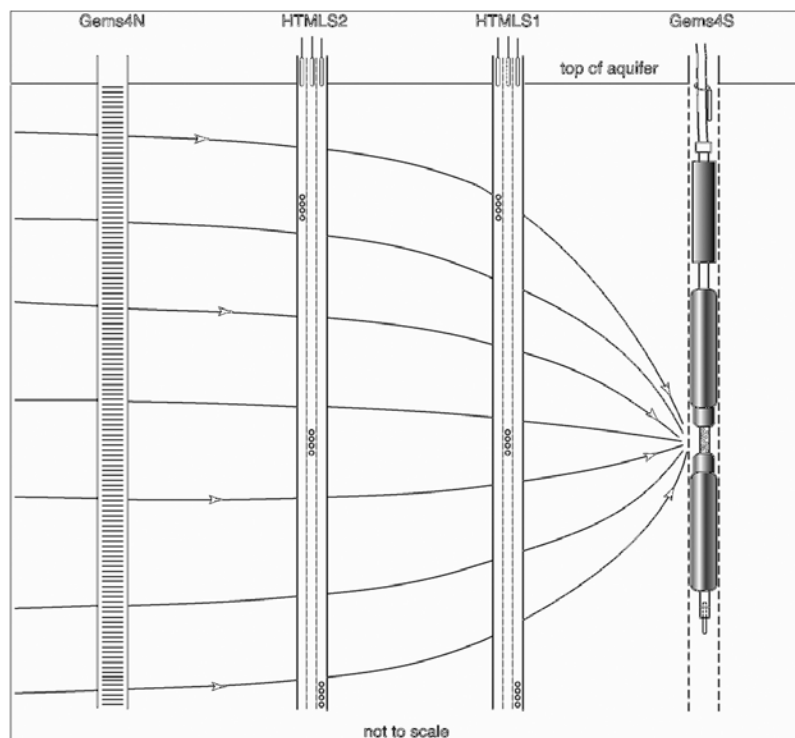


Figure 4. Experimental setup of tomographic pumping tests (see Figure 2 for well location map).

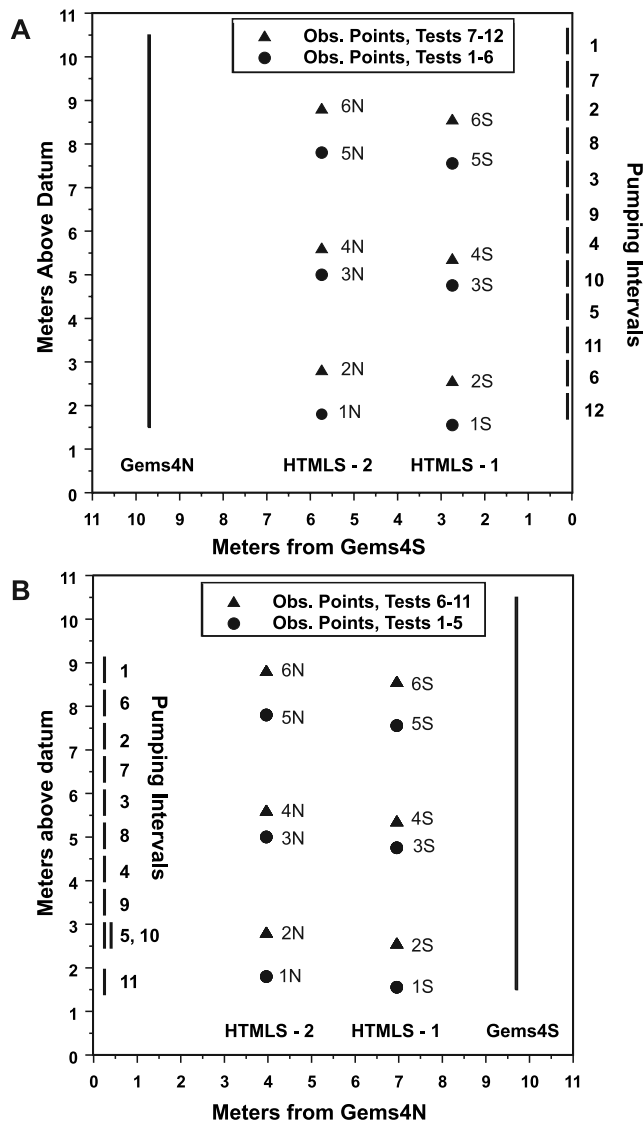


Figure 5. Pumping intervals and observation point locations for (a) 12 tests with pumping in Gems4S and (b) 11 tests with pumping in Gems4N.

just one chamber at each sample depth (Solinst continuous multichannel tubing [Einarson and Cherry, 2002]). The multichamber PVC pipe has an outer diameter of 41 mm and was installed inside a direct-push pipe with an outer diameter of 83 mm, again allowing the formation to collapse naturally upon withdrawal of the direct-push pipe. Changes in drawdown (pressure) at a particular depth were measured with a pressure transducer (Druck PDCR 35/D 103.42 kPa (15 psi) gauge sensor) in the corresponding chamber. Flow rate was measured electronically with a paddle wheel flowmeter (Omega FP-5800), and manually with a calibrated bucket and a stopwatch. The flowmeter and all of the pressure transducers were connected to the same data logger, a Campbell Scientific 23X with an acquisition rate of two samples per second (2 Hz).

[11] In this study, we compare K estimates obtained from the tomographic pumping tests with estimates obtained from direct-push slug tests [Butler *et al.*, 2002; McCall *et*

al., 2002; Sellwood *et al.*, 2005] at the locations labeled HP1, HP8, and DP808 in Figure 2, and to a K profile developed from GEMSTRAC1, an induced-gradient tracer test [Bohling, 1999]. As shown in Figure 2, HP1 is about 0.6 m from HTMLS2 and very close to the line connecting Gems4S and Gems 4N, DP808 is about 1.8 m to the southwest of HP1, and HP8 is about 2.6 m southwest of Gems4S. The induced-gradient tracer test was performed just to the northeast of the Gems4S-Gems4N line, with tracer injection in the well labeled IW and extraction from the well labeled DW in Figure 2. Figure 1 shows the K profiles derived from the direct-push slug tests and GEMSTRAC1.

[12] Direct-push slug tests can provide accurate estimates of the hydraulic conductivity in the immediate vicinity of each test interval, but cannot yield definitive information on strata connectivity. A tracer test, however, probes the material between wells, so the analysis of tracer-test data can provide high-resolution information regarding the connectivity of potential transport pathways. For the GEMSTRAC1 tracer test, we monitored the pumping-induced movement of a bromide tracer through a network of multilevel sampling wells between wells IW and DW over the course of 1 month [Bohling, 1999]. We estimated a vertical profile of relative flux rates based on tracer breakthrough curves measured in a number of multilevel sampling ports in the network, and then multiplied the relative flux rate profile by an estimate of the average horizontal conductivity at the site (130 m/d) to obtain a K profile. The flux rate profile was developed by fitting a modified version of an analytical radial transport model proposed by Moench [1989] to tracer breakthrough curves at individual ports and then developing a composite vertical profile of the flux terms from the individual fits, assuming that the test was dominated by horizontally stratified flow. This assumption appeared to be reasonably well satisfied in the upgradient half of the tracer network, and the composite flux profile is strongly weighted toward the results at this northern end of the network.

[13] As detailed by Bohling [1999], analysis of this test was complicated by a nonuniform vertical distribution of the injected tracer mass; most of the mass was apparently drawn into high-conductivity zones lower in the aquifer and traveled quite rapidly through those zones to the extraction well, a problem which could possibly have been avoided or at least reduced by injecting the tracer in equal amounts over a sequence of short packed-off intervals. Our inability to introduce sufficient tracer into the upper half of the aquifer resulted in an undersampling of the properties of that region and possibly resulted in an overestimation of K values in the lower half of the aquifer. In addition, the nature of the tracer test did not allow us to estimate porosity variations independently of flux rate variations. Although the conversion of the relative flux rate profile to the K profile did not require an explicit estimate of porosity, it is possible that unaccounted for variations in porosity could have influenced the resulting K estimates. Core sample measurements indicate porosities generally ranging between 20% and 30%, but a very high degree of short-scale variability in these measurements makes it difficult to discern systematic variations that could be accounted for in computing the K profile.

[14] Figure 5 represents the experimental sequence for the tomographic pumping tests. Figure 5a shows the locations of the pumping intervals and observation points for the 12 tests with pumping in Gems4S. Tests 1–6 were performed with increasing pumping interval depth, then the packer string was pulled back up and tests 7–12 were performed. Observations were obtained at six locations during each test, with three transducers installed in each of the two multilevel sampling wells. Between tests 6 and 7, the transducers were relocated (moved to different chambers), so that tests 7–12 used a different set of observation locations than tests 1–6.

[15] Figure 5b shows the locations of pumping intervals and observation points for the 11 tests with pumping in Gems4N. Although the sequence of operations for these tests was similar to those in Gems4S, with the transducers being relocated between two sets of tests, the sequence of tests in Gems4N is more irregular for two reasons: (1) Both tests 5 and 10 employed the same pumping interval (2.5–3 m above datum), and (2) the pumping test in the shallowest isolated interval at this well was subject to some logistical difficulties and so has not been included in this sequence.

[16] Datum in Figures 5a and 5b corresponds roughly with the bottom of the aquifer, and is therefore taken as the aquifer bottom for analysis purposes. An aquifer thickness of 10.67 m (35 feet) was used in all of the analyses.

3. Assessment of Steady Shape

[17] For each test, data were recorded at half-second intervals over a period of at least 100 s from initiation of pumping. For the analyses described here, we have employed a subset of data running from 20 to 70 at 2-s increments, a total of 26 observations for each observation point in each test. Overall, drawdown data from this time interval exhibited a constant, common slope versus log time at all observation points for each test and the differences between drawdowns at different observation locations were roughly constant over time. This behavior is the signature of steady shape conditions [Bohling *et al.*, 2002].

[18] For constant-rate pumping from a well fully penetrating a confined aquifer, Butler [1988] demonstrates that the time required to reach steady shape conditions is the same as the time at which the Cooper and Jacob [1946] semilog approximation becomes valid. This time is given by $t > 100 r^2 S / 4T$, where r is distance from the pumping well, S is storativity, and T is transmissivity. For the alluvial aquifer at GEMS, T is approximately $0.016 \text{ m}^2/\text{s}$ and S is of the order of 4×10^{-4} , implying that steady shape conditions would be achieved at around 5 s under fully penetrating pumping at a distance of 2.7 m, the smallest observation radius in our tests, and around 30 s at 7.0 m, the largest observation radius. Effects of partial penetration and heterogeneity will modify the time to steady shape somewhat, but this estimate indicates that we could expect to be at or very close to steady shape conditions at 20 s into each test at all observation locations.

[19] In order to assess the attainment of steady shape conditions in the field tests, involving partially penetrating pumping under heterogeneous conditions, we examine here the full set of data (at 1-s intervals out to 900 s) for two tests in Gems4N: test 1, with a pumping interval about 8.5–9 m

above datum, and test 5, with a pumping interval from 2.5–3 m above datum. These two tests involve some of the largest separations between pumping interval and observation point. The distance from the pumping interval for test 1 and observation point 1S is about 10 m (Figure 5). We also examine the behavior of drawdowns predicted for these tests using one of the estimated K profiles developed later in the paper, namely, the profile derived from transient analysis of the 23 tests using a seven-layer zonation derived from the cross-hole radar profiles (dashed line in the seven-zone plot of Figure 12). Examining the modeled drawdowns for these two tests using this K field, one that we feel is reasonably representative of conditions at the site, represents an a posteriori assessment of the validity of the steady shape assumption.

[20] The points in Figure 6a represent the full set of observed drawdowns versus the logarithm of time for tests 1 and 5 in Gems4N. The arrangement of panels in the plot reflects the spatial positioning of the observation points (Figure 5). Between about 20 and 100 s the data for both tests display a nearly constant slope versus log time. Past 100 s the slopes increase, indicating the influence of an additional mechanism, most likely interference from intermittent pumping at nearby wells. The lines on the plots represent the modeled drawdowns for the two tests. (It is important to keep in mind that the K profile and storage coefficient used in the model have not been adjusted solely to match the data sequence shown here, but instead represent a compromise fit to 20–70 s data from all 23 tests.) Figure 6b shows the differences at each observation time between the drawdown at each observation point and the drawdown at observation point 3N, chosen somewhat arbitrarily as the reference point for this presentation. These differences are roughly constant over the 20–70 s time interval. Clearly, there is some slight drift in the differences over this time interval, and more so over the full time span, but this drift is quite possibly due to factors other than lack of attainment of steady shape. The lines in Figure 6b represent the differences of the corresponding modeled drawdowns and by 20 s these differences are already very close to their final values at all observation locations, and essentially at their final values from 30 s onward. Thus our model supports the assumption that drawdown differences over the 20–70 s time interval are essentially the same as those that would exist under steady state conditions.

[21] One motivation for including data starting at 20 s, rather than using a later data segment (for example, 50–70 s) is that several of the tests were influenced by initiation or termination of pumping at neighboring high-capacity wells and these changes tended to introduce trends into the later time data. Using earlier time data helped to reduce the influence of these trends.

4. Cooper-Jacob Analyses

[22] Figure 7 shows observed drawdowns versus the logarithm of time for all 12 tests with pumping in Gems4S. The drawdown versus time plots for the Gems4N tests are similar in character. In all cases, the drawdown versus log-time plots exhibit a nearly constant slope from approximately 20 to 70 s after test initiation. In addition, there is an approximately constant offset between drawdown plots from different observation points. This behavior corre-

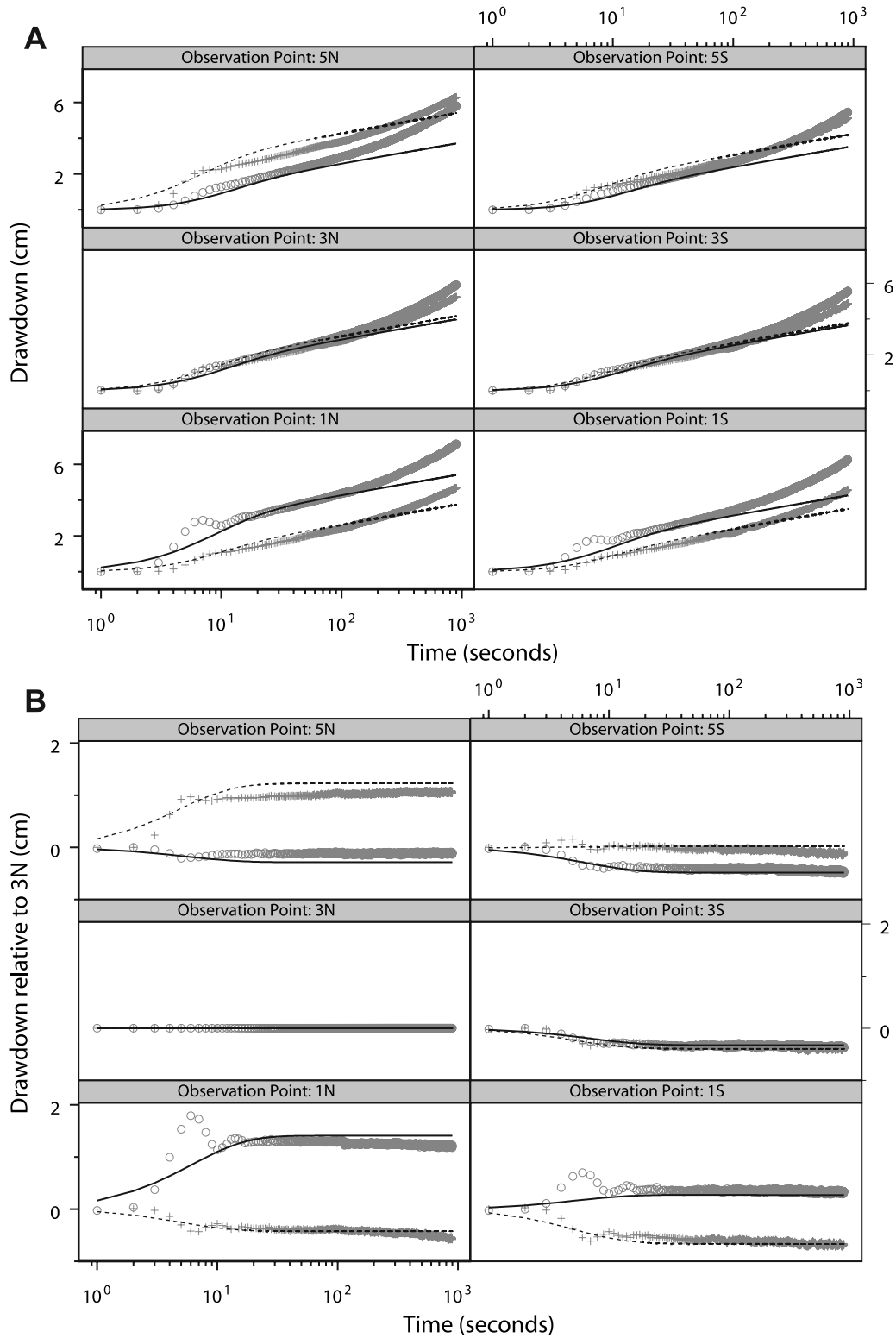


Figure 6. (a) Full set of drawdown data for tests 1 (plus signs) and 5 (circles) in Gems4N. (b) Drawdowns relative to drawdown at observation point 3N at same time. Curves (dashed for test 1, solid for test 5) represent drawdowns and drawdowns relative to 3N predicted by one of the seven-layer models developed in this study (see discussion in text). See Figure 5 for pumping interval and observation point locations.

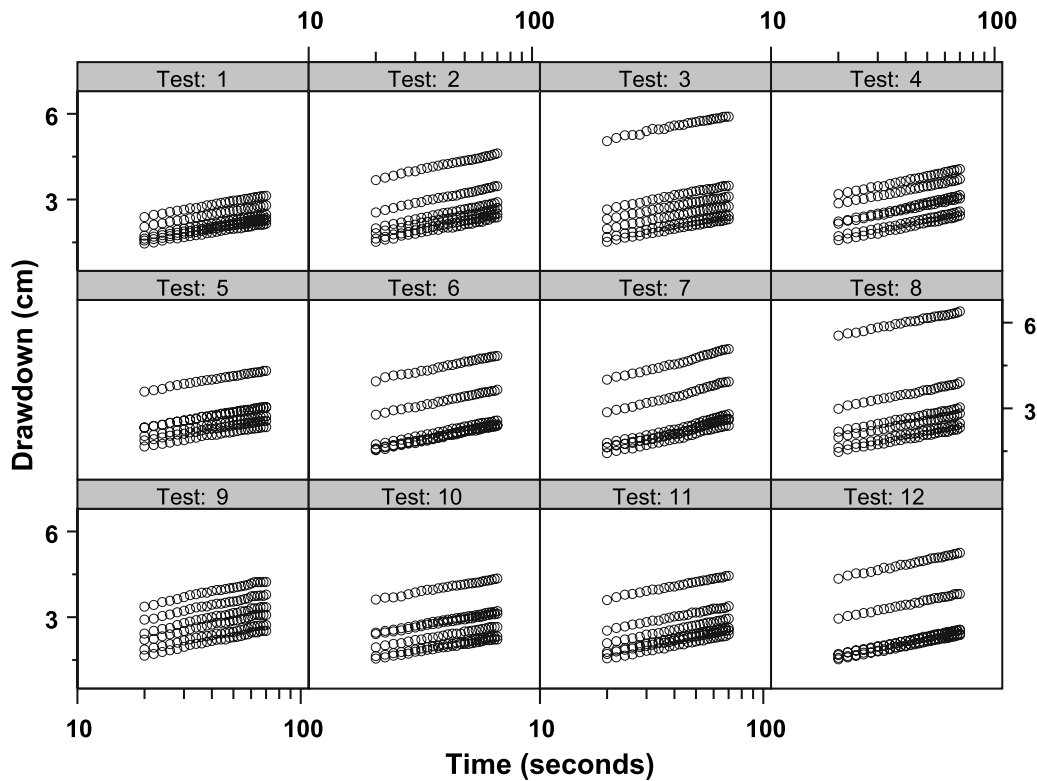


Figure 7. Drawdowns measured at six observation points (each line of points) over the 12 tests involving pumping in Gems4S.

sponds to a “steady shape” drawdown configuration: Gradients in the region of investigation are no longer changing, although drawdown is still increasing overall (see Figure 2 of Bohling *et al.* [2002]). Under steady shape conditions, the slope of the drawdown versus log time plot is controlled by the bulk average horizontal hydraulic conductivity [Butler, 1990]. We have estimated a set of K values from Cooper-Jacob analyses [Cooper and Jacob, 1946; Butler, 1990] of the drawdown versus log-time plots for all the tomographic pumping tests, one for each of the six observation points in each of the 23 tests, for a total of 138 estimates. The Cooper-Jacob analyses of the 138 slopes yield K estimates with a mean of 131 m/d and a standard deviation of 16 m/d. The variations around the mean are due in part to violations of the assumption of homogeneity used in the analyses and in part to subtle variations in slope produced by the chance turning on or off of distant pumping wells during a test. Fitting a model with a common slope but separate intercepts to the 138 records yields a slope of $0.0152 \text{ m}/\log_{10}(s)$, corresponding to a hydraulic conductivity of 129 m/d. Thus 130 m/d seems to be a reasonable estimate for the overall average horizontal hydraulic conductivity. This is equivalent to the average conductivity value, derived from earlier large-scale pumping tests, used to convert the GEMSTRAC1 relative flux profile into the K profile shown in Figure 1.

5. Data Analysis Methodology

[23] The primary analysis method used here was the steady shape approach of Bohling *et al.* [2002]. Under

steady shape conditions, the drawdown differences between observation points in the region of investigation are controlled by the K distribution in that region and are the same as the differences that would exist under steady state conditions, assuming the pumping rate remains unchanged. In this case, the drawdown differences can be modeled and fit using steady state simulations of the pumping tests, rather than transient simulations. This decreases computational time for the analyses by 1–2 orders of magnitude relative to a transient simulation and reduces the influence of poorly known boundary conditions relative to analyzing the drawdowns themselves using a steady state model [Bohling *et al.*, 2002].

[24] The observations to be matched with the steady shape analysis are the differences in drawdown between all 15 possible pairs of the six observation points for each of the 26 observation times (2-Hz data sampled from 20 to 70 s at 2-s increments) for each test. Since the differences are approximately constant over time (Figure 7), we obtain 26 repeat measurements of each of the 15 pairwise drawdown differences between the six observation ports for each test. This leads to 390 observed drawdown differences per test, for a total of 8970 observations over all 23 tests.

[25] To simulate and fit the data, we use a two-dimensional radial-vertical finite difference flow model coupled with the Levenberg-Marquardt nonlinear regression algorithm [Bohling and Butler, 2001]. The flow model utilizes a logarithmically transformed radial coordinate given by $r' = \ln(r/r_w)$, where r is the actual radial distance from the center of the pumping well and r_w is the pumping well radius. This

transformation allows radial flow to be simulated with a regular Cartesian (rectangular) grid. The model employed here uses 60 cells in the horizontal at a spacing of $\Delta r' = 0.2$ (dimensionless) and 70 cells in the vertical at a spacing of $\Delta z = 0.152$ m. We also include an extra (zeroth) column of nodes to simulate the well bore, with a combination of high- and low-conductivity cells representing the open well bore and packer configuration for each test. This enables the model to incorporate the damping of vertical gradients that results from bypass flow along the well bore, a potentially important mechanism for hydraulic tomography experiments performed in porous formations.

[26] Previous work has found that conventional slug and pumping tests at GEMS are influenced by inertial effects due to the high permeability of the aquifer [Butler and Zhan, 2004]. Although the early time data from the tomographic pumping tests show evidence of the oscillatory behavior indicative of inertial effects (Figure 6a), we only analyzed data obtained after this early oscillatory period, following the findings of Butler and Zhan [2004]. Thus it was not necessary to incorporate inertial effects into the flow model.

[27] We used layered zonations of the model aquifer for this study, with all cells in a given layer assigned a single K value. Tests were first run with equal-thickness layers and then with variable-thickness layers based on a zero-offset radar survey run between Gems4S and Gems4N. In this paper, we present results based on analyzing all 23 tests, the suite of 12 Gems4S and the suite of 11 Gems4N tests, simultaneously. The assumption of a perfectly layered aquifer allows us to treat the parameter models as equivalent between the two sets of tests, despite the fact that the radial coordinate system is centered on Gems4S in one case and on Gems4N in the other, allowing for the simultaneous analysis of all 23 tests.

[28] In the steady shape approach, the nonlinear regression algorithm adjusts the hydraulic conductivity values for the different layers in an attempt to minimize a chi-square objective function given by

$$\chi^2 = \sum_{i=1}^n \left[\left(dd_i^{obs} - dd_i^{prd} \right) / \sigma_{dd} \right]^2, \quad (1)$$

where n is the number of drawdown differences considered, dd_i^{obs} is the i th observed drawdown difference, dd_i^{prd} is the corresponding difference predicted by the model, and σ_{dd} is the estimated measurement error (standard deviation) for the drawdown differences. On the basis of pressure transducer characteristics and the residuals from the Cooper-Jacob analyses mentioned above, we have estimated the measurement error for the drawdown differences as $\sigma_{dd} = 1 \times 10^{-4}$ m. This value is quite small relative to actual deviations between observed and predicted drawdown differences, leading to large χ^2 values, meaning that none of the models comes close to matching the data to within measurement error. It is possible that a larger value for σ_{dd} would be more appropriate, taking into account more factors than just measurement error. However, the value is constant for all observations, so the estimated K values are the same as those that would be obtained from unweighted regression. In addition, the relative variations in χ^2 between different

analyses would be the same regardless of the choice of σ_{dd} . Accordingly, we will use the raw root-mean-square (RMS) residual between observed and predicted drawdown differences as the summary fit statistic, rather than the χ^2 value. The RMS residual is given by

$$s = \sqrt{\left(\sum_{i=1}^n \left(dd_i^{obs} - dd_i^{prd} \right)^2 \right) / (n - p)}, \quad (2)$$

where p is the number of estimated parameters (K values), so that $n - p$ represents the degrees of freedom for the fit.

[29] For comparison purposes, we have also analyzed the transient drawdown through time responses measured at each observation point. In the transient analysis, the objective function is the more conventional sum of squared (scaled) residuals between the observed and predicted drawdowns. One potential difference between the transient and steady shape approaches is that the transient approach “sees” the change in drawdown through time, which is governed by the large-scale average hydraulic conductivity, whereas the steady shape approach filters out this time variation and thus might be less constrained to produce a K profile that reproduces the large-scale average.

[30] For both the steady shape and transient analyses, we assume isotropy for hydraulic conductivity in each layer, since we have not seen any evidence of significant anisotropy in these sediments. Thus, for the steady shape analyses, the only fitting parameters are the single K values for each layer. For the transient analyses, we also fit a single specific storage (S_s) value for the entire aquifer. Previous investigators [e.g., Zhu and Yeh, 2005] have estimated specific storage as part of an analysis of synthetic hydraulic tomography data. However, the pumping-induced drawdowns measured in tomography experiments are primarily a function of the hydraulic diffusivity (K/S_s) of the material between the observation point and the pumping interval, and the large-scale average K of the aquifer [Brauchler et al., 2003]. Extracting information about K and S_s from the diffusivity parameter can be difficult without a priori estimates of both, which are rarely available at any scale (S_s) or at the scale appropriate for a tomography analysis (K). Thus, as with conventional pumping tests [e.g., Butler, 1990; Schad and Teutsch, 1994; Sanchez-Vila et al., 1999], obtaining a reliable estimate of specific storage from the tomography drawdown data presents a significant challenge. The most promising approach for obtaining reliable estimates of S_s would be to estimate K with a steady shape analysis, and then extract information on S_s from the hydraulic diffusivity determined in the transient analysis using the estimated K distribution. For this initial field assessment, however, we choose to avoid this additional complexity by invoking the pragmatic assumption of a constant specific storage.

6. Regular Layer Zonations: Analysis Results

[31] Figure 8 shows the estimated K profiles obtained from steady shape and transient analyses of the tomographic pumping tests using four different layered zonations, in which the 10.67-m-thick aquifer is divided into 5, 7, 10, and 14 equal-thickness layers (henceforth, regular-layer zonation).

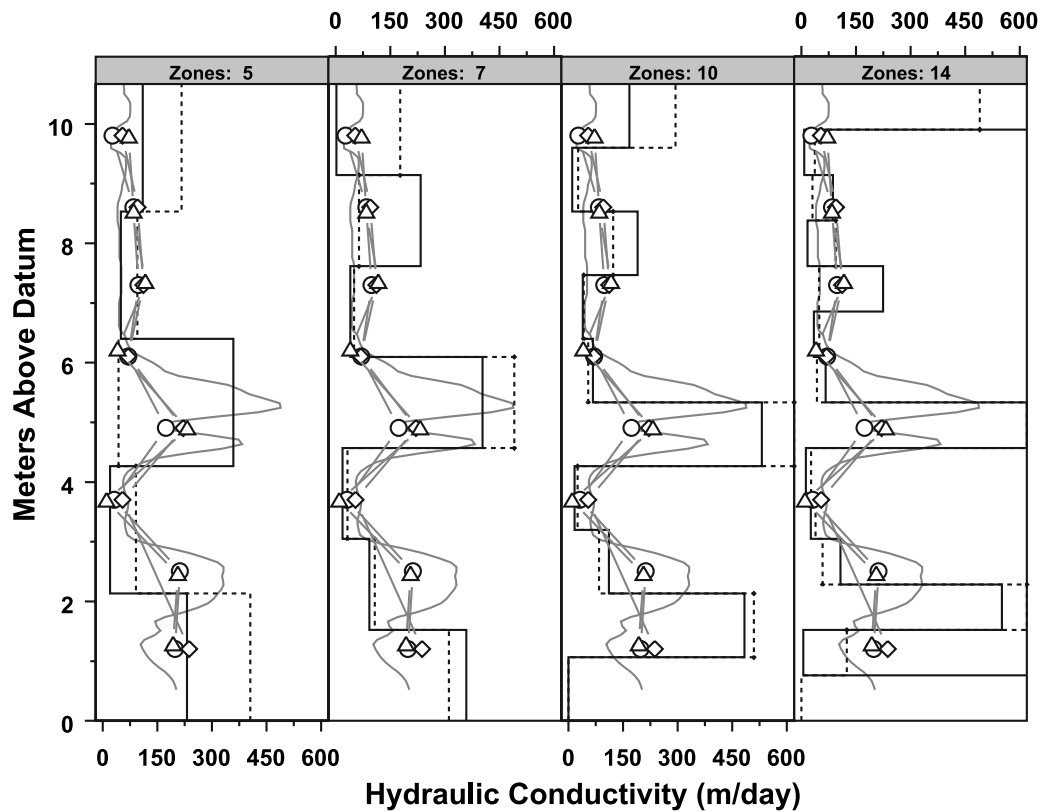


Figure 8. Estimated hydraulic conductivities from steady shape hydraulic tomography (solid curve) and transient hydraulic tomography (dashed curve) using 5, 7, 10, and 14 equal-thickness zones (layers) compared with K profiles from the direct-push slug tests (HP1, triangles; HP8, circles; DP808, diamonds) and the GEMSTRAC1 tracer test (shaded curve).

tion). Since the flow model is discretized into 70 equal-thickness cells in the vertical direction, each zone of the five-layer zonation comprises fourteen cells in the vertical, and so forth down to five cells per layer for the 14-layer zonation. These results are compared with K profiles from the GEMSTRAC1 tracer test and the direct-push slug tests at HP1, HP8, and DP808. Table 1 contains the corresponding summary statistics for each analysis, including the overall

thickness-weighted average of the hydraulic conductivity estimates and the RMS residual between observed and predicted drawdown differences for steady shape tomography, or between observed and predicted drawdowns for transient tomography. The homogeneous (single-layer) analysis statistics are included in Table 1 for comparison, as are the results from the radar-based zonations discussed below. Figure 9 shows the RMS residual versus

Table 1. Summary Fit Statistics for Hydraulic Tomography With Regular-Layered and Radar-Based Zonations^a

Layering	Number of Zones	Average K, m/d		RMS Residual, cm		Rcond(J'J)		Log10[det(cov)]	
		Steady Shape	Transient	Steady Shape	Transient	Steady Shape	Transient	Steady Shape	Transient
NA	1	95.8	124.5	0.606	0.475	1.0E+00	7.0E-03	-10.9	-11.3
Regular	5	153.9	170.1	0.410	0.407	2.7E-02	1.7E-03	-9.7	-7.2
Regular	7	164.2	176.3	0.301	0.342	3.9E-03	1.3E-03	-7.0	-4.7
Regular	10	161.6	188.1	0.246	0.280	9.6E-12	6.3E-06	5.7	2.2
Regular	14	278.7	195.4	0.203	0.273	4.9E-04	4.1E-07	4.5	12.8
Radar	5	145.0	168.3	0.497	0.408	3.9E-02	1.4E-03	-9.0	-7.4
Radar	7	159.1	171.0	0.326	0.340	5.2E-03	2.8E-03	-7.3	-5.1
Radar	10	165.7	185.8	0.286	0.261	7.8E-04	5.1E-04	-1.2	0.3
Radar	13	151.4	187.5	0.324	0.205	8.7E-07	2.4E-05	7.5	5.9

^aAverage K is the thickness-weighted average hydraulic conductivity over the set of zones (layers); RMS Residual is the square root of the mean-square residual between observed and predicted drawdown differences for steady shape tomography or between observed and predicted drawdowns for transient tomography; Rcond(J'J) is the reciprocal condition number of the inner product of the final Jacobian for the fit; and Log10[det(cov)] is the base 10 logarithm of the determinant of the estimated parameter covariance matrix for the fit.

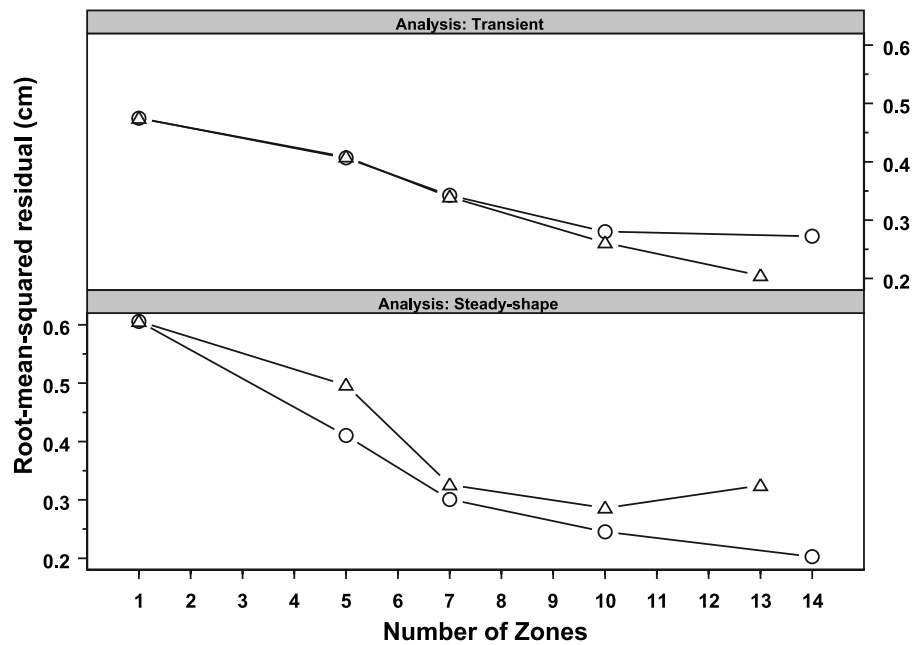


Figure 9. Root-mean square (RMS) residual versus number of zones (layers) for steady shape and transient analyses of tomographic pumping tests using both regular-layered zonations (circles) and zonations based on zero-offset radar profiles (triangles). Residual is between observed and predicted drawdown differences for steady shape analyses and between observed and predicted drawdowns for transient analyses.

number of layers for both the transient and steady shape analyses. Figure 9 also includes the results for the radar-based zonations.

[32] For both the transient and steady shape approaches, the RMS residual decreases as the number of zones increases (Table 1 and Figure 9). This is as expected because the model should match the data more closely as the number of adjustable parameters increases. Ideally, plots of the RMS residual versus number of layers would cease to decrease significantly beyond a certain number of zones, indicating a point of diminishing returns from increasing model complexity. For the regular-layered zonations, the RMS residuals for the transient analyses do not decrease significantly going from the 10- to 14-layer zonations. However, as can be seen in Figure 8, the K estimates from the 14-layer case do not compare well with the existing K data, as estimates for a number of the layers are very high and off the scale of the plot. Upper and lower bounds were set for the K estimates in the optimization process, but these bounds were well outside the expected K range for the aquifer and did not constrain the optimization procedure in any of the analyses. Because of the highly significant lack of fit (relative to estimated measurement error) for all zonations, statistics including terms adjusting for the number of fitted parameters, such as the Akaike information criterion or Bayesian information criterion [Hastie et al., 2001; Carrera and Neuman, 1986], do not provide any more guidance for model selection beyond that provided by the RMS residuals shown here.

[33] Table 1 also contains diagnostic information on the overall reliability of the fits in each case, estimated from the Jacobian (sensitivity) matrix evaluated at the final parameter vector. This includes the reciprocal condition number of the

inner product of the Jacobian, $J'J$, and the base 10 logarithm of the determinant of the estimated parameter covariance matrix. Note that the estimated parameters for the transient fits include the global storage coefficient estimate, so sensitivity to storage is included in the Jacobians for these fits. A reciprocal condition number close to machine precision (about 10^{-16} , since we are using double precision) would indicate that the estimation problem is close to singular, a problem that often results from attempting to estimate indistinguishable parameters [Press et al., 1992]. The determinant of the covariance matrix is a summary measure of the overall uncertainty in the parameter estimates, accounting for parameter correlation as well as individual parameter variances and, in general, we would expect this to increase with the number of estimated parameters. For most of the regular-layer fits, the condition numbers of the final Jacobian (sensitivity) matrices are quite mild and therefore do not show significant evidence of overparameterization. In addition, the estimated pair-wise parameter correlations for most of the fits are quite mild, usually less than 0.5 in magnitude. However, the steady shape fit using 10 regular layers is clearly problematic, with a reciprocal condition number of 9.6×10^{-12} and with a covariance determinant that is out of trend, higher than that for the 14-layer steady shape fit. The 14-layer transient fit is also problematic, with a very large covariance determinant. In fact, of the regular-layer fits, only three show correlations with magnitudes greater than 0.9: the single-zone transient fit, with a correlation of -0.9457 between the global estimates for storage and K; the 10-regular-layer steady shape fit, with a correlation of 0.9310 between the K estimates for layers 1 and 2 at the bottom of the aquifer; and the 14-regular-layer transient, with a correlation of

Table 2. Run Times for Steady Shape and Transient Analysis of 23 Tomographic Pumping Tests Using Different Zonations

Layering	Number of Zones	Run Time, min	Number of Parameter Iterations	Number of Forward Simulations	Run Time per Forward Simulation, min
<i>Steady Shape</i>					
NA	1	1.98	4	9	0.22
Regular	5	12.67	11	70	0.18
Regular	7	17.28	11	96	0.18
Regular	10	24.84	12	139	0.18
Regular	14	68.33	25	399	0.17
Radar	5	20.23	19	116	0.17
Radar	7	23.36	15	133	0.18
Radar	10	32.57	16	186	0.18
Radar	13	17.11	5	89	0.19
<i>Transient</i>					
NA	1	221.33	4	16	13.83
Regular	5	657.42	13	95	6.92
Regular	7	731.65	10	104	7.04
Regular	10	757.23	7	88	8.60
Regular	14	3902.35	32	540	7.23
Radar	5	598.68	11	88	6.80
Radar	7	543.67	7	77	7.06
Radar	10	1511.46	17	222	6.81
Radar	13	2255.69	22	337	6.69

–0.9997 between the K estimates for layers 2 and 3 near the bottom of the aquifer. Parameter correlations near 1 in magnitude indicate an inability to obtain independent estimates of the corresponding parameters from the data, resulting in an ill-conditioned inverse problem.

[34] Although the diagnostic statistics described in the previous paragraph do not show strong indications of over-parameterization in most of the fits, this does not mean that the estimated parameters can be accepted as either unique or globally optimal. These diagnostics are computed on the basis of a quadratic approximation of the objective function about the final parameter vector, also assuming that the inner product of the Jacobian matrix provides a reasonable approximation for the Hessian (second derivative) matrix. The true objective function could be far from quadratic, globally, and the inner product of the Jacobian could provide a poor approximation for the Hessian, especially since the residuals are fairly large for this problem [Aster *et al.*, 2005; Press *et al.*, 1992]. For the steady shape analyses, we have explored the issues of uniqueness and global optimality more thoroughly using a simulated quenching approach described in section 7.

[35] The average K values (Table 1) generally increase as the number of layers increases, exceeding the average of 130 m/d obtained from the Cooper-Jacob analyses for most zonations. The estimated K values for the most permeable layers tend to be much larger than the K values for the direct-push slug tests, but are relatively consistent with the magnitude of K values from the GEMSTRAC1 tracer test, at least for the five- and seven-layer zonations. In particular, the seven-layer analyses yield a good match to the peak conductivity value in the GEMSTRAC1 profile, although not to the overall character of that profile.

[36] Table 2 shows the run times for the steady shape and transient analyses of all 23 tests, for both the regular-layered and radar-based zonations. Direct comparison of the overall run times is complicated by the fact that different inverse runs will require differing numbers of parameter iterations

to reach convergence and therefore differing numbers of forward simulations. Here a single forward simulation means simulation of all 23 tests with a single parameter vector. In Table 2, the results are resolved to an average run time per forward simulation in each analysis. Disregarding the results for the single-layer (homogeneous) fits, for which a significant fraction of the overall run time consists of the overhead of reading and writing data, the results stabilize to about 0.18 min per forward simulation for the steady shape approach and 7.14 min per forward simulation for the transient approach, meaning the transient forward simulation takes about 40 times longer than the steady state simulation used in the steady shape approach. Clearly, the time savings rendered by the steady shape approach will depend on simulation details, and particularly the time discretization that would be used in the transient simulations. We have used 118 time steps in our transient simulations, with an initial time step of 1×10^{-4} s and a time step acceleration factor of 1.1, reaching out to a simulation time just past our final measurement time of 70 s. (Small initial time steps are required to get an accurate representation of the entire drawdown curve.) Regardless of the details, the steady shape approach will clearly have a significant computational advantage over conventional time stepping transient simulation in most situations.

[37] The steady shape and transient approaches tend to produce fairly similar results overall, implying that the steady shape analysis is capturing most of the information available in the transient data. More accurately, since we are analyzing the same drawdown observations in both approaches, the steady shape “view” of the data, considering only drawdown differences between different locations at common observation times, is yielding very similar results as the transient view of the data. We speculated earlier that the steady shape approach might be less inclined to produce a K profile whose average matches the large-scale K of the aquifer, since the steady shape analysis is not constrained to reproduce the drawdown versus time rela-

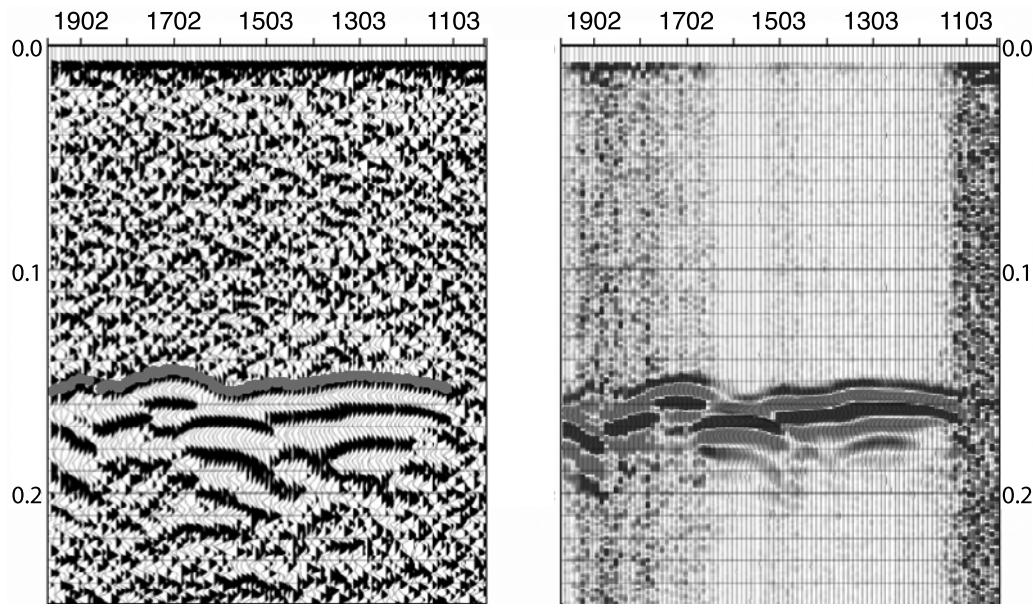


Figure 10. Zero-offset radar profile between Gems4S and Gems4N, shown as (left) variable-gain wiggle traces and (right) gray scale-coded amplitudes with no gain. Horizontal axis is depth below a datum at ground surface in centimeters, increasing to the left, and vertical axis is travel time in microseconds. First-break picks are highlighted on left.

tionship. For these tests, both of the approaches tend to overestimate K values relative to our prior expectations, so it is difficult to judge whether the steady shape approach is any less constrained to reproduce the overall average K than the transient approach. However, the single-layer (homogeneous) transient analysis produces a K estimate of 124.5 m/d (Table 1), fairly close to the estimate of 130 m/d obtained from the Cooper-Jacob analyses and previous pumping tests, while the single-layer steady shape analysis produces a K value somewhat lower than expected, 95.8 m/d.

[38] Note that these analyses employ no regularization terms to constrain the K values to match a priori estimates or smoothness criteria. For example, we could have constrained the analyses to yield K profiles with averages matching our prior estimate for the bulk average K , and this almost certainly would have reduced the tendency to produce such high K estimates for some layers. However, the unconstrained inversion approach used here provides a more independent comparison between the K values estimated from hydraulic tomography and those derived from other testing techniques. These inversions do have some dependence on the prior bulk K estimate through the use of that value (130 m/d) as the initial K value for each layer, but the K estimates are free to range widely from that initial value.

7. Radar-Based Zonations

[39] The underlying assumption of the radar-based zonations used here is that the hydraulic and electromagnetic property distributions in the subsurface are governed to some extent by the same lithologic factors. Unlike a major thrust of some previous work [e.g., Rubin *et al.*, 1992; Coptly *et al.*, 1993; Hubbard *et al.*, 2001; Chen *et al.*, 2004], we do not attempt to use direct correlations between hydrogeological and geophysical properties. Instead, we attempt

to exploit a weaker connection, assuming both sets of properties show some correspondence with lithology. Our work is thus in keeping with previous studies that used geophysical data for the characterization of lithofacies geometry [e.g., Hyndman *et al.*, 1994; Coptly and Rubin, 1995; Eppstein and Dougherty, 1998; Hubbard *et al.*, 1999]. One new element of our work is the investigation of the effectiveness of an automated clustering procedure for guiding zonation using an approach similar to that of Tronicke *et al.* [2004].

7.1. Interpretation of Zero-Offset Radar Profile

[40] Figure 10 shows two different representations of a zero-offset radar profile run between Gems4S and Gems4N, in the same vertical plane as the hydraulic tomography tests. This survey employed 100-MHz antennas, with the transmitting antenna in Gems4N and the receiving antenna in Gems4S. The two antennas were lowered together in 10-cm increments, thus remaining at approximately equal depths in the two wells. Figure 11 shows the propagation velocity and signal attenuation values derived from the first break travel times and peak amplitude variations, respectively. These values are plotted versus meters above datum, together with a set of block averages for different layered zonations of the velocity and attenuation data, with layer boundaries adjusted to the nearest cell boundary in the flow model. The zonal boundaries are the same for both properties, although not all breaks are clearly expressed in both variables.

[41] The 13-layer zonation, at the right in Figure 11, is an expert interpretation of the radar data. In addition to using a zonation based on expert interpretation, we investigated zonations derived with a more automated approach. The 5-, 7-, and 10-layer zonations shown in Figure 11 are derived from hierarchical depth-constrained cluster analysis [Gill *et al.*, 1993; Bohling *et al.*, 1998] of the velocity and attenuation values. Hierarchical depth-constrained cluster

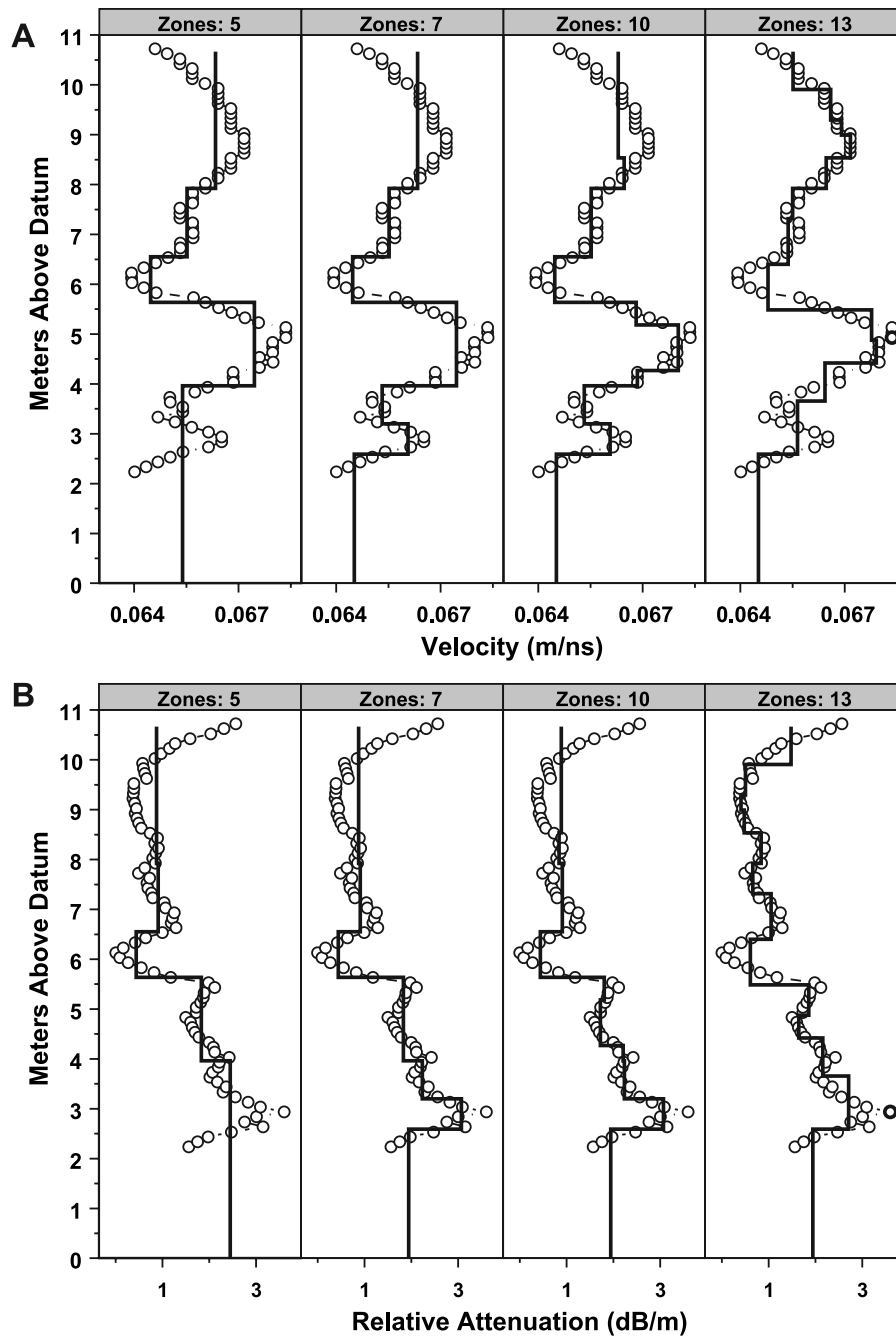


Figure 11. (a) Radar velocity. (b) Attenuation values from zero-offset survey with clustering-based zonations (5–10 zones) and 13-zone expert interpretation. Zone boundaries are the same for both properties.

analysis is simply an application of *Ward's* [1963] multi-variate clustering algorithm subject to the constraint that only vertically adjacent objects may be joined. In this case the algorithm begins by considering each measurement point (depth) in the velocity-attenuation profiles as a separate object and then joins the two most similar vertically adjacent objects, where similarity is measured in terms of the Euclidean distance between the points in velocity-attenuation space (variables are standardized to zero mean and unit standard deviation to equalize their influence). The object, or zone, created by the merger replaces the original two measurements in the series and the process repeats until

all measurements are joined into a single cluster or zone. Each merger creates the least possible increase in the overall within-zone variance of the velocity and attenuation values, meaning the process attempts to keep the zones as homogeneous as possible at each step.

[42] The 5-, 7-, and 10-layer zonations were selected based on an examination of the relative increases in within-zone variance at each merger. Reasoning that large relative increases indicate the merger of fairly dissimilar zones, we selected the zonations immediately preceding such increases as somewhat natural divisions of the sequence. The zonations shown in Figure 11 do not exactly

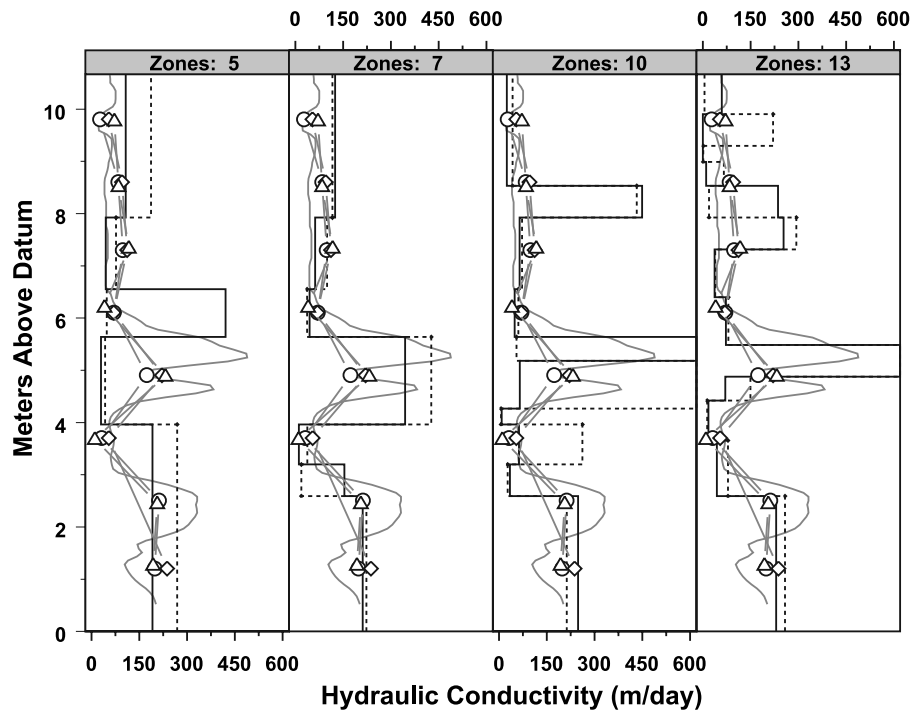


Figure 12. Estimated hydraulic conductivities from steady shape hydraulic tomography (solid curves) and transient hydraulic tomography (dashed curves) using radar-based zonations compared with K profiles from the direct-push slug tests (HP1, triangles; HP8, circles; DP808, diamonds) and the GEMSTRAC1 tracer test (shaded curve).

correspond to the zonations selected on the basis of within-zone variance due to the adjustment of the zone boundaries to the flow model grid and the elimination of two thin zones, including a 0.3-m-thick zone near the top of the aquifer. Thus the cluster analysis does not provide absolute objectivity in interpreting the radar profiles; it simply serves as a tool to aid in that interpretation. As shown in Figure 11, the resulting zonations provide fairly reasonable representations of the radar velocity and attenuation profiles, at least up to about 8.5 m above datum. Note that due to the hierarchical nature of the clustering, the 5-, 7-, and 10-layer zonations are nested. The number of flow model cells per layer is variable in this case. For the 10-layer zonation, the number of cells per layer ranges from two (0.3 m) to 17 (2.6 m).

7.2. Analysis Results

[43] Figure 12 compares the estimated K profiles obtained from the tomography analyses using the radar-based zonations with the slug-test and GEMSTRAC1 K profiles. As for the regular-layered results, none of the radar-based layering models shows significant signs of overparameterization as judged by the Jacobian matrix condition numbers or estimated parameter correlations (Table 1), except perhaps the 13-layer steady shape fit, for which three correlation values exceed 0.9 in magnitude but which, again, does not exhibit a particularly troubling reciprocal condition number (9×10^{-7}). Table 1 contains RMS residuals and thickness-weighted average K values for each case. These RMS residuals are plotted in Figure 9, together with the results for the regular-layered zonations. The RMS residuals are, in general, slightly better for the

regular-layered zonations than for the radar-based zonations, but do not provide strong evidence for choosing one zonation style over the other. Both approaches result in similar overestimates of average K relative to our prior expectations. Comparing the K profiles in Figure 12 with those in Figure 8, however, shows some evidence that the radar-based zonations provide a better match to the K field structure than do the regular-layered zonations. In both cases, the five-zone results appear to be too coarse to produce a good match to the slug-test and GEMSTRAC1 K profiles, while the results for 10 or more zones seem to exhibit instability with anomalously high K estimates for some layers.

[44] The seven-layer results appear to produce the most satisfactory estimates, in terms of the consistency between transient and steady shape approaches and their agreement with the slug-test and GEMSTRAC1 K profiles. Comparing the 10-layer results with the seven-layer results, and the 13-layer results with the 10-layer results, one can see a significant increase in the discrepancy between steady shape and transient results as certain portions of the aquifer are divided into more zones, with the K estimate for the coarser zonation tending to be replaced by an oscillating high- and low-K pattern in the finer zonation. This behavior probably indicates that the more detailed representations are in fact overparameterized in these portions of the aquifer. In addition, the simulated quenching results discussed in section 7.3 indicate a much higher level of variability in the 10-layer K estimates than in the seven-layer K estimates, giving further evidence that the seven-layer zonation is perhaps the most optimal of those shown. For these seven-layer results, the radar-based layering produces

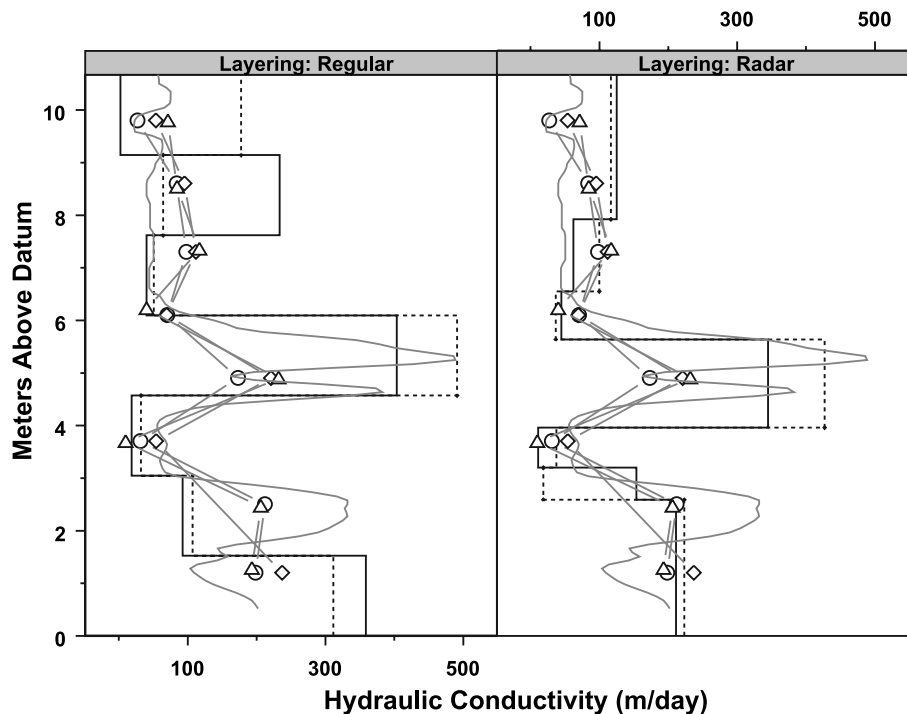


Figure 13. Comparison of K profiles from regular and radar-based seven-layer hydraulic tomography analysis with the direct-push slug tests and the GEMSTRAC1 tracer test (symbols as in Figure 12).

notably better matches to the slug-test K profiles and, overall, a better match to the GEMSTRAC1 K profile. Also, the plots of RMS residual versus number of zones (Figure 9) for the steady shape, radar-based analyses appear to show a break at seven layers, perhaps indicating that this zonation has some correspondence with the true K field structure. In the next section we examine the seven-layer analyses in more detail.

7.3. Detailed Examination of Seven-Layer Analyses

[45] Figure 13 shows a side-by-side comparison of the regular-layered and radar-based seven-zone results for the steady shape and transient analyses of all 23 tomographic pumping tests, with both the slug-test and GEMSTRAC1 K profiles included for comparison. For the regular-layered zonation, the steady shape analysis required about 17 min of processing time and the transient analysis required about 732 min (12.2 hours), both on a 3.4-GHz PC. For the radar-based zonation, the steady shape analysis ran for 23 min and the transient analysis ran for 544 min (9.1 hours) on the same 3.4-GHz PC. The run time per forward transient simulation of all 23 tests is 7.06 min in this case, while the run time per forward steady state simulation (employed in the steady shape analysis) is 0.18 min (Table 2). Thus the most striking aspect of the results displayed in Figure 13 is the fact that the steady shape analyses have achieved results very similar to the transient analyses in roughly 2–4% of the CPU time. The ramifications of this computational advantage are discussed further shortly.

[46] Figure 14 shows the observed and predicted drawdowns for the transient analyses, and the observed and predicted drawdown differences for the steady shape analyses for the two zonation strategies. The correlations between observed and predicted values are very high in

all four cases: 0.9062 for the transient analysis with regular layering, 0.9069 for the transient analysis with radar-based layering, 0.9633 for the steady shape analysis with regular layering, and 0.9567 for the steady shape analysis with radar-based layering. Distinct streaks of data points on the cross plots for the transient analyses are associated with particular observation points during particular tests. That is, the discrepancies between observed and predicted drawdowns are dominated by consistent over- and under-predictions at specific observation locations during each test. Similarly, many of the apparent single points on the cross plots for the steady shape analyses are, in fact, clusters of points corresponding to consistent over- or under-prediction for the 26 repeat measurements of a particular drawdown difference. These systematic discrepancies are consistent with the highly significant lack-of-fit statistics mentioned earlier. Clearly, the perfectly layered configurations cannot reproduce the hydraulic conductivity variations at GEMS in sufficient detail to remove the systematic deviations between the observed and predicted values. Despite that, however, analyses with these simplified representations can yield some insight into the viability of the tomography procedure and the relative advantages of the different zonation strategies.

[47] Judged relative to the GEMSTRAC1 K profile, the regular-layered zonation appears to provide a reasonable fit in the middle of the aquifer, particularly in terms of characterizing the highest K zone in that profile (Figure 13). The radar-based zonation produces a better match to the GEMSTRAC1 profile in the upper and lower portions of the aquifer. Relative to the slug-test profiles, the radar-based zonation clearly provides a better match. In fact, the match is quite striking apart from the discrepancies in the estimates

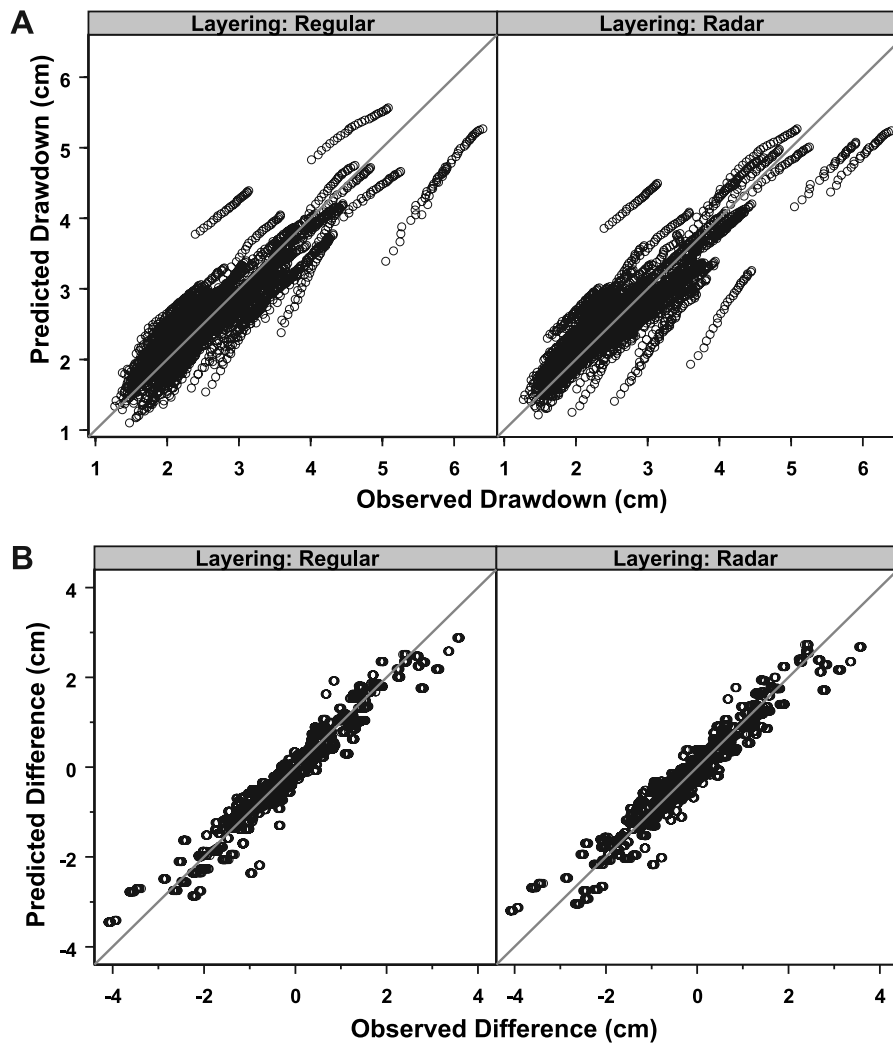


Figure 14. (a) Observed and predicted drawdowns from transient analysis. (b) Observed and predicted drawdown differences for steady shape analysis using regular and radar-based seven-layer zonations.

for the highest K zone. Moreover, the magnitude of the K estimates from both the transient and steady shape tomography analyses using the radar-based zonation are not inconsistent with those obtained from the GEMSTRAC1 analysis.

[48] Table 3 contains the K estimates and their coefficients of variation for each layer for all four analyses (steady

shape and transient, regular and radar-based layering). The coefficient of variation is the standard deviation of the K estimate divided by the estimate itself, with the standard deviation computed as the square root of the corresponding diagonal element of the estimated parameter covariance matrix (a linearized estimate about the final vector of

Table 3. Hydraulic Conductivity Estimates and Coefficients of Variation for Seven-Zone Analyses of All 23 Tomographic Pumping Tests, Using Seven Equal-Thickness Zones and Seven Radar-Based Zones With Drawdown Data Analyzed in Both Steady Shape and Transient Modes^a

Zone	Regular, Steady Shape		Regular, Transient		Radar, Steady Shape		Radar, Transient	
	K, m/d	CV, %	K, m/d	CV, %	K, m/d	CV, %	K, m/d	CV, %
7	2.2	9.88	177.5	4.36	125.1	1.86	116.7	3.61
6	233.6	1.60	64.1	4.64	62.2	2.48	99.8	4.70
5	40.2	1.54	50.8	3.11	44.7	2.42	36.6	3.73
4	403.4	1.81	491.0	3.15	344.7	1.88	427.0	2.83
3	18.5	1.71	32.2	2.46	11.2	1.85	37.4	4.08
2	92.7	6.25	107.2	10.13	153.8	10.54	18.6	4.98
1	358.8	2.35	311.5	3.92	211.3	2.51	223.2	2.29

^aZone (layer) 7 is at the top of the aquifer and zone 1 is at the bottom. Zone boundaries differ between regular and radar-based layerings and thus are not strictly comparable. CV indicates standard deviation of estimate divided by estimate.

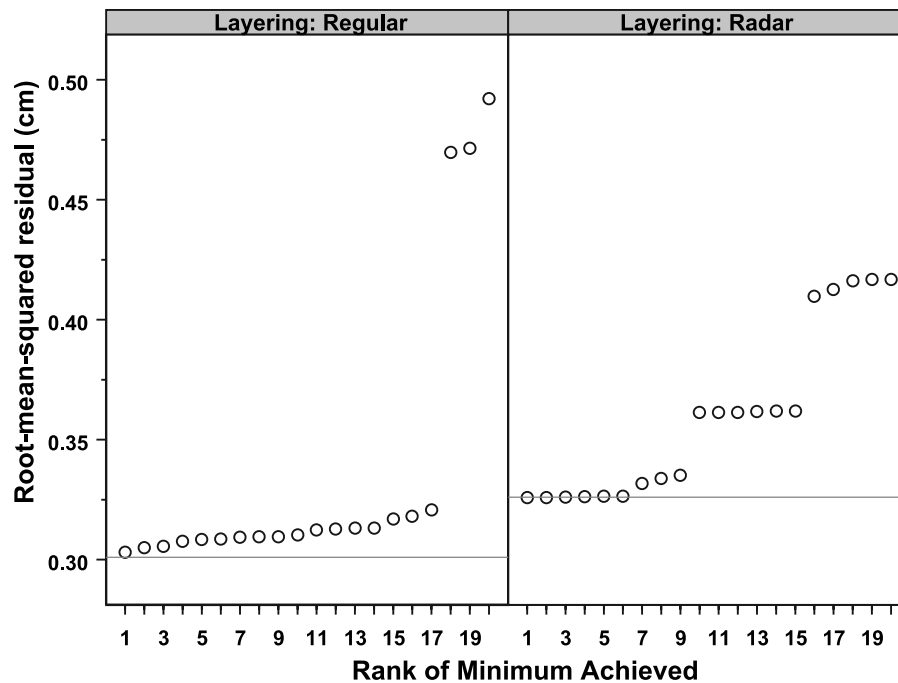


Figure 15. Minimum RMS residual values achieved by each trial of simulated quenching analysis, in order from smallest to largest for each layering style. Horizontal lines are minima achieved by gradient-based inversion.

estimated K values). Overall, the uncertainty estimates are fairly comparable; neither zonation scheme yields clearly superior (smaller) coefficients of variation.

[49] *Zhu and Yeh* [2005] raise some concerns about the gradient-based optimization method used here. In order to assess the reliability of this method (Levenberg-Marquardt with a finite difference Jacobian) and examine the influence of local minima, we repeated the analysis using a more computationally intensive random-search method. This alternative approach could be called “simulated quenching” [Carle, 1997], as it is basically a zero-temperature version of simulated annealing [Tarantola, 2005]. Each trial of the quenching process begins from a random initial vector of K values, and at each iteration, one of the K values, selected at random, is perturbed by a random ratio, between 75% and 125% of its current value. If the perturbation results in a reduction of the objective function, then it is accepted and the parameter vector is updated. Otherwise, the change is rejected. We have used the steady shape objective function (equation (1)) for these analyses and have run 20 trials, each starting from a different random initial K vector, for each zonation style. Each trial proceeds for either 500 iterations or until 50 consecutive iterations fail to improve the objective function. Thus each analysis could involve up to 10,000 forward solutions, for each of which the steady state flow equation is solved for 23 different tests. Most of the trials ran for the full 500 iterations but produced insignificant reductions in the objective function after approximately 300 iterations. For either zonation style (regular or radar-based), the quenching analysis took about 27 hours on the same 3.4-GHz PC. A similar analysis based on transient simulations would not be feasible on a standard desktop computer, providing yet a further demonstration of the computational advantages of the steady shape approach.

[50] Figure 15 shows the minimum values achieved in each quenching trial for the two zonation strategies. The horizontal shaded line in each panel represents the RMS residual attained by the gradient-based optimization, 0.301 cm for the regular layering and 0.326 cm for the radar-based layering. A number of the trials in each case reached minima very close to that obtained by the Levenberg-Marquardt algorithm, but none reached a better solution, possibly indicating that the Levenberg-Marquardt algorithm has converged on a global minimum. For both zonations, some of the quenching trials ended up in local minima. Although more of the quenching trials for the regular-layered zonation reached minima close to the (apparent) global minimum, the radar-based zonation results are much more clearly divided into clusters associated with the apparent global minimum and two local minima. The first six quenching results (ordered by minimum RMS residual, as shown in Figure 15) for the radar-based zonation reached a minimum value essentially indistinguishable from the gradient-based minimum. In contrast, the minimum residuals for the trials using regular layering rise gradually from the gradient-based minimum.

[51] For each zonation strategy, Figure 16 shows the K profiles associated with the first six quenching results along with the K profile from the Levenberg-Marquardt optimization. The six K profiles yielded by simulated quenching for the radar-based zonation are almost indistinguishable from that reached by the Levenberg-Marquardt algorithm, while there is a significant amount of variation between profiles for the regular-layered zonation. Note that the seventh smallest minimum for the radar-based zonation, that associated with the first departure from the “global minimum” line in Figure 15, is notably different from the first six shown here. Thus the radar-based zonation appears

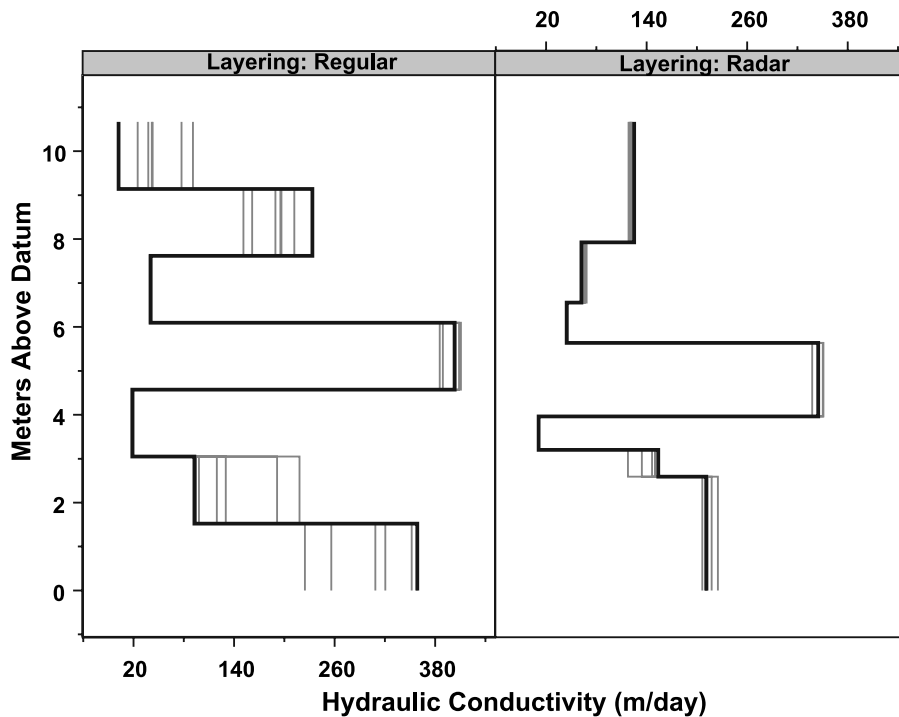


Figure 16. K profiles from steady shape gradient-based optimization for seven-layer zonation (solid curve) and K profiles for the six smallest minima achieved over 20 simulated quenching trials (shaded curves) for each layering style.

to provide a more clearly defined objective function surface for this optimization problem.

[52] Simulated quenching analyses for the 10- and 13- or 14-layer zonations yield similar results, in that the quenching routine is not able to improve on the fit obtained by the Levenberg-Marquardt algorithm, indicating that the latter is quite possibly finding a global minimum. Not surprisingly, the variability in the final K profiles and range of final objective function values for the different simulated quenching trials shows a greater level of variability for these fits than for the seven-layer zonations, which could perhaps be taken as indication of overparameterization for these zonations.

[53] In all cases, we assume a perfectly layered aquifer. Obviously, this conceptualization is a highly simplified representation of reality. Some insight into the appropriateness of this conceptualization can be gained by comparison of the K profiles resulting from the simultaneous analysis of all 23 pumping tests with the profiles resulting from analysis of the “individual-well” results, i.e., analyses only of the pumping tests at Gems4S (12 tests) or only the tests at Gems4N (11 tests). For the sake of brevity, we do not present detailed comparisons between the analyses of different combinations of tests in this paper. In summary, the K profiles from analyses of only the 12 tomographic tests at Gems4S and of all 23 tests at Gems4S and Gems4N yield fairly consistent results. The analysis of the 11 tests at Gems4N alone, however, yields somewhat different results, which are also less consistent with the existing K estimates. This is almost certainly a result of significant lateral variations that violate our assumption of perfect layering. In fact, multilevel slug tests in Gems4S and Gems4N [Butler, 2005] indicate lower K values in the upper portion of the aquifer at Gems4N, compared with those at Gems4S

and the three direct-push slug test sites discussed in this paper. We have also performed a tomographic radar survey between Gems4S and Gems4N, and the resulting tomogram shows some evidence of lateral variations between the two wells, particularly a decrease in porosity from south to north near the top of the aquifer (M. Knoll, unpublished data, 2003). Although the hydraulic tomography results do not show a tendency to produce relatively lower K estimates near the top of the aquifer for the analysis of the Gems4N tests, neglected lateral variations almost certainly contribute to the discrepancies in results for the different combinations of tests. The layer K estimates must be considered “apparent” or “effective” parameters, which are dependent on the imposed geometry. However, this is true of essentially all K estimates, since K is actually a phenomenological parameter and not an inherent property of the medium [Beckie, 1996]. Nevertheless, a layered zonation appears to be a reasonable “first-order” approximation for the K field structure and thus provides a reasonable framework for the investigations of the issues discussed in this paper.

8. Further Verification of Hydraulic Tomography Results

[54] A large number of small-scale hydraulic tests were performed at GEMS prior to the hydraulic tomography experiments reported here, but none of those tests revealed the presence of the very high K layers indicated by our analysis of the GEMSTRAC1 tracer test (Figure 1), causing us initially to discount the K values obtained for those layers as overestimates due to logistical problems with the tracer test. The hydraulic tomography results, however, appeared to corroborate the tracer-test analysis, giving us more

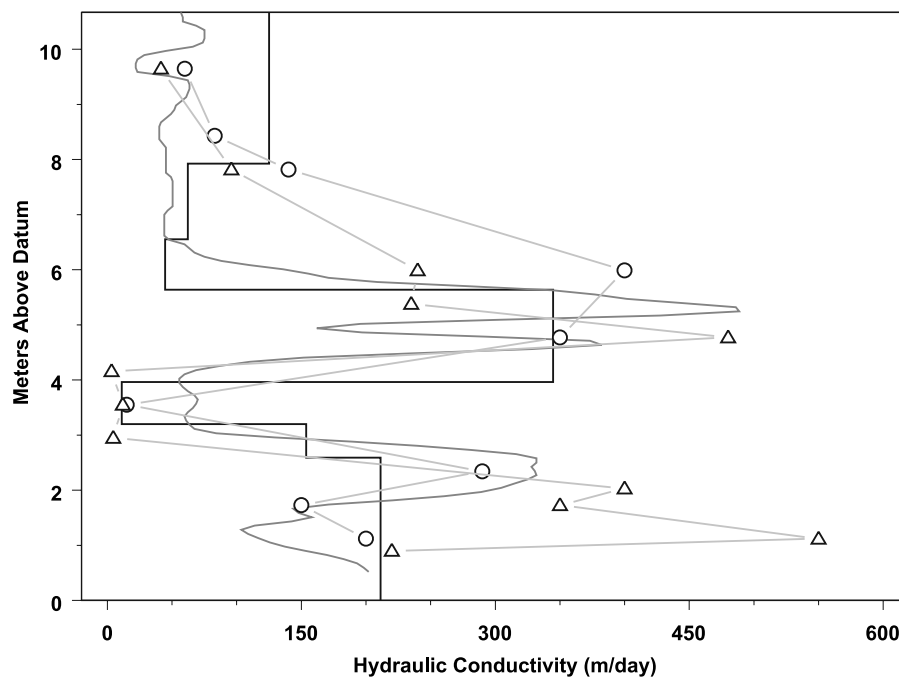


Figure 17. Comparison of K profile from steady shape hydraulic tomography using radar-based zonation (solid curve) with K profiles from the GEMSTRAC1 tracer test (shaded curve) and the direct-push permeameter tests (CP1029a, triangles; CP1029b, circles).

confidence in the existence of the high-K layers. Potential reasons for our inability to detect the high-K layers with small-scale hydraulic tests could be lateral heterogeneity (i.e., the most permeable layers were nonexistent at the few wells for which we had near-continuous K information) and the lack of an approach for getting near-continuous K information away from existing wells.

[55] A promising approach for obtaining near-continuous profiles of hydraulic conductivity in the absence of wells is the direct-push permeameter [Butler and Dietrich, 2004; Butler, 2005; Butler et al., 2007]. Two direct-push permeameter (DPP) K profiles were obtained just to the southwest of the Gems4S-Gems4N plane (locations CP1029a and CP1029b in Figure 2) after completion of the tomography experiments. Although initially we questioned the DPP K estimates due to the low signal-to-noise ratios in the transducer readings [Butler, 2005], follow-up work on the DPP methodology at a heavily studied field site in Germany indicated that the DPP K values should be representative of in situ conditions [Butler and Dietrich, 2004; Butler et al., 2007]. Figure 17 shows that the GEMS DPP K estimates are in reasonable agreement with those from GEMSTRAC1 and from the tomographic pumping tests, providing an important further confirmation of the existence of the apparent high-K zones. This comparison also demonstrates the limitations of point-based measurements for detection of relatively thin zones of extreme properties, as only two of the K profiles based on the many small-scale hydraulic tests performed at GEMS detected the existence of the high-K pathways that played such a critical role in the tracer movement. Clearly, there is a pressing need for a method that can provide information about conditions between wells at a high enough resolution for transport applications. The field comparisons presented here indicate that hydraulic

tomography appears to be able to fulfill that need. However, even with a very high data density, hydraulic tests alone cannot be expected to yield a definitive representation of subsurface conditions; they must always be interpreted in conjunction with additional independent information, such as that provided by geophysical surveys and conceptual geological models.

9. Conclusions

[56] Previous simulation and laboratory studies have found that hydraulic tomography has considerable potential for providing information about spatial variations in hydraulic conductivity in saturated porous media at a scale of relevance for contaminant transport investigations. In this paper, we followed up that promising initial work by presenting the results of the first field assessment of the approach in a sand and gravel aquifer, a setting similar to those commonly faced in practical field investigations. As part of the field assessment, we examined three critical issues regarding hydraulic tomography: zonation strategies, analysis methodology, and viability of resulting parameter estimates. The primary conclusions of our study are discussed in the following sections.

9.1. Zonation Strategies

[57] A major focus of our zonation investigation was to assess whether information on electromagnetic property variations obtained from cross-hole radar surveys can help to guide the characterization of the K field geometry, on the premise that underlying lithological variations are a significant control on both the hydraulic and electromagnetic property variations. K estimates for layered zonations derived from clustering of radar velocity and attenuation profiles exhibit a better overall match to previous K

estimates at the site than do estimates from arbitrary evenly spaced zonations, providing some indication of the potential utility of radar surveys for aquifer characterization.

[58] The results presented here were based on a simplified representation of a heterogeneous aquifer as a system of homogenous layers extending the full width of the domain of our investigation. Although the assumption of perfect layering does not appear to have been particularly unreasonable for the system we investigated here, it clearly would not be appropriate in many systems. A possibly more effective zonation strategy to consider for future investigations would involve utilizing the computationally efficient hydraulic diffusivity tomography approach of *Brauchler et al.* [2003] and cluster analysis to construct zones of constant diffusivity. The steady shape analysis would then determine the hydraulic conductivity for each zone, after which the specific storage of each zone could be calculated from the hydraulic conductivity and diffusivity estimates.

9.2. Analysis Methodology

[59] Our primary focus was on the steady shape analysis method of *Bohling et al.* [2002]. In order to develop a better understanding of the relative advantages of the different analysis methodologies, we also analyzed the tests in a fully transient mode, adjusting hydraulic conductivities to match the observed drawdowns versus time. Overall, the K profiles estimated using the steady shape approach are quite similar to those estimated using the transient analysis. The run times for the steady shape analyses are typically a few percent of those for the fully transient analyses. Thus the steady shape approach appears to extract the vital information from the test data in a fraction of the run time of the transient approach. This study, to the best of our knowledge, is the first demonstration of the practical utility of the steady shape approach using actual field data.

[60] For both the steady shape and transient approaches, the parameter estimates have been obtained using a traditional gradient-based optimization algorithm (specifically, Levenberg-Marquardt with a finite difference Jacobian estimate), starting from an initial K estimate of 130 m/d for every zone. This initial value is close to the bulk average K determined from previous pumping tests at the site. For the steady shape approach, we also used a more brute force random-search algorithm to address previously raised concerns about the gradient-based optimization algorithm [*Zhu and Yeh*, 2005] and to explore the uncertainty and non-uniqueness in parameter estimates in more detail. This exercise demonstrated one of the great strengths of the steady shape approach: The efficiency of the steady state forward simulations allows the use of more computationally intensive probabilistic parameter estimation techniques, and thus a better characterization of the uncertainty in the final parameter estimates.

[61] We have used a fairly simple alternating direction implicit scheme for the solution of the two-dimensional radial-vertical flow model employed here. While it is possible that more efficient solution methods would reduce the differences in run time between the transient and steady shape approaches, there would almost certainly still be a significant computational advantage for the steady shape approach because of its use of steady state simulations. It is important to note again the two major advantages of the steady shape approach over a conventional steady state

method: (1) A steady shape approach does not require attainment of true steady state conditions, and (2) it is less sensitive to the specification of outer boundary conditions, which are often poorly known. It should also be noted that the steady shape approach focuses only on the formulation of the objective function for inverse analysis and thus could be incorporated into any number of approaches for estimating the K distribution, including the sequential inversion of tomographic pumping tests advocated by *Zhu and Yeh* [2005].

9.3. Field Comparison

[62] We performed a series of experiments in an extensively studied sand and gravel aquifer to assess the value of the information obtained using the hydraulic tomography procedure. The results of those experiments showed that the hydraulic tomography K profiles were in reasonable agreement with those obtained from an induced-gradient tracer test and direct-push permeameter tests. The vast majority of small-scale hydraulic tests previously performed at the site, however, failed to reveal the presence of the high-K layers detected by these three approaches. The knowledge of these preferential pathways would be critical for transport investigations. The most likely reasons for our previous inability to detect these critically important features are (1) the layers are nonexistent at the few wells for which we have near-continuous K information, and (2) we did not have reliable methods for obtaining near-continuous K information away from those wells. This field comparison clearly demonstrates the value of the information that can be gained from the hydraulic tomography approach. Given the logistical challenges and costs (in terms of time, labor, and money) associated with the performance of multiwell tracer tests, hydraulic tomography appears to be an exciting and practically feasible alternative for obtaining information about spatial variations in hydraulic conductivity on the scale of relevance for contaminant transport applications.

9.4. Limitations

[63] One major challenge to effective implementation of hydraulic tomography is our limited ability to induce measurable vertical differences in pressure over significant distances, especially in high-permeability media. This study involved a small region of investigation, with only a few to several meters separating pumping and observation wells. Contaminant transport problems, on the other hand, can easily involve distances of the order of hundreds of meters or more. In high-permeability media, such as the alluvial aquifer at GEMS, it would be nearly impossible to measure pumping-induced drawdowns to an adequate degree of accuracy over more than a few tens of meters, due to sensor limitations, interference from other sources, and, more fundamentally, the highly dissipative nature of groundwater flow.

[64] A specific limitation of the present study is our use of a two-dimensional flow model for an inherently three-dimensional problem. However, the flow model correctly represents the physics of two-dimensional radial-vertical flow and allowed us to accurately represent flow in the vicinity of the well, including damping of vertical gradients due to bypass flow along the well bore. The model will accurately represent flow to a well in the absence of significant angular variations in boundary conditions or aquifer properties. Our primary justification for using such

a model in a clearly heterogeneous aquifer is that there is a relatively large degree of lateral continuity at our field site (Figures 1, 13, and 17), so the magnitude of angular variations in K should not be great. Although a truly three-dimensional model would clearly allow a more accurate representation of the aquifer heterogeneity, we would not expect a large impact on the results reported here.

[65] **Acknowledgments.** This research was supported by the Hydrologic Sciences Program of the National Science Foundation under grant 9903103 and the Kansas Geological Survey Applied Geohydrology Summer Research Assistantship Program. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF. Field assistance provided by John Healey of the Kansas Geological Survey and Greg Davis and Sam Cain, participants in the Applied Geohydrology Summer Research Assistantship Program, is gratefully acknowledged. This manuscript greatly benefited from comments by the three reviewers.

References

- Aster, R. C., B. Borchers, and C. H. Thurber (2005), *Parameter Estimation and Inverse Problems*, 301 pp., Elsevier, New York.
- Beckie, R. (1996), Measurement scale, network sampling scale, and groundwater model parameters, *Water Resour. Res.*, 32(1), 65–76.
- Bohling, G. C. (1993), Hydraulic tomography in two-dimensional, steady-state groundwater flow, *Eos Trans. AGU*, 74, 141.
- Bohling, G. C. (1999), Evaluation of an induced gradient tracer test in an alluvial aquifer, Ph.D. dissertation, 224 pp., Univ. of Kansas, Lawrence.
- Bohling, G. C., and J. J. Butler, Jr. (2001), Lr2dinv: A finite-difference model for inverse analysis of two-dimensional linear or radial groundwater flow, *Comput. Geosci.*, 27(10), 1147–1156.
- Bohling, G. C., J. H. Doveton, B. Guy, W. L. Watney, and S. Bhattacharya (1998), PteFFER 2.0 Manual, Kans. Geol. Surv., Lawrence.
- Bohling, G. C., X. Zhan, J. J. Butler, Jr., and L. Zheng (2002), Steady shape analysis of tomographic pumping tests for characterization of aquifer heterogeneities, *Water Resour. Res.*, 38(12), 1324, doi:10.1029/2001WR001176.
- Brauchler, R., R. Liedl, and P. Dietrich (2003), A travel time based hydraulic tomographic approach, *Water Resour. Res.*, 39(12), 1370, doi:10.1029/2003WR002262.
- Butler, J. J., Jr. (1988), Pumping tests in nonuniform aquifers—The radially symmetric case, *J. Hydrol.*, 101(1–4), 15–30.
- Butler, J. J., Jr. (1990), The role of pumping tests in site characterization: Some theoretical considerations, *Ground Water*, 28(3), 394–402.
- Butler, J. J., Jr. (2005), Hydrogeological methods for estimation of hydraulic conductivity, in *Hydrogeophysics*, edited by Y. Rubin and S. Hubbard, pp. 23–58, Springer, New York.
- Butler, J. J., Jr., and P. Dietrich (2004), New methods for high-resolution characterization of spatial variations in hydraulic conductivity, in *Proceedings of the International Symposium on Hydrogeological Investigation and Remedial Technology*, pp. 42–45, Natl. Central Univ., Jhongli, Taiwan.
- Butler, J. J., Jr., and X. Zhan (2004), Hydraulic tests in highly permeable aquifers, *Water Resour. Res.*, 40, W12402, doi:10.1029/2003WR002998.
- Butler, J. J., Jr., J. M. Healey, L. Zheng, W. McCall, and M. K. Schulmeister (1999a), Hydrostratigraphic characterization of unconsolidated alluvial deposits with direct-push sensor technology, *Kans. Geol. Surv. Open File Rep.*, 99–40, 30 pp. (Available at www.kgs.ku.edu/Hydro/Publications/OFR99_40/index.html)
- Butler, J. J., Jr., C. D. McElwee, and G. C. Bohling (1999b), Pumping tests in networks of multilevel sampling wells: Motivation and methodology, *Water Resour. Res.*, 35(11), 3553–3560.
- Butler, J. J., Jr., J. M. Healey, G. W. McCall, E. J. Garnett, and S. P. Loheide II (2002), Hydraulic tests with direct-push equipment, *Ground Water*, 40(1), 25–36.
- Butler, J. J., Jr., P. Dietrich, V. Wittig, and T. Christy (2007), The direct-push permeameter for characterizing spatial variations in hydraulic conductivity, *Ground Water*, in press.
- Carle, S. F. (1997), Implementation schemes for avoiding artifact discontinuities in simulated annealing, *Math. Geol.*, 29(2), 231–244.
- Carrera, J., and S. P. Neuman (1986), Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, *Water Resour. Res.*, 22(2), 199–210.
- Chen, J., S. Hubbard, Y. Rubin, C. Murray, E. Roden, and E. Majer (2004), Geochemical characterization using geophysical data and Markov Chain Monte Carlo methods: A case study at the South Oyster bacterial transport site in Virginia, *Water Resour. Res.*, 40, W12412, doi:10.1029/2003WR002883.
- Cooper, H. H., Jr., and C. E. Jacob (1946), A generalized graphical method for evaluating formation constants and summarizing well field history, *Eos Trans. AGU*, 27, 526–534.
- Coptly, N., and Y. Rubin (1995), A stochastic approach to the characterization of lithofacies from surface seismic and well data, *Water Resour. Res.*, 31(7), 1673–1686.
- Coptly, N., Y. Rubin, and G. Mavko (1993), Geophysical-hydrological identification of field permeabilities through Bayesian updating, *Water Resour. Res.*, 29(8), 2813–2825.
- Einarson, M. D., and J. A. Cherry (2002), A new multilevel ground water monitoring system using multichannel tubing, *Ground Water Monit. Rem.*, 22(4), 52–65.
- Eppstein, M. J., and D. E. Dougherty (1998), Efficient three-dimensional data inversion: Soil characterization and moisture monitoring from cross-well ground-penetrating radar at a Vermont test site, *Water Resour. Res.*, 34(8), 1889–1900.
- Gill, D., A. Shomrony, and H. Fligelman (1993), Numerical zonation of log suites and logfacies recognition by multivariate clustering, *AAPG Bull.*, 77(10), 1781–1791.
- Gottlieb, J., and P. Dietrich (1995), Identification of the permeability distribution in soil by hydraulic tomography, *Inverse Probl.*, 11, 353–360.
- Hastie, T., R. Tibshirani, and J. Friedman (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 533 pp., Springer, New York.
- Hubbard, S. S., Y. Rubin, and E. Majer (1999), Spatial correlation structure estimation using geophysical and hydrogeological data, *Water Resour. Res.*, 35(6), 1809–1825.
- Hubbard, S. S., J. Chen, J. Peterson, E. L. Majer, K. H. Williams, D. J. Swift, B. Mailloux, and Y. Rubin (2001), Hydrogeological characterization of the South Oyster bacterial transport site using geophysical data, *Water Resour. Res.*, 37(10), 2431–2456.
- Hyndman, D. W., J. M. Harris, and S. M. Gorelick (1994), Coupled seismic and tracer test inversion for aquifer property characterization, *Water Resour. Res.*, 30(7), 1965–1977.
- Illman, W. A., and S. P. Neuman (2003), Steady-state analyses of cross-hole pneumatic injection tests in unsaturated fractured tuff, *J. Hydrol.*, 281, 36–54.
- Kruseman, G. P., and N. A. de Ridder (1990), *Analysis and Evaluation of Pumping Test Data*, ILRI Publ. 47, 377 pp., Int. Inst. for Land Reclam. and Impr., Wageningen, Netherlands.
- Liu, S., T.-C. J. Yeh, and R. Gardiner (2002), Effectiveness of hydraulic tomography: Sandbox experiments, *Water Resour. Res.*, 38(4), 1034, doi:10.1029/2001WR000338.
- McCall, W., J. J. Butler Jr., J. M. Healey, A. A. Lanier, S. M. Sellwood, and E. J. Garnett (2002), A dual-tube direct-push method for vertical profiling of hydraulic conductivity in unconsolidated formations, *Environ. Eng. Geosci.*, 8(2), 75–84.
- Moench, A. F. (1989), Convergent radial dispersion: A Laplace transform solution for aquifer tracer testing, *Water Resour. Res.*, 25(3), 439–447.
- Neuman, S. P. (1987), Stochastic continuum representation of fractured rock permeability as an alternative to the REV and fracture network concepts, in *Proc. U. S. Rock Mech. Symp.*, 28, 533–561.
- Parker, R. L. (1994), *Geophysical Inverse Theory*, 386 pp., Princeton Univ. Press, Princeton, N. J.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992), *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed., 963 pp., Cambridge Univ. Press, New York.
- Rubin, Y., G. Mavko, and J. Harris (1992), Mapping permeability in heterogeneous aquifers using hydrologic and seismic data, *Water Resour. Res.*, 28(7), 1809–1816.
- Sánchez-Vila, X., P. M. Meier, and J. Carrera (1999), Pumping tests in heterogeneous aquifers: An analytical study of what can be obtained from their interpretation using Jacob's method, *Water Resour. Res.*, 35(4), 943–952.
- Schad, H., and G. Teutsch (1994), Effects of the investigation scale on pumping test results in heterogeneous porous aquifers, *J. Hydrol.*, 159, 61–77.
- Schulmeister, M. K., J. J. Butler Jr., J. M. Healey, L. Zheng, D. A. Wysocki, and G. W. McCall (2003), Direct-push electrical conductivity logging for high-resolution hydrostratigraphic characterization, *Ground Water Monit. Rem.*, 23(3), 52–62.

- Sellwood, S. M., J. M. Healey, S. Birk, and J. J. Butler Jr. (2005), Direct-push hydrostratigraphic profiling: Coupling electrical logging and slug tests, *Ground Water*, 43(1), 19–29.
- Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, 342 pp., Soc. for Ind. and Appl. Math., Philadelphia, Pa.
- Tosaka, H., K. Masumoto, and K. Kojima (1993), Hydropulse tomography for identifying 3-D permeability distribution, in *High Level Radioactive Waste Management: Proceedings of the Fourth Annual International Conference of the ASCE*, pp. 955–959, Am. Soc. Civ. Eng., Reston, Va.
- Tronicke, J., K. Holliger, W. Barrash, and M. D. Knoll (2004), Multivariate analysis of cross-hole georadar velocity and attenuation tomograms for aquifer zonation, *Water Resour. Res.*, 40, W01519, doi:10.1029/2003WR002031.
- Vesselinov, V. V., S. P. Neuman, and W. A. Illman (2001a), Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff: 1. Methodology and borehole effects, *Water Resour. Res.*, 37(12), 3001–3017.
- Vesselinov, V. V., S. P. Neuman, and W. A. Illman (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 Resour. Res.*, 37(12), 3019–3041.
- Ward, J. H. (1963), Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.*, 58, 236–244.
- Yeh, T.-C. J., and S. Liu (2000), Hydraulic tomography: Development of a new aquifer test method, *Water Resour. Res.*, 36(8), 2095–2105.
- Zemansky, G. M., and C. D. McElwee (2005), High-resolution slug testing, *Ground Water*, 43(2), 222–230.
- Zhu, J., and T.-C. J. Yeh (2005), Characterization of aquifer heterogeneity using transient hydraulic tomography, *Water Resour. Res.*, 41, W07028, doi:10.1029/2004WR003790.
-
- G. C. Bohling and J. J. Butler Jr., Kansas Geological Survey, University of Kansas, 1930 Constant Avenue, Lawrence, KS 66057, USA. (geoff@kgs.ku.edu; jbutler@kgs.ku.edu)
- M. D. Knoll, Center for Geophysical Investigation of the Shallow Subsurface, Boise State University, 1910 University Drive, MG206, Boise, ID 83725-1536, USA. (mknoll@cgiss.boisest)
- X. Zhan, MACTEC Engineering and Consulting, Inc., 600 Grand Avenue, Suite 300, Oakland, CA 94610, USA. (xyz316@gmail.com)