Hydraulic Tomography:

Field and Laboratory Studies

by

Steven James Berg

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Earth Sciences

Waterloo, Ontario, Canada, 2011

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Steven J. Berg

Abstract

Accurately characterizing the distribution of hydraulic parameters is critical for any site investigation, particularly those dealing with solute or contaminant transport. Despite the fact that many tools are currently available for both characterizing (e.g. soil core analysis, slug and pumping tests, direct push techniques, etc.,) and modeling (e.g. geostatistical interpolators, construction of geological models, etc.,) heterogeneous aquifers, this still remains a challenge. In this thesis, hydraulic tomography (HT), a recently developed tool for characterizing and modeling heterogeneous aquifers is evaluated under both laboratory and field conditions.

To date, both steady state hydraulic tomography (SSHT) and transient hydraulic tomography (THT) have been demonstrated at the laboratory scale, however, only SSHT has been rigorously validated through the prediction of independent tests (those not used for estimating the distribution of hydraulic parameters), and comparison to other characterization/modeling techniques. Additionally, laboratory and field validations of HT using comparisons other than the prediction of independent pumping tests (e.g. prediction of solute transport) are lacking.

The laboratory studies performed in this thesis address some of these gaps by: i) rigorously validating THT through the prediction of independent pumping tests, and comparison to other characterization techniques; ii) using HT estimated parameter distributions to predict the migration of a conservative tracer in a heterogeneous sandbox aquifer; and, iii) predicting the flow of water to a well in a heterogeneous, unconfined, sandbox aquifer. For all three cases, HT was compared to more traditional characterization/modeling approaches, such as; the calculation of homogeneous effective parameters, kriging of point data, or the creation and calibration of a geological model. For each study the performance of HT was superior to the other characterization methods. These laboratory

experiments demonstrated both the ability of HT to map aquifer heterogeneity, and the critical need for accurately understanding heterogeneity in order to make accurate predictions about a system. In this regard, HT is a powerful tool at the laboratory scale where the forcing functions (i.e., boundary conditions, flow rates, etc.,) are accurately known.

While several field scale HT studies have been reported in the literature, none attempt to validate 3D THT through the prediction of independent pumping tests, or through comparison to known geology. The application of THT at the field scale presents unique challenges not faced in the laboratory setting. For example, boundary conditions are not accurately known and it is not possible to instrument a field site as densely as a sandbox aquifer. In the field studies conducted as part of this thesis, THT was validated by comparing estimated hydraulic parameter fields to known geology (borehole data) and simulating 9 pumping tests that were performed at the site. The THT analysis was able to capture the salient features of the aquifer (the presence of a double aquifer separated by an aquitard), and was able to reasonably reproduce most of the pumping tests. For comparison purposes, a homogeneous model and three additional heterogeneous models were created: i) permeameter estimates of hydraulic conductivity from soil cores were interpolated via kriging; ii) the transition probability/Markov Chain approach was used to interpret material classifications from borehole logs; and iii) a stratigraphic model was created and calibrated to pumping test data. Of these cases, THT and the calibrated stratigraphic model performed best, with THT performing slightly better.

This work indicates that it is possible to interpret multiple pumping tests using hydraulic tomography to estimate the 3D distribution of hydraulic parameters in heterogeneous aquifer systems. Also, since hydraulic tomography does not require the collection and analysis of a large number of point samples, it is likely comparable in cost to other characterization/modeling approaches.

Acknowledgements

This research was supported by the Strategic Environmental Research & Development Program (SERDP), the National Science Foundation (NSF), the Natural Sciences and Engineering Research Council (NSERC), the Canadian Foundation for Innovation (CFI), the Ontario Research Foundation, TD Bank, the Ontario Graduate Scholarship program, and the University of Waterloo.

A project with this many components is not possible without the help and support of many people. I would like to thank my committee: Dr. Bob Gillham, Dr. Dave Rudolph, Dr. Ed Sudicky, and Dr. Jim Yeh, for providing me with constructive feedback and guidance over the years. I especially thank my supervisor, Dr. Walter Illman who has mentored me over the past four years and provided me with the opportunity to gain invaluable experience in groundwater modeling, laboratory experiments, and field work. These experiences will always be with me. Thanks also to my external reviewers, Dr. Neil Thomson and Dr. Brent Sleep, for offering their time to review this thesis.

The many components of this thesis required the expertise and experience of many of the staff and technicians within the Department of Earth and Environmental Sciences at the University of Waterloo. I am grateful to; Rob McLaren, Dr. Young-Jin Park, Paul Johnson, Bob Ingleton, Marianne Vandergriendt, Shirley Chatten, Scott Piggott, and Wayne Noble.

I have been fortunate to have had to opportunity to supervise, and work with, many co-op and summer students over the past four years. I think in some ways I learned more from them than they learned from me. I am sure they will never forget their field work experience, whether it was in the dead of winter or the blistering heat of summer. Without the help of these students this work would have progressed much slower and been less enjoyable; to them I am grateful. I thank: Juzer Beawerwala; Andrew Price; Eric Poon; Aaron Vandenhoff; and William Sellier.

My fellow students have also been an invaluable resource over the years. Our many discussions have been critical to solving problems and pointing me in new directions. I thank; Matt Alexander (the permeameter master), Steve Frey, Marcelo Sousa, Justin Clark, Fatemeh Vakili, Rubaiat Sharmeen, Mike Makahnouk, Colby Steelman, Melissa Bunn, Deqiang Mao, Junfeng Zhu, Michelle Fraser, Jason Cole, Eric Sykes, Delcio Camara, and Joscelyn Bailey.

Lastly, I thank my family. My parents, Alan and Bev who instilled in me a strong work ethic and an inherent curiosity about the world. It is this curiosity that guided me towards the sciences and ultimately to hydrogeology. I thank my brother Dave and sister-in-law Joanna, who were always there to support me, especially with late night coffee runs during my 24 hr pumping tests. And most importantly I thank my wife Suzanne, without her love, support, and commitment to education, this Ph.D. would not have been possible. As promised, it is now your turn to go back to school!

Dedication

To my wife Suzanne

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Despite many decades of research, the characterization and modeling of heterogeneous aquifer systems remains an ever present challenge for practitioners and researchers alike. Obtaining accurate estimates of hydraulic parameters and understanding how they vary in space is critical for making predictions about future site behaviour (e.g. response to pumping, or predicting the migration path of a contaminant). While the importance of accurately characterizing and modeling heterogeneity is generally accepted, assumptions of homogeneity (or layered homogeneity) are often made as the number of samples required to accurately characterize/model a heterogeneous aquifer can be prohibitive, both in terms of time and money. While the homogeneity assumption may be sufficient for some investigations, many investigations benefit from an accurate understanding of subsurface heterogeneity. Examples include: predicting the fate and transport of a contaminant plume; designing effective and efficient remediation systems; designing nuclear waste disposal repositories; creating well head protection plans based on likely capture zones; etc.

Traditional characterization techniques can be grouped into two broad categories; i) those that rely on the collection of point scale data, and ii) those that rely on the interpretation of hydraulic tests. Point scale estimates are often obtained by collecting soil cores and analyzing these samples using either permeameter tests or grain size analyses. While the collection of point data (e.g. permeameter test) along a borehole provides excellent information about the 1-dimensional nature of the site at the location of the borehole, various interpolation tools are needed to turn these discrete points into a heterogeneous model. Alternatively, pumping tests can be performed to estimate larger scale values of hydraulic conductivity (K) and specific storage (S_s). These tests are often interpreted using analytical solutions that assume aquifer homogeneity (e.g., *Theis* [1935]; *Hantush* [1960], and

Neuman [1972]) and as a result provide effective, or homogeneous estimates of K and S_s. However, for heterogeneous aquifers, the homogeneity assumption is violated and the resulting hydraulic parameter estimates may be ambiguous and of questionable quality [*Wu et al*., 2005].

1.1 Review of Heterogeneity Characterization Methods

Traditionally the collection of a large number of point data has been the default approach for characterizing heterogeneous aquifer systems in detail. These approaches often rely on the collection of soil cores which are analyzed using various laboratory tests to estimate K (permeameter tests or grain size analyses) and less frequently S_s (consolidation tests). If a sufficient number of samples are collected it is possible to determine the statistical parameters of the system, such as, geometric mean, variance, and correlation lengths. Such high density data sets based on permeameter tests have been collected and analyzed by *Sudicky* [1986] and *Sudicky et al.* [2010]. By interpolating these point data, heterogeneous representations of the aquifers can be created (see Figures 1.1 and 1.2). If incorporated into a groundwater model these heterogeneous fields can be used to make predictions about the aquifer (e.g., predicting solute transport or the response of the system to pumping).

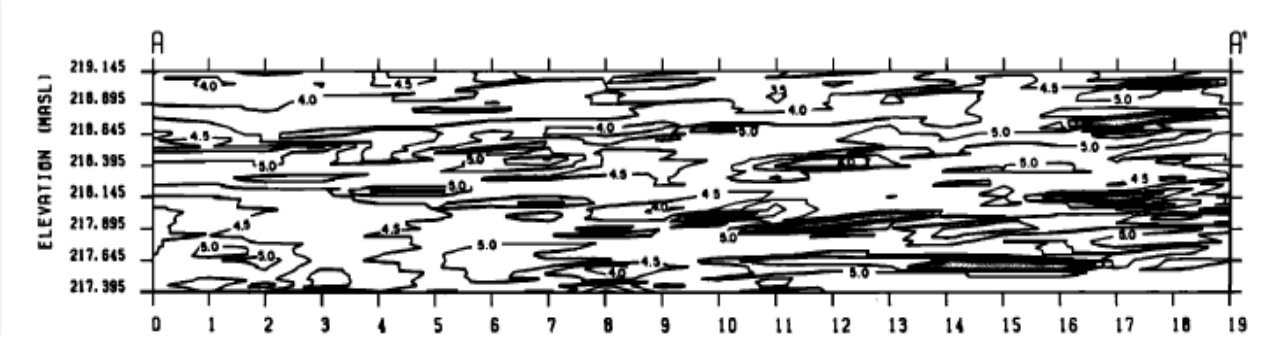


Figure 1.1. Distribution of ln (K) based on permeameter samples collected at CFB Borden (after *Sudicky* [1986]).

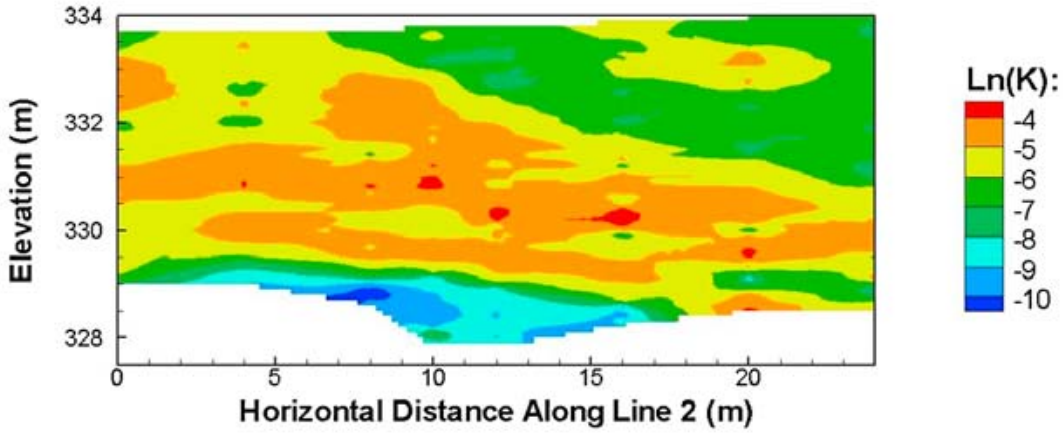


Figure 1.2. Distribution of ln (K) based on permeameter samples collected at the North Bay Site (after *Sudicky et al.,* [2010]).

Several other sites in the literature have also been characterized through intensive sampling to estimate their statistical parameters. These include: the Cape Code Site [*Wolf,* 1988]; the Columbus Site [*Rehfeldt et al.,* 1992] and the University of Waterloo North Campus Research Site (NCRS) [*Alexander et al.*, 2011]. Table 1 provides a summary of their statistical properties. For all of these sites, anisotropy is evident in the correlation lengths, with the horizontal correlation lengths being significantly greater than the vertical. The ln *K* variance ranges from 0.14 at the Cape Cod site [*Wolf,* 1988] to 6.5 at the University of Waterloo North Campus Research Site (NCRS) [*Alexander et al.,* 2011]. The sparsity of sites characterized to the degree presented in Table 1 is due to the considerable time and money required for the collection and analysis of 100's to 1000's of soil samples.

A large number of point scale *K* estimates have also been obtained using flow meter tests [e.g., *Hufschmied,* 1986; *Molz et al*., 1989; *Boman et al*., 1997], steady-state dipole flow tests [e.g., *Kabala,* 1993; *Zlotnik et al*., 2001], slug tests [e.g., *Rehfeldt et al*., 1992] and single-hole pumping or injection tests. More recently, several researchers have developed direct push slug and permeameter test approaches [e.g., *Butler et al*., 2007; *Dietrich and Leven,* 2005; *Dietrich et al*., 2008 and *Liu et*

al., 2009] which allow for the collection of a large number of *K* measurements as the instrument is advanced into unconsolidated sediments. Because tests can be conducted at narrow intervals, high resolution characterization of vertical variations in *K* is possible.

Table 1. Summary of statistical properties for highly characterized heterogeneous sites in the literature.

	n	Geometric Mean, K_G (m/s)	Minimum	Maximum	$\sigma^2 \ln K$	λ_1	λ_2	λ_3	Source
Cape Cod	825	3.5×10^{-4}	-	-	0.14	2.4	1.4	0.26	*Wolf*, 1988
Borden	1279	7.2×10^{-4}	6×10^{-6}	2×10^{-4}	0.29	2.8	2.8	0.12	*Sudicky*, 1986
Columbus	2187	5.5×10^{-4}	4.1×10^{-7}	1.5×10^{-2}	3.5-5.6	7.5-22.6		1-2.3	*Rehfeldt et al*., 1992
North Bay	1878	3.5×10^{-5}	1.4×10^{-7}	4.2×10^{-4}	2	17.19	7.39	1	*Sudicky et al.*, 2010
NCRS	471	3.0×10^{-7}	5.8×10^{-10}	2.8×10^{-4}	6.5	(4*)	-	0.15	*Alexander et al.*, 2011

* Based on omni-directional variogram

Provided a sufficient number of samples are collected and analyzed, it is possible to capture the salient features of aquifer heterogeneity. However, this is often time and cost prohibitive. For example, in order to confidently predict the transport of tracers in an alluvial aquifer at the Columbus research site, *Rehfeldt et al.* [1992] estimated that approximately 400,000 *K* measurements would be required. Clearly, this level of sampling is not practical (or even possible). As a consequence, various interpolation methods are often used to estimate the distribution of hydraulic parameters between boreholes [*de Marsily et al.*, 2005].

1.2 Review of Heterogeneity Modeling Approaches

One of the most common approaches for interpolating point data is to use geostatistical or stochastic estimators [*de Marsily et al.*, 2005]. These methods often assume that the hydraulic parameters are log-normally distributed and stationary. The user then fits a variogram model (exponential, spherical, etc.,) to the experimental variogram based on observed data (permeameter data, slug tests, etc.,). Some stochastic estimators, such as ordinary kriging, generate smoothed fields that are unable to accurately represent abrupt changes in parameter values (e.g. the contact between two distinct material types). To overcome this, categorical interpolation methods such as indicator kriging [*Journel*, 1983; *Journel and Isaaks*, 1984; *Journel and Alabert*, 1990; *Journel and Gomez-Hernandez*, 1993] and transition probability Markov chain methods [*Carle*, 1999; *Carle and Fogg*, 1997; *Weissmann et al.*, 1999] have been developed. These approaches interpolate categories, as opposed to discrete values, making it possible to reproduce abrupt material changes. For cases where very high resolution conditioning data are available, very realistic hydrofacies models can be constructed [e.g. *Maji and Sudicky*, 2008]. Figure 1.3 shows an example of a transition probability/ Markov Chain realization generated using TPROGs [*Carle*, 1999] for a fluvial deposit in southwest Germany [*Maji and Sudicky, 2008*].

Geostatistical and stochastic inversion methods are another class of techniques developed to delineate subsurface heterogeneity. In particular, the geostatistics-based inverse technique (cokriging) has received much attention. This approach produces the first and second statistical moments of hydrogeologic variables, representing their most likely estimates and their uncertainty, respectively, conditioned on available observations. Cokriging relies on the classical linear predictor theory that considers spatial correlation structures of flow processes (such as hydraulic head and velocity) and the subsurface hydraulic property, and cross-correlation between the flow processes and the hydraulic

property. In the past few decades, many researchers [e.g., *Kitanidis and Vomvoris*, 1983; *Hoeksema and Kitanidis*, 1984 and 1989; *Rubin and Dagan*, 1987; *Gutjahr and Wilson*, 1989; *Sun and Yeh*, 1992; *Harvey and Gorelick*, 1995; *Yeh et al.*, 1995 and 1996] have demonstrated its ability to estimate *K*, head, velocity, and solute concentrations in heterogeneous aquifers. However, the approach can face difficulties caused by the problem of non-uniqueness in estimated parameter fields. In addition, since cokriging is a linear estimator, its applicability to heterogeneous systems is limited to those with mild heterogeneity (variance of ln *K* ($\sigma^2_{\ln K}$) = 0.1). For highly heterogeneous systems ($\sigma^2_{\ln K}$ > 1) cokriging is unable to provide good estimates of mean *K*. [*Yeh et al.*, 1996; *Yeh and Liu*, 2000].

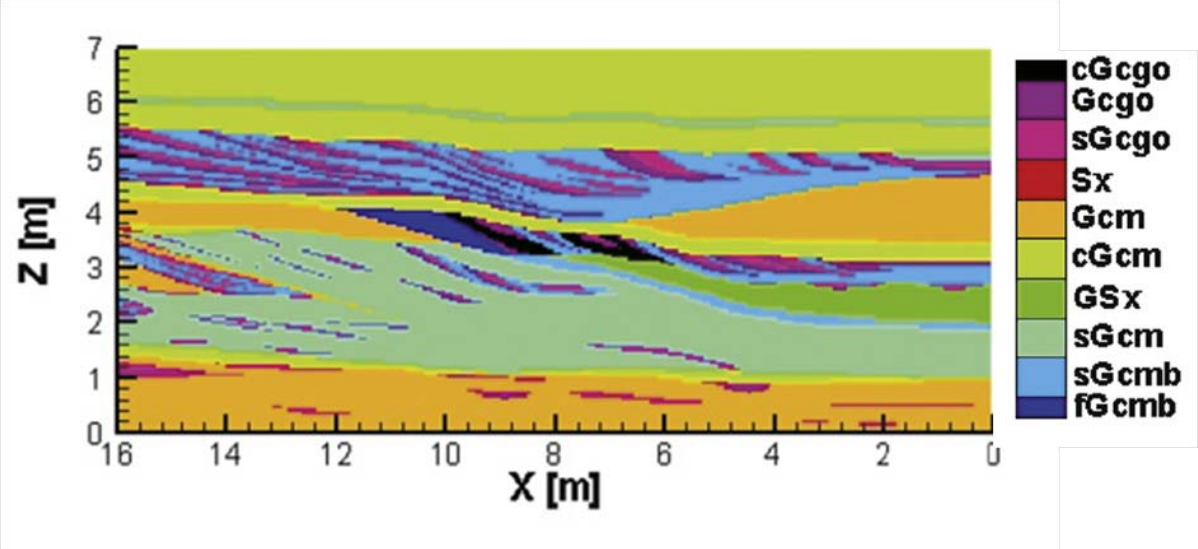


Figure 1.3. Cross-section through a TPROGs generated field for a fluvial deposit in southwest Germany (after *Maji and Sudicky* [2008]). Material types range from fine gravel (fGcmb, K = 6.0 x 10^{-07} m/s) to open framework cobbles-gravel (cGcgo, K = 1.0 m/s).

1.3 Hydraulic Tomography

Recently, hydraulic tomography (HT) has been developed to obtain information on subsurface heterogeneity of K and S_s through the analysis of multiple pumping tests. Hydraulic tomography was first suggested by *Neuman* [1987] who proposed it as an analogue to geophysical tomography. Since then, various inverse methods have been developed for HT, which use data from multiple pumping tests either simultaneously or sequentially [e.g., *Gottlieb and Dietrich*, 1995; *Yeh and Liu*, 2000; *Bohling et al.*, 2002; *Brauchler et al.*, 2003; *McDermott et al.*, 2003; *Zhu and Yeh*, 2005, 2006; *Li et al*., 2005; *Fienen et al*., 2008; *Castagna and Bellin*, 2009; *Xiang et al*., 2009]. Numerous laboratory [e.g., *Liu et al*., 2002; *Illman et al*., 2007, 2008, 2010; *Liu et al*., 2007; *Yin and Illman*, 2009; *Liu and Kitanidis*, 2011] and field experiments [e.g., *Bohling et al*., 2007; *Straface et al*., 2007; *Illman et al*., 2009; *Cardiff et al*., 2009] have been conducted to demonstrate the ability of HT to characterize/model subsurface heterogeneity. However, a rigorous study which compares the results of HT to more traditional characterization and modeling methods is generally lacking.

For the purposes of the studies presented in this thesis, the stochastic inverse methods of *Yeh and Liu* [2000] and *Zhu and Yeh* [2005] are used for steady state hydraulic tomography (SSHT) and transient hydraulic tomography (THT) respectively. Both approaches rely on the Sequential Successive Linear Estimator (SSLE) to estimate the distribution of hydraulic parameters from available hydraulic head data. SSHT uses steady state drawdown values to estimate the distribution of K, and THT uses the transient record to estimate the distribution of both K and S_s. The heterogeneous parameter distributions generated by HT are referred to as tomograms (e.g. K- *and* S_s-tomograms). For the purposes of illustration, the theory behind THT as implemented by *Zhu and Yeh* [2005], is briefly presented here. The estimation procedure assumes that transient flow in fully saturated, three-dimensional, heterogeneous, porous media is described by:

$$\nabla \cdot [K(\mathbf{x})\nabla H] + Q(\mathbf{x}_p) = S_s(\mathbf{x})\frac{\partial H}{\partial t} \quad (1.1)$$

where: H is total head [L], $\mathbf{x}$ is the spatial coordinate x = {x_1,x_2,x_3} [L], $Q(\mathbf{x}_p)$ is the pumping rate [1/T] at location $\mathbf{x_p}$, $K(\mathbf{x})$is the hydraulic conductivity [L/T], and $S_s(\mathbf{x})$ is specific storage [L^{-1}], and subject to boundary conditions:

$$H|_{\Gamma_1} = H_1, \ [K(\mathbf{x})\nabla H] \cdot \mathbf{n}|_{\Gamma_2} = q \quad (1.2)$$

where H_1 is prescribed total head at the Dirichlet boundary (Γ_1), q is the specified flux [L/T] at the Neumann boundary Γ_2, $\mathbf{n}$ is a unit vector normal to the union of Γ_1 and Γ_2. Equation (1.1) is subject to the initial condition:

$$H|_{t=0} = H_0 \quad (1.3)$$

where H_0 represents the initial total head.

The SSLE approach [*Yeh and Liu*, 2000; *Zhu and Yeh*, 2005) is an extension of the SLE (successive linear estimator) which estimates parameters for one pumping test only [*Yeh et al., 1996*; *Yeh and Zhang*, 1996; *Zhang and Yeh*, 1997; *Hanna and Yeh*, 1998; *Vargas-Guzman and Yeh*, 1999, 2002; *Hughson and Yeh*, 2000]. The SLE approach is similar to cokriging [*Yeh et al*., 1995] which seeks mean parameter fields using point data and geologic/hydrogeologic structures (i.e. the spatial covariance and cross-variance functions of parameters and hydraulic heads) to condition the estimates. SLE is different from cokriging in that is uses a "*linear estimator based on differences*

between observed and simulated hydraulic heads successively to update both conditional means and covariances of the estimates" [*Zhu and Yeh*, 2005]. This allows the non-linear relation between head and K to be handled more effectively, thus allowing systems with higher $\sigma^2_{\ln K}$ to be handled. Additionally, the iterative nature of SLE allows for the flow process to be incorporated through the use of a groundwater flow model.

SSLE uses the SLE approach and sequentially includes data sets, thus reducing the computational burden that would be encountered if all of the data sets were included simultaneously. Since hydraulic parameters are generally considered to be log-normally distributed, the SSLE assumes the natural logs of K and S_s can be represented as stochastic processes. It is therefore assumed that $lnK = \overline{K} + f$ and $lnS_s = \bar{S} + s$ where $\overline{K}$ and $\bar{S}$ are mean values, and f and s are perturbations. In addition, the transient hydraulic head is represented by $H = \overline{H} + h$, where $\overline{H}$ is the mean and h is the perturbation [*Zhu and Yeh*, 2005].

By substituting these stochastic variables into the groundwater flow equation (1.1) and taking the conditional expectation and conditioning with some observations (hydraulic head and parameter estimates) the mean flow equation is:

$$\nabla \cdot [\overline{K}_{con}(\mathbf{x})\nabla\overline{H}_{con}] + Q(\mathbf{x}_p) = \bar{S}_{con}(\mathbf{x}_p)\frac{\partial \overline{H}_{con}}{\partial t} \qquad (1.4)$$

where $\overline{K}_{con}$, $\overline{H}_{con}$, and $\bar{S}_{con}$ are conditional effective hydraulic conductivity, hydraulic head, and specific storage respectively [*Zhu and Yeh*, 2005].

The following paragraph from *Zhu and Yeh* [2005] very succinctly describes the estimation process: *"The estimation procedure starts with a weighted linear combination of direct measurements of the parameters and transient head data at different locations to obtain the first estimate of the*

parameters. The weights are calculated based on statistical moments (namely, means and covariances) of parameters, the covariances of heads in space and time, the cross covariance between heads and parameters. The first estimate is then used in the mean flow equation (1.4) to calculate the heads at observation locations and sampling times (i.e., forward simulation). Differences between the observed and simulated heads are determined subsequently. A weighted linear combination of these differences is then used to improve the previous estimates."

After the SSLE reaches set convergence criteria for a specific pumping test, the next data set is included. This in continued until all of the pumping test data sets have been included. Further details to this method are in *Zhu and Yeh* [2005].

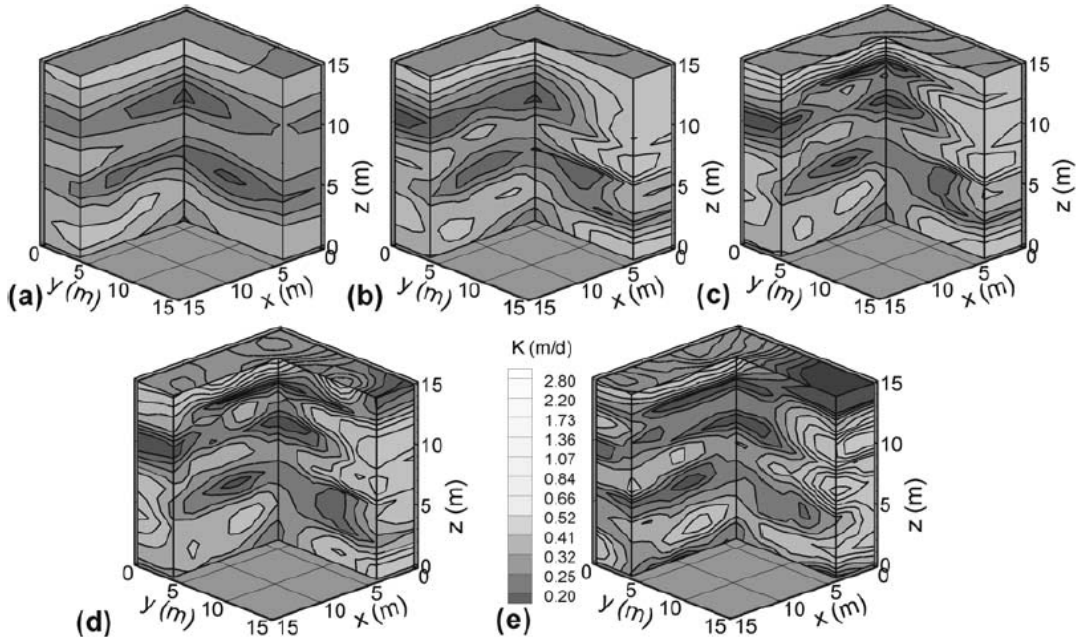


Figure 1.4. *K*-tomograms estimated using THT after the inclusion 2 tests (a), 4 tests (b), 6 tests (c), and 8 tests (d), and the true *K* field (e) [after *Zhu and Yeh*, 2005].

With THT, it is possible to estimate both K and S_s through the analysis of drawdown curves as the early-time portion of a drawdown curve is more sensitive to S_s, and the late time-portion is more sensitive to K. Additionally, since THT relies on transient data, it is not necessary to run pumping tests to steady state, which is often not possible to achieve at all observation points for field scale investigations. It should be noted that the computational requirements of THT are greater than those of SSHT, however, recent advances in parallelization and computer clustering have made THT analysis much more manageable. As part of the implementation of THT into SSLE, *Zhu and Yeh* [2005] presented a numerical illustration of the ability of SSLE to map the heterogeneous distribution of K and S_s through the use of multiple pumping tests. Figures 1.4 and 1.5 show the evolution of K- and S_s-tomograms respectively as additional pumping tests are included in the inversion.

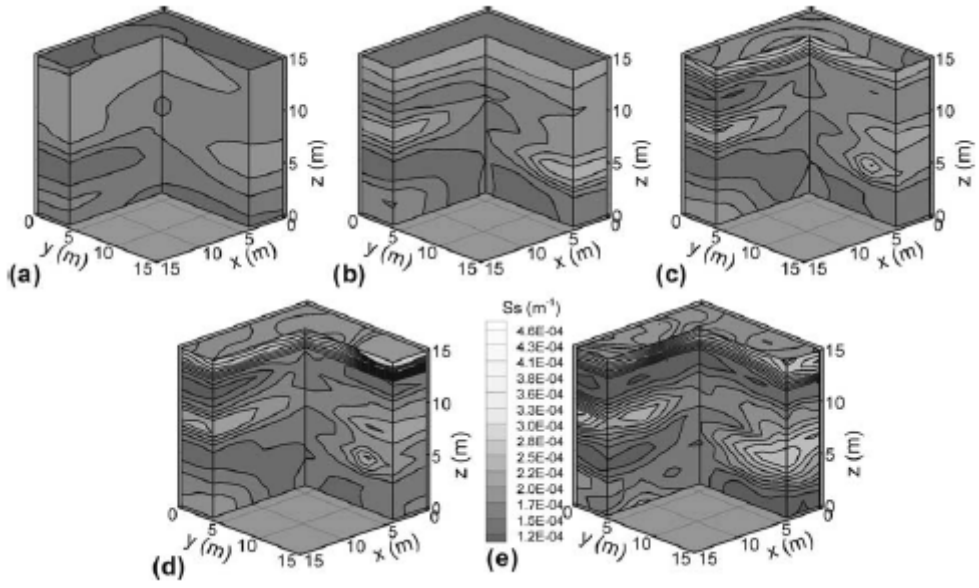


Figure 1.5. S_s-tomograms estimated using THT after the inclusion of 2 tests (a), 4 tests (b), 6 tests (c), 8 tests (d), and the true S_s field (e) [after *Zhu and Yeh*, 2005].

1.3.1 Laboratory Studies of SSHT and THT

Several laboratory studies have been performed to further assess SSHT [e.g., *Liu et al.*, 2002; *Illman et al.*, 2007, 2008, 2010] and THT [*Liu et al.*, 2007]. In particular, *Illman et al.* [2010] rigorously validated SSHT by characterizing a heterogeneus sandbox aquifer (Figure 1.6) with SSHT and then, using the estimated *K*-tomogram (Figure 1.7), predicted the reponse of the system to 16 independent pumping tests (pumping tests not included in the creation of the *K*-tomogram). In addition, *Illman et al.* [2010] applied other characterization approaches such as creating effective *K* fields, and kriging point data in which the 16 independent pumping tests were also performed.

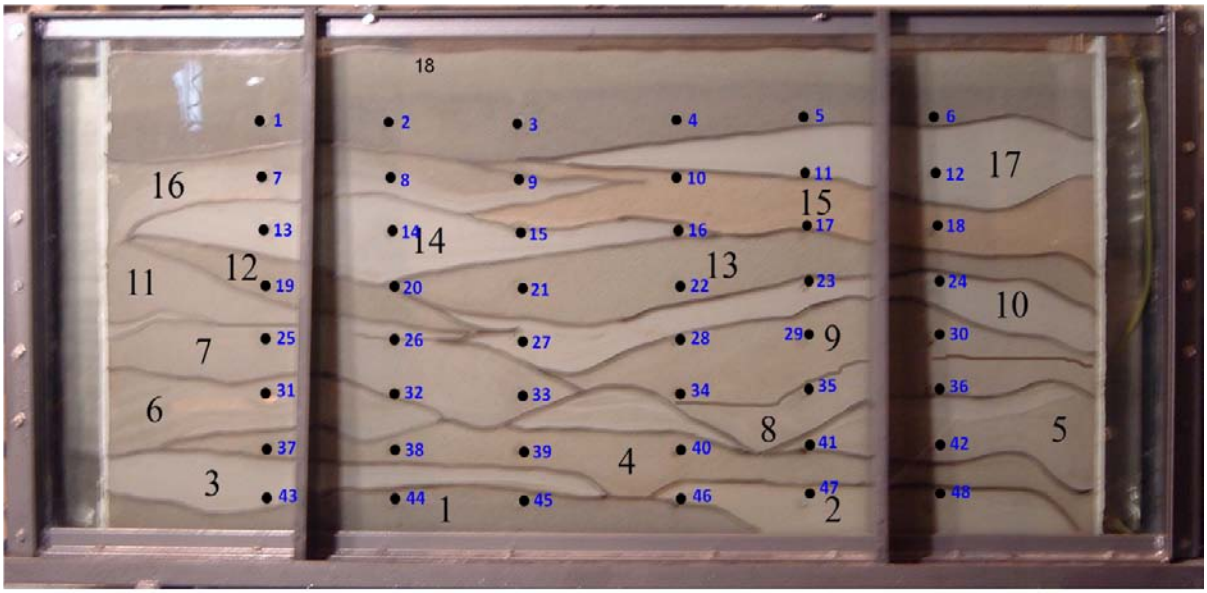


Figure 1.6. Heterogeneous sandbox created through the cyclic flux of sediment laden water [after *Illman et al.*, 2010].

Results showed that the mean *K*, and heterogeneous *K* fields estimated through kriging of small scale *K* data (core and single-hole tests) were unable to accurately predict the drawdowns

observed in the synthetic heterogeneous sandbox aquifer for the 16 independent tests. In contrast, the heterogeneous *K*-tomogram provided excellent predictions of drawdowns of independent pumping tests. Based on these results, *Illman et al*. [2010] suggested that steady-state groundwater model validation for this laboratory sandbox aquifer was possible, provided the heterogeneous *K* distribution and forcing functions (boundary conditions and source/sink terms) were sufficiently characterized. To date, a rigorous comparison of THT to other heterogeneous characterization/modeling approaches, similar to that of *Illman et al.*[2010] for SSHT, has yet to be performed at either the laboratory or field scales.

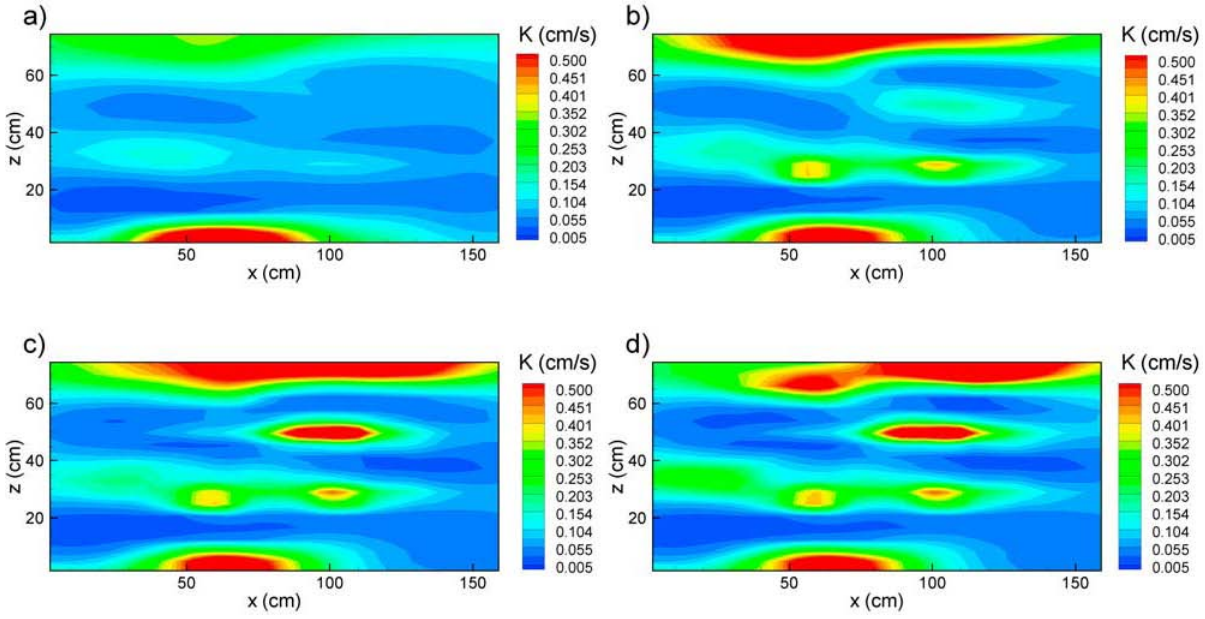


Figure 1.7. Evolution of *K*-tomogram after the inclusion of 2 tests (a), 4 tests (b), 6 tests (c), and 8 tests (d) [after *Illman et al.*, 2010].

1.3.2 Field Studies

Some of the earliest field studies demonstrating the tomographic concept were performed using pneumatic tests in fractured rock systems [*Illman and Neuman*, 2001; 2003; *Vesselinov et al.*, 2001a;

b]. In particular, *Vesselinov et al.* [2001a; b] conducted 3D pneumatic tomography in unsaturated fractured tuffs at the Apache Leap Research Site (ALRS) using 3 cross-hole pneumatic injection tests performed by *Illman and Neuman* [2001; 2003]. They used the pressure records from these tests to estimate equivalent permeability (k) and porosity values, as well as their heterogeneous distributions. The results of the pneumatic tomography were compared to kriged permeability fields based on single-hole pneumatic injection tests [*Chen et al.*, 2000] and were found to share a similar internal structure, however, no other methods were used to validate their results. Figure 1.8 shows the $\log_{10}k$ field estimated by simultaneously inverting 3 pneumatic tests.

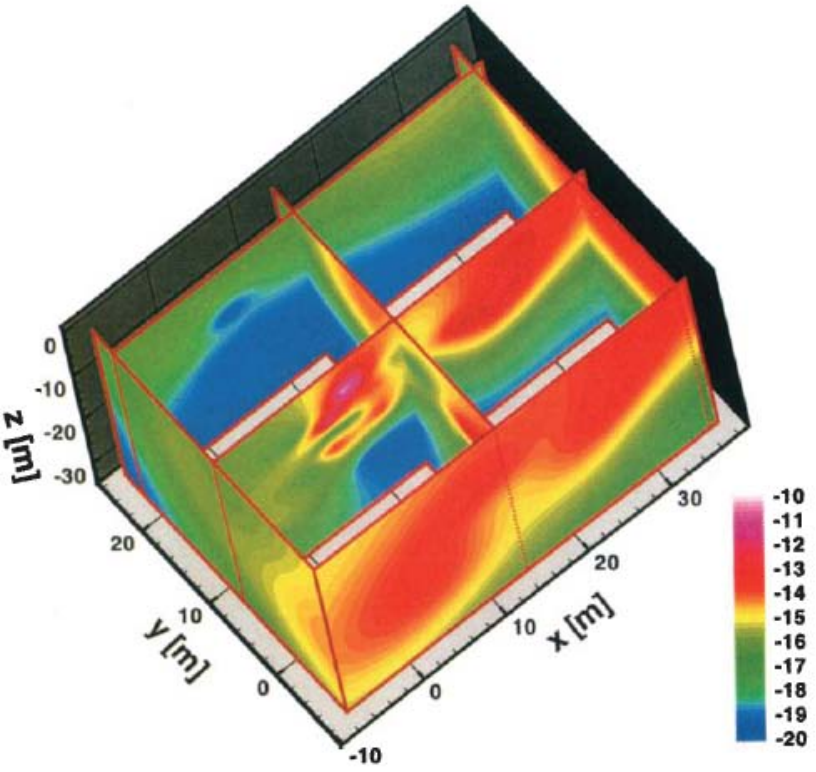


Figure 1.8. Estimated $\log_{10}$ k [m^2] field from the simultaneous inversion of 3 cross-hole pneumatic injection tests [after *Vesselinov et al.*, 2001b].

In a different study, *Bohling et al.* [2007] assessed steady shape hydraulic tomography [*Bohling et al*., 2002] in an alluvial aquifer at the Geohydrologic Experimental and Monitoring Site (GEMS) of the Kansas Geological Survey. They analyzed a total of 23 pumping tests performed at discrete intervals within 2 wells that were several meters apart. Between these 2 wells were 2 monitoring wells, each with 6 vertical observation points all aligned in a 2D plane, to record the drawdown responses. The tomographic analysis produced a 1D vertical profile of K between the 2 pumping wells which agreed reasonably with profiles obtained from an induced gradient tracer test, and direct push permeameter tests. *Bohling and Butler* [2010] followed up on their own study showing that heterogeneity in hydraulic parameters could not be obtained perpendicular to the planar configuration of wells in which they conducted the hydraulic tomography survey. That is, *Bohling and Butler* [2010] reported that heterogeneity in hydraulic parameters cannot be estimated accurately in areas where drawdown data or other information is not available.

Hydraulic tomography based on the stochastic or geostatistical inversion framework may be able to overcome some of the limitations cited by *Bohling and Butler* [2010] as it considers spatial correlations in the estimated parameters as well as cross-correlations between the parameter and heads. A field example of the use of a geostatistically based inverse model was first published by *Straface et al.* [2007] who analyzed 6 pumping tests performed sequentially within a 6-well network using the transient hydraulic tomography (THT) code developed by *Zhu and Yeh* [2005] to estimate the transmissivity (T) and storage coefficient (S) tomograms in two-dimensions. Despite the small number of wells, they concluded that the T and S tomograms were reasonable representations of the aquifer based on the geological setting of the site, however, no attempts were made to validate the tomograms. Figure 1.9 shows the estimated T and S tomograms for this study.

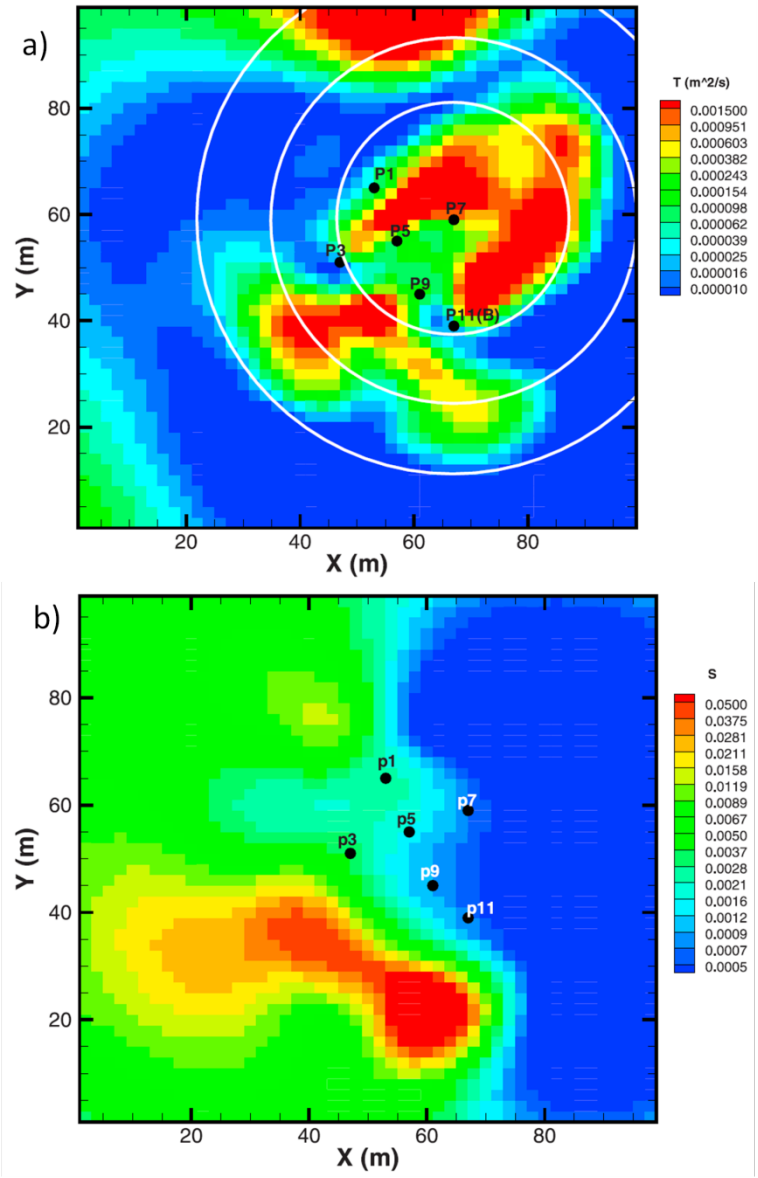


Figure 1.9. Estimated transmissivity (a) and storativity (b) fields using hydraulic tomography at a well field in Montalto Uffugo, Scalo, Italy [after *Straface et al.*, 2007].

At the Krauthausen test site in Germany, *Li et al.* [2008] used a geostatistical inverse approach to jointly analyze steady-state drawdown and borehole flow meter data from multiple pumping tests to estimate the *K* distribution in three-dimensions. They found that jointly inverting

both steady-state drawdown and flow meter data produced an improved 3D structure when compared to just inverting pumping test data.

The first SSHT survey in an unconfined aquifer was performed by *Cardiff et al.* [2009] using 9 pumping tests at the Boise Hydrogeophysical Research Site (BHRS) to estimate the distribution of depth-averaged *K*. They found that the *K*-tomogram contained expected geological features, additionally, the uncertainty bounds on the estimation indicated that *K* was well constrained within the central portion of the research site where the pumping and observation well network was located. Figure 1.10 shows the estimated *K*-tomograms as well as the uncertainty associated with these estimates.

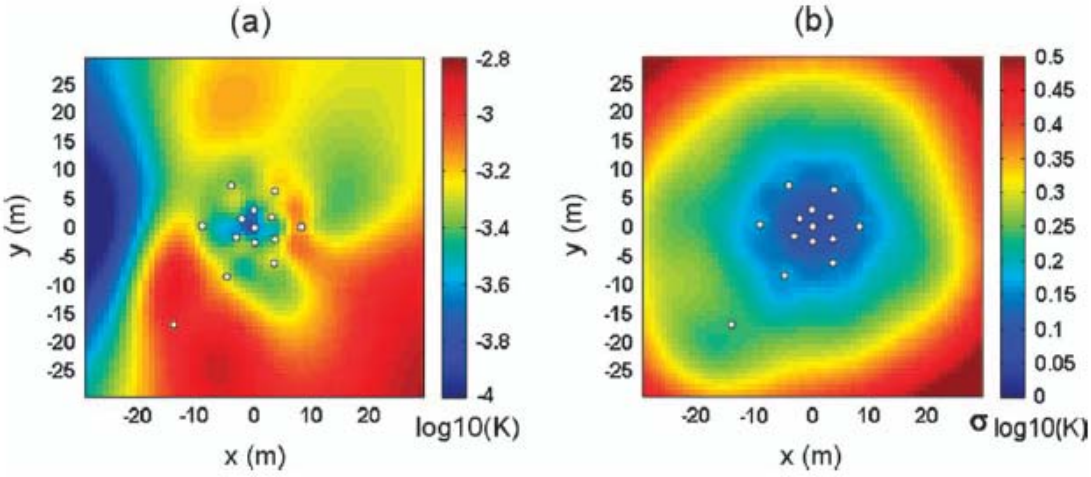


Figure 1.10. Estimated *K* field (a) and associated uncertainty (b) for HT performed at the BHRS [after *Cardiff et al.*, 2009].

A large-scale application of THT in fractured rock was demonstrated by *Illman et al.* [2009] at the Mizunami Underground Research site in Japan. Using 2 pumping tests, they estimated 3D *K*- and S_s-tomograms as well as their uncertainties (see Figure 1.11 for the estimated *K*-tomogram). This was the first application of three-dimensional hydraulic tomography at the field scale utilizing

transient drawdown data. Several continuous high K and low S_s zones were identified and interpreted as possible fault zones. This field investigation highlighted the potential use of hydraulic tomography in fractured rock environments to identify hydraulic connections between boreholes. However, the evaluation of the K- and S_s-tomograms was limited to available fault data, several drawdown data from one of the cross-hole pumping tests, and coseismic responses in wells. That is, the validity of K- and S_s-tomograms was not confirmed rigorously through the prediction of independent drawdown inducing events such as shown previously by *Illman et al*. [2007, 2008, 2010], and *Liu et al*. [2007] through sandbox studies. While these field studies demonstrate the utility of hydraulic tomography for mapping heterogeneous aquifer systems, a comprehensive study designed to validate THT in three-dimensions is lacking.

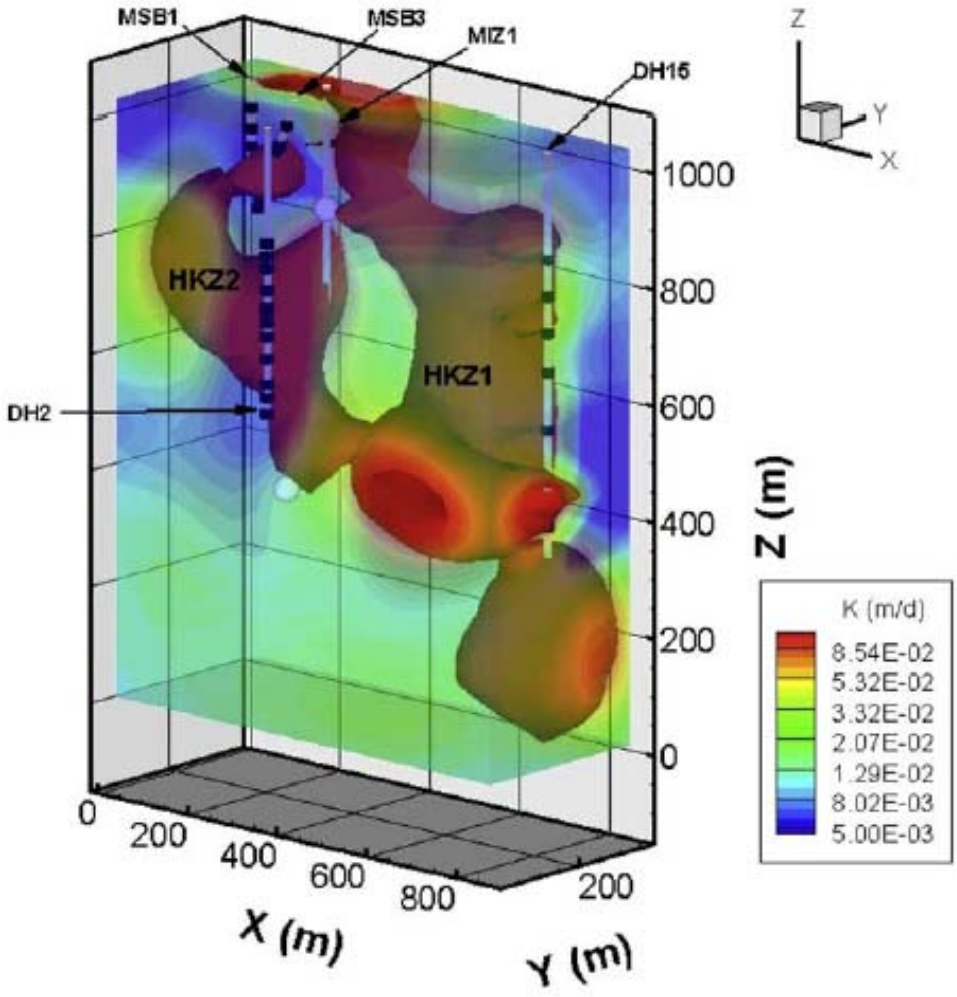


Figure 1.11 *K*-tomogram estimated from two pumping tests performed at the Mizunami Underground Research Site in Japan [after *Illman et al.*, 2009].

1.3.3 Other Applications of Hydraulic Tomography

Solute Transport

Recently, tomograms estimated by HT have been validated by predicting solute transport. For example, using a synthetic heterogeneous aquifer *Ni et al.* [2009] compared the ability of various heterogeneity characterization approaches to predict the migration of solute transport. These

approaches included; calculating the geometric mean of the system and treating it to be homogenous, creating a kriged field based on point data, and performing a hydraulic tomography survey. They found that HT was best able to reproduce the true flow field as well as the breakthrough curves (BTCs) measured at a number of down gradient locations. To date, no laboratory, or field studies have been conducted that use HT to map aquifer heterogeneity for the purposes of predicting solute transport.

Heterogeneous Unconfined Aquifer Systems

For unconfined heterogeneous aquifer systems, *Mao et al.* [*in press*] have demonstrated that the drawdown response to pumping is more sensitive to the distribution of saturated zone hydraulic parameters (K and S_s) than unsaturated zone parameters (saturated water content, residual water content, and the parameters describing the primary drainage curve). This suggests that provided accurate effective estimates of unsaturated zone parameters are available, it may be possible to use these along with K- and S_s-tomograms estimated by HT to predict the response of an unconfined aquifer to pumping. To date, no laboratory or field studies have been published that investigate these observations experimentally.

Mapping of Non-Aqueous Phase Liquid (*NAPL*) *Distribution*

Hydraulic tomography has also been used to map aquifer heterogeneity for partitioning tracer tomography in a method referred to as hydraulic/partitioning tracer tomography (HPTT) [*Yeh and Zhu*, 2007]. This method first uses hydraulic tomography to map aquifer heterogeneity, followed by a series of partitioning tracer tests. These partitioning tracer tests, if performed in different configurations (tomographically) can provide information on the distribution of Dense Non-Aqueous Phase Liquid (DNAPL) mass within a source zone, potentially leading to more focused and efficient

remediation efforts. To date, only numerical studies have been performed to validate the method and a laboratory assessment of the method is lacking.

1.4 Research Objectives

The primary objective of this thesis is to assess hydraulic tomography in a variety of settings (laboratory and field) using a number of different methods (comparison to known geology, prediction of independent pumping tests, prediction of solute transport, and prediction of flow to a well in a heterogeneous unconfined aquifer). The majority of this research focuses on performing THT at the field scale in 3D, and validating the estimated K- and S_s-tomograms. Prior to the presentation of the field research, laboratory studies that examine the predictive ability of HT are presented. The results of these studies provide a deeper understanding of the utility, and potential limitations of hydraulic tomography for the characterization/modeling of heterogeneous aquifers.

Based on the areas identified in the literature review that require further research, the goals of this thesis are:

1) Assess the ability of THT to map both K and S_s in a laboratory sandbox by rigorously validating the method through the prediction of independent pumping tests and comparison to other characterization techniques (similar to the work of *Illman et al.,* [2010] for SSHT);

2) At the laboratory scale, determine whether K tomograms estimated using SSHT can be used to accurately predict the migration of a conservative tracer within a laboratory sandbox, and compare the performance of SSHT to other characterization techniques;

3) Validate THT by, estimating K- and S_s-tomograms for a heterogeneous sandbox aquifer while fully saturated; then, in conjunction with estimates of unsaturated zone parameters use the tomograms to predict the response of the system to drainage caused by an unconfined pumping test;

4) Analyze partitioning tracer tests performed in a tomographic manner at the laboratory scale to assess the ability of HPTT to map the distribution of NAPL in a laboratory sandbox with a simple heterogeneity pattern;

5) Use THT to map the 3D heterogeneous distribution of K and S_s in a highly heterogeneous aquifer at the field scale;

6) Determine whether 3D K- and S_S-tomograms estimated for a heterogeneous aquifer at the field scale are accurate, by: i) comparing the estimated tomograms to known geology; ii) assessing their ability to predict pumping tests; and iii) comparing the tomograms to other subsurface heterogeneity characterization/mapping methods; and

7) Determine whether it is possible to interpret pumping test data impacted by poroelastic effects using traditional approaches to estimate K and S_s (this last goal (explained in more detail in chapter 3) is a result of anomalous drawdown behaviour referred to as the Noordbergum effect observed during field tests).

1.5 Thesis Organization

This thesis contains many individual studies (both laboratory and field) that were written in manuscript format. To avoid repetition of material, and to avoid excess length, each manuscript is briefly summarized in chapters 2 (laboratory studies) and 3 (field studies). The objective of these summaries is to provide the reader with an overview of the work performed and how the conclusions

were reached without the need to read the full version of each manuscript. In addition, for similar studies (e.g. several field studies performed at the same site), this approach avoids repetition of material (e.g. site description, field methods, etc.,). All manuscripts (in their entirety) are provided as appendices should the reader require additional information. The laboratory studies presented in chapter 2 focus on validating hydraulic tomography through: 1) the prediction of independent transient pumping tests (section 2.1); 2) using hydraulic tomography to map heterogeneity for the prediction of solute transport (section 2.2); and, 3) by using hydraulic tomography to predict the response of a heterogeneous unconfined sandbox to pumping (section 2.3). The last section of chapter 2 (section 2.4) analyzes multiple partitioning tracer tests using hydraulic/partitioning tracer tomography to estimate the distribution of DNAPL within a heterogeneous laboratory sandbox.

All of the field studies described in chapter 3 were performed at the same field site. To avoid introducing the site for each summary, section 3.1 provides details of the site (description, instrumentation, and the pumping tests performed). Section 3.2, though a numerical study, is included with the field studies because the motivation for this study was the observation of anomalous drawdown responses caused by poroelastic effects during pumping tests at the field site. Based on these observations a numerical study was conducted to assess the impact of poroelastic effects on the estimation of hydraulic parameters. The remainder of chapter 3 presents various characterization and modeling efforts. Section 3.3 analyzes drawdown data from multiple pumping tests using transient hydraulic tomography, and section 3.4 compares transient hydraulic tomography to other characterization techniques. Chapter 4 presents a discussion on hydraulic tomography based on the findings of chapters 2 and 3, and chapter 5 presents the main conclusions from these individual studies, and makes recommendations for future research.

1.6 Thesis Contributions

This section summarizes my contributions to each of these individual sections.

1.6.1 Capturing Heterogeneity in Groundwater Flow Parameters: Comparison of Approaches Through Controlled Sandbox Experiments

My contribution to this research was the inverse modeling of previously performed pumping tests in a laboratory sandbox aquifer, the creation of various homogeneous and heterogeneous model representations of the heterogeneous sandbox aquifer, all accompanying data analysis, and writing this manuscript.

1.6.2 Comparison of Approaches for Predicting Solute Transport: Sandbox Experiments

My contribution to this research was the completion of dipole pumping tests for HT analysis, the tracer experiment, all inverse and forward modeling, analysis of results, editing this manuscript.

1.6.3 Flow to a Well in a Heterogeneous Unconfined Aquifer: Insights from an Intermediate Scale Sandbox

My contribution to this research was the packing and instrumentation of the sandbox, completion of all pumping tests, drainage tests, inverse modeling forward modeling, as well as writing this manuscript.

1.6.4 Hydraulic/Partitioning Tracer Tomography for DNAPL Source Zone Characterization: Small-Scale Sandbox Experiments

For this study the laboratory experiments were completed by *Massi* [2007]. My contribution was the inverse modeling performed with this data, and editing this manuscript.

1.6.5 Estimating Hydraulic Parameters When Poroelastic Effects are Significant

For this study I received assistance from Dr. Hsieh of the USGS with model setup and design of numerical experiments. However, I conducted all model runs, PEST analysis, data analysis, and writing of this manuscript.

1.6.6 North Campus Research Site (NCRS) Field Experiments

The two field studies presented in this study were performed at the NCRS. I was responsible for the design and installation of the wells and instrumentation at the site and performed all pumping tests. Additionally, I performed all data analysis, inverse and forward modeling presented in these studies, as well as writing these manuscripts.

Chapter 2

Laboratory Studies of Hydraulic Tomography

2.1 Capturing Heterogeneity in Groundwater Flow Parameters: Comparison of Approaches Through Controlled Sandbox Experiments

Prior to conducting THT at the field scale, a series of laboratory experiments were conducted to further validate the approach in comparison to other more common heterogeneity characterization and modeling techniques. This section provides a summary of the paper entitled "Capturing heterogeneity in groundwater flow parameters: comparison of approaches through controlled sandbox experiments", by *Berg, S.J. and W.A. Illman* [*submitted manuscript*]. The manuscript is included as Appendix A.

To date SSHT has been validated in the lab through the prediction of independent pumping tests and comparison to other heterogeneity modeling approaches [*Illman et al.,* 2010a]. During the pumping tests performed by *Illman et al.* [2010a] the transient drawdown records were also collected; it is these data that were used for this study.

Using the heterogeneous sandbox aquifer presented in Figure 1.6, 5 groundwater models were created using various characterization approaches; 2 homogeneous, and 3 heterogeneous. The homogeneous cases were constructed by: 1) calculating the geometric mean K and S_s from 48 single-hole pumping tests; and 2) estimating effective values of K and S_s by simultaneously matching the response at all ports during a pumping test through inverse modeling. The three heterogeneous cases were created by; 1) kriging K and S_s estimates made from the analysis of 48 single-hole pumping tests (case1), 2) creating a geological model based on the layering evident in the sandbox and calibrating each layer to the drawdown response observed during pumping at port 20 (case 2), and 3) performing THT using the transient drawdown record from 8 pumping tests (performed at ports 2, 5, 14, 17, 32,

35, 44, and 47) (case 3). For the purposes of this summary only the heterogeneous cases are examined. Figures 2.1a, b, and c show the K and S_s fields for the heterogeneous cases 1, 2, and 3 respectively.

Table 2. Characteristics of each layer used to create a synthetic heterogeneous aquifer (after *Illman et al.*, 2010).

Layer	Sand	d_{50} (mm)	K (cm/s) Shepherd	Core Permeameter K (cm/s)*	Single-hole K (cm/s)*	Single-hole S_s (cm^{-1})*
1	20/30	0.75	1.03×10^{-1}	3.20×10^{-2}	5.32×10^{-2}	2.12×10^{-4}
2	4030	0.35	2.99×10^{-2}	5.29×10^{-2}	5.67×10^{-2}	2.60×10^{-4}
3	F-85	0.15	7.29×10^{-3}	7.14×10^{-2}	5.70×10^{-2}	5.00×10^{-4}
4	20/40	0.58	6.68×10^{-2}	5.68×10^{-2}	5.10×10^{-2}	2.22×10^{-4}
5	mix	0.46	N/A	N/A	N/A	N/A
6	mix	0.46	N/A	8.16×10^{-2}	5.00×10^{-2}	4.00×10^{-4}
7	#12	0.52	5.70×10^{-2}	1.27×10^{-1}	7.35×10^{-2}	4.20×10^{-4}
8	F32	0.5	5.33×10^{-2}	1.34×10^{-1}	4.50×10^{-2}	1.75×10^{-4}
9	20/40	0.58	6.68×10^{-2}	8.69×10^{-2}	4.60×10^{-2}	2.15×10^{-4}
10	F-65	0.2	1.20×10^{-2}	1.13×10^{-1}	8.25×10^{-2}	1.14×10^{-3}
11	#12	0.52	5.70×10^{-2}	1.37×10^{-1}	2.05×10^{-1}	2.15×10^{-4}
12	16/30	0.87	1.32×10^{-1}	3.40×10^{-2}	4.95×10^{-2}	6.32×10^{-4}
13	20/30	0.75	1.03×10^{-1}	2.60×10^{-1}	1.05×10^{-1}	9.80×10^{-4}
14	F-75	0.17	9.22×10^{-3}	9.79×10^{-2}	5.70×10^{-2}	9.80×10^{-4}
15	20/40	0.58	6.68×10^{-2}	8.58×10^{-2}	7.50×10^{-2}	2.00×10^{-3}
16	mix	0.46	N/A	4.16×10^{-2}	2.69×10^{-2}	7.11×10^{-4}
17	F-85	0.15	7.29×10^{-3}	4.51×10^{-2}	4.47×10^{-2}	1.14×10^{-3}
18	20/30	0.75	1.03×10^{-1}	1.45×10^{-1}	1.16×10^{-1}	3.38×10^{-3}

*** If multiple ports are in the same layer then the geometric mean is presented**

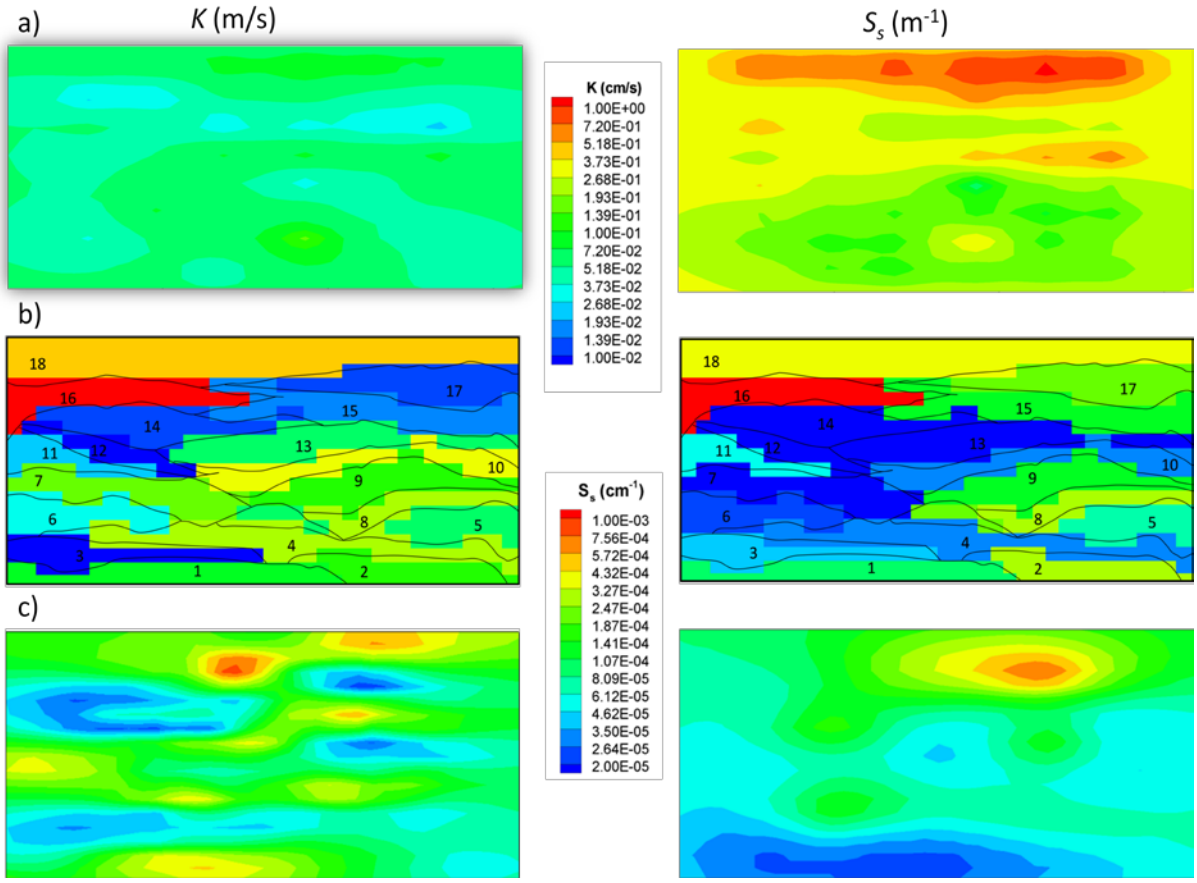


Figure 2.1. *K* (left) and S_s fields (right) for the three heterogeneous cases; a) Case 1 - kriged single-hole data, b) Case 2 - calibrated geological model, and c) Case 3 - THT.

While initially appearing quite different, there are some consistent features in all three heterogeneous fields in Figure 2.1. For example, all three *K* fields have a low *K* zone located near the top third of the sandbox that almost extends across the entire width of the sandbox. For the S_s fields, a consistent pattern of lower S_s near the bottom of the sandbox, and higher S_s near the top are present in all three fields. This pattern has been observed in other sandbox studies [*Liu et al.,* 2007] and is attributed to the bottom of the sandbox being more compacted than the top due to the overlying sand.

The simulation of independent pumping tests as a method for validating groundwater models has been used in other HT studies [*Illman et al.*, 2007, 2008, 2010; *Liu et al.*, 2007]. If a heterogeneous model is able to accurately predict a number of independent pumping tests (ones not used for characterization purposes), then it can be stated that the heterogeneous model is a reasonable representation of the true heterogeneity. To determine which heterogeneous realization best represents the heterogeneous sandbox, 16 independent pumping tests were simulated for each case. These tests were performed at ports 8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42. For comparison purposes, only the results from 4 of these pumping tests are presented for the two best performing cases; the calibrated geological model (case 2), and the THT analysis (case 3). The results for the homogeneous cases and the kriged case can be found in Appendix A. Figures 2.2 and 2.3 show scatterplots for the 4 independent pumping tests for the calibrated geological model, and for the THT analysis respectively. The scatterplots show observed vs. simulated drawdown for all observation ports at four different times (t = 0.5, 2, 5, and 10 seconds).

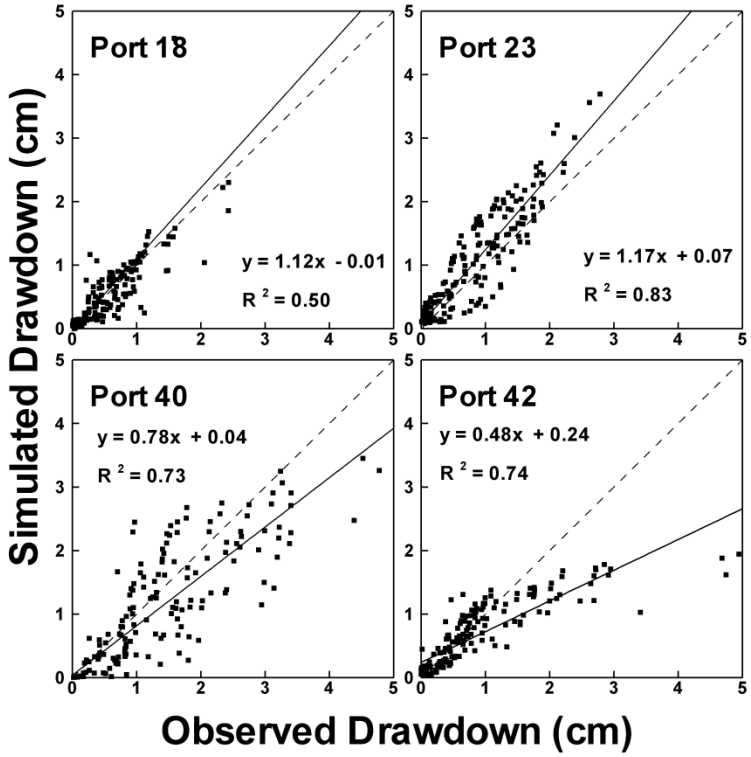


Figure 2.2. Scatterplots of observed vs simulated drawdowns from independent pumping tests 18, 23, 40, and 42 at t = 0.5, 2, 5, and 10 seconds. Simulated drawdown values were obtained through numerical simulations with a calibrated geological model of K and S_s. The dashed line is a 1:1 line indicating a perfect match. The solid line is a best fit line, and the parameters describing this line are on each plot.

The scatterplots for the calibrated geological model (Figure 2.2) tend to cluster around the 1:1 line (with the exception of Port 42) indicating the calibrated geological model captures the drawdown in an average sense, however, the degree of scatter suggests that heterogeneity is not accurately represented. In contrast, Figure 2.3 (THT analysis) shows a significant improvement for all pumping tests. The best fit lines for the all of the independent pumping tests are much closer to the 1:1 line and the degree of scatter is significantly less, suggesting that the K- and S_s-tomograms from THT analysis accurately capture the heterogeneity in both K and S_s for this sandbox.

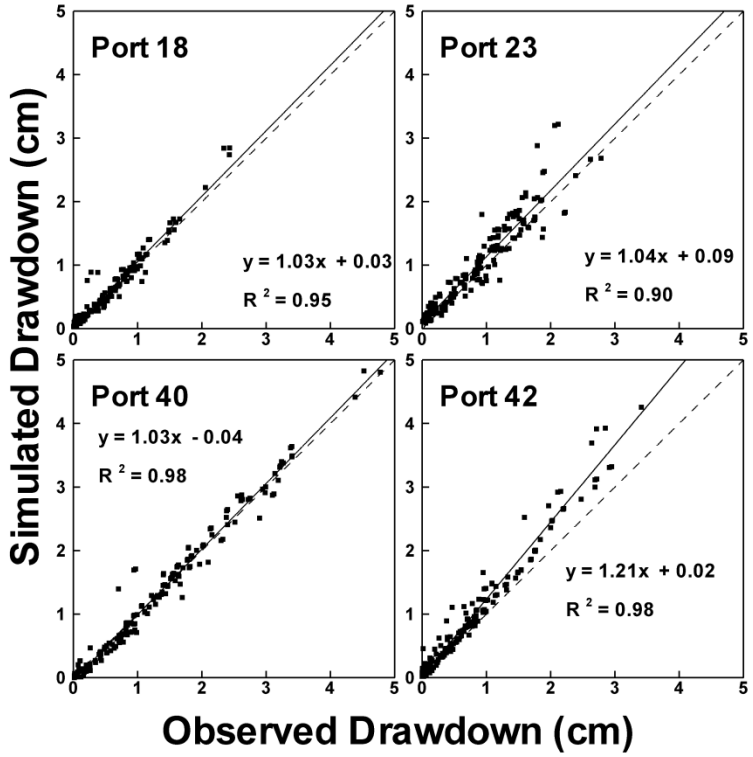


Figure 2.3. Scatterplots of observed vs. simulated drawdowns from independent pumping tests 18, 23, 40, and 42 at t = 0.5, 2, 5, and 10 seconds. Simulated drawdown values were obtained through numerical simulations with *K* and S_s tomograms from transient hydraulic tomography. The dashed line is a 1:1 line indicating a perfect match. The solid line is a best fit line, and the parameters describing this line are on each plot.

In addition to the scatterplot comparison, a direct drawdown comparison was made for the independent pumping test performed at port 40. Figures 2.4 and 2.5 show the simulated and observed drawdown curves at 16 ports for the calibrated geological model and THT respectively. The dots indicate the observed drawdown, and the solid line is the simulated drawdown. For the calibrated geological model (Figure 2.4), the simulated results agree reasonably well at most observation ports, with the exception of the shallow ports (ports 1-13). This possibly indicates that the hydraulic parameters estimated for the upper portion of calibrated geological model were in error.

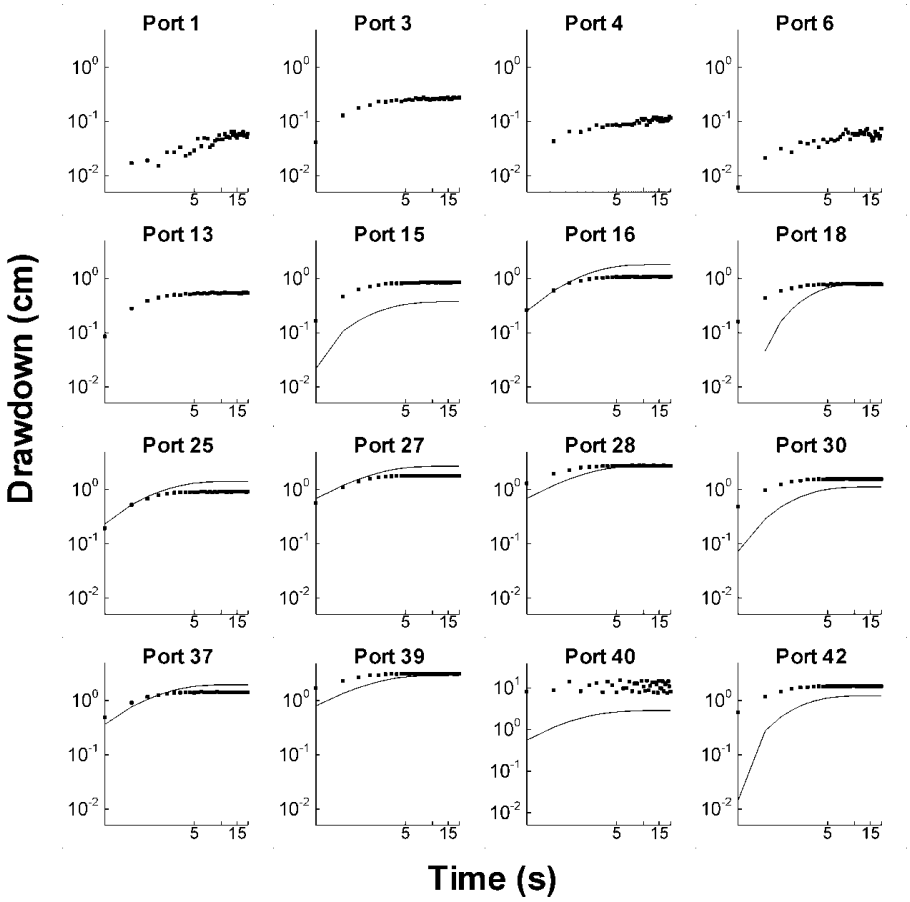


Figure 2.4. Observed and simulated drawdown comparison at 16 ports for the calibrated geological model for the pumping test performed at port 40. Squares indicate the observed drawdown and the solid line is the simulated drawdown.

As with the scatterplots, a significant improvement is seen for the THT case. Particularly for the ports located in the upper region of the sandbox (ports 1, 3, 4 and 6). The forward numerical modeling of the 16 independent pumping tests using the K- and S_s-tomograms computed via THT convincingly showed that these tomograms led to the smallest discrepancy between observed and simulated drawdowns. This sandbox experiment demonstrated the robustness of THT conducted in a controlled laboratory environment.

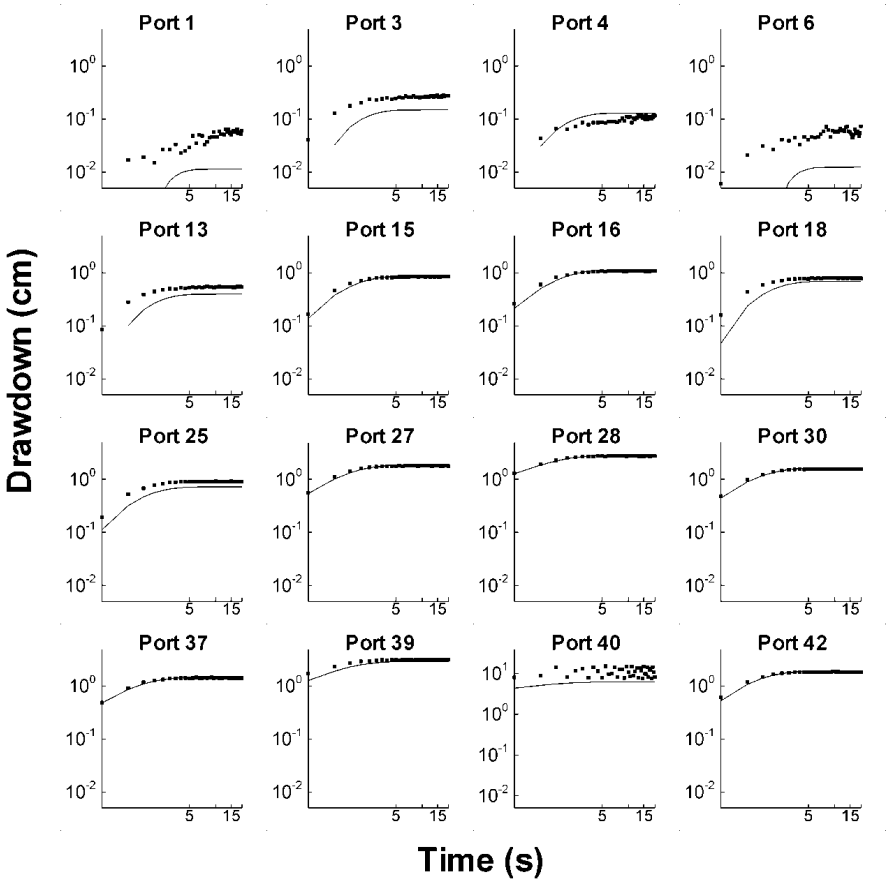


Figure 2.5. Observed and simulated drawdown comparison at 16 ports for THT for the pumping test performed at port 40. Squares indicate the observed drawdown and the solid line is the simulated drawdown.

2.2 Comparison of Approaches for Predicting Solute Transport: Sandbox Experiments

Traditionally K- and S_s-tomograms obtained from HT have been validated by examining their ability to match drawdowns observed during pumping tests, or by comparing the estimated fields to known geology. In this study, "Comparison of Approaches for Predicting Solute Transport: Sandbox Experiments" by *Illman, W.A., S.J. Berg, and T-C. J. Yeh* [*submitted manuscript*] various heterogeneous characterization approaches were validated by assessing their ability to predict the migration of a solute in a heterogeneous aquifer. This manuscript is included as Appendix B. The author's contribution to this study was; conduction the laboratory experiments, all inverse and forward modeling.

Three different representations were created for the heterogeneous sandbox shown in Figure 1.6. These include: 1) a homogeneous K field based on the effective parameter approach of *Gelhar and Axness* [1983]; 2) a heterogeneous kriged K field (similar to Figure 2.1); and, 3) a K-tomogram created using 8 dipole pumping tests (See Figure 2.6) estimated via SSHT. Appendix B provides additional details on how these different cases were created.

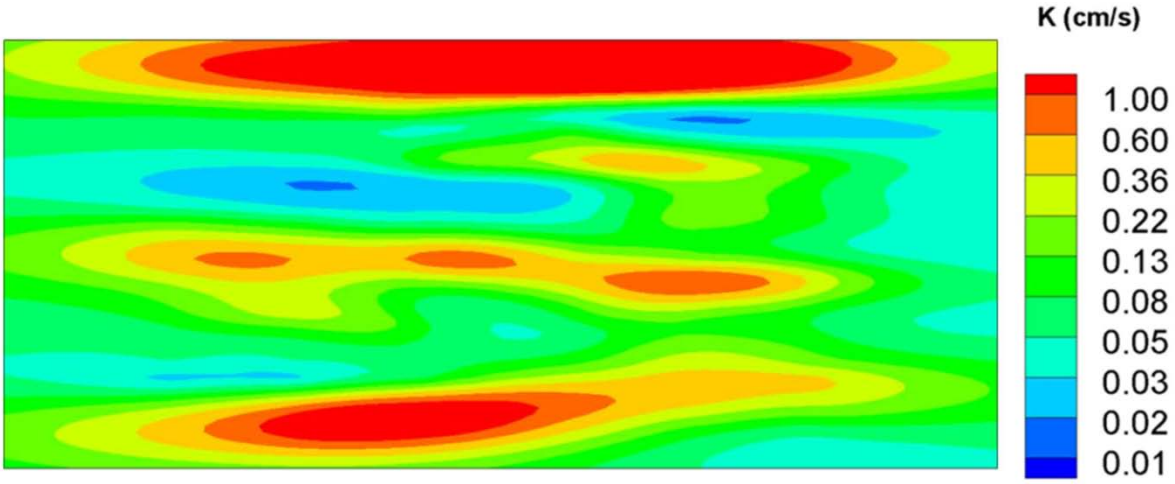


Figure 2.6. *K*-tomogram generated from 8 dipole pumping test using the steady state hydraulic tomography (SSHT) algorithm of *Yeh and Liu* [2000].

After the sandbox was characterized using these various approaches, a dipole tracer test was conducted by injecting a conservative tracer solution dyed with food colouring at port 42 at a rate of 372.4 mL/min and extracting at port 7 at 304 mL/min. A water mass balance was then achieved by setting 12 sampling lines at a cumulative rate of 68.4 mL/min.. The tracer was injected for 10 minutes followed by approximately 5 hrs of clean water. Figure 2.7 shows the injection, extraction and sampling ports for the dipole conservative tracer test, and Figure 2.8 shows the location of the conservative tracer within the sandbox at 6 different times.

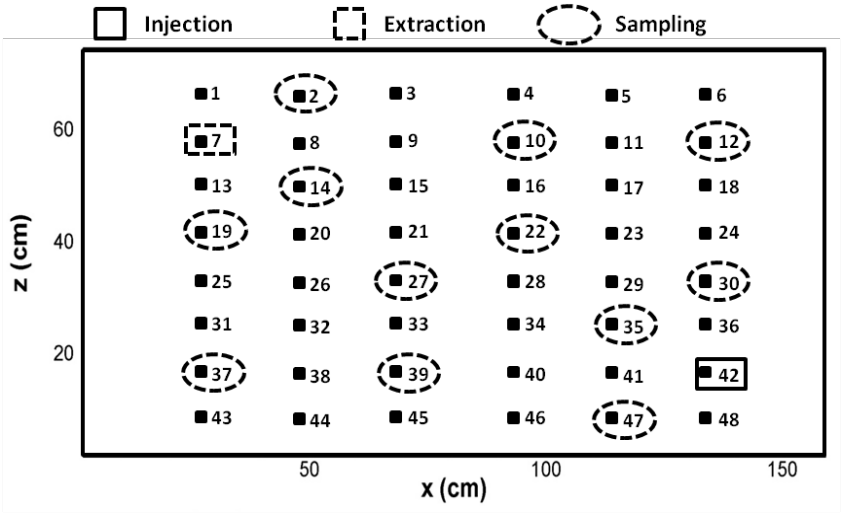


Figure 2.7. Schematic diagram showing the injection, extraction, and sampling ports during the dipole conservative tracer test.

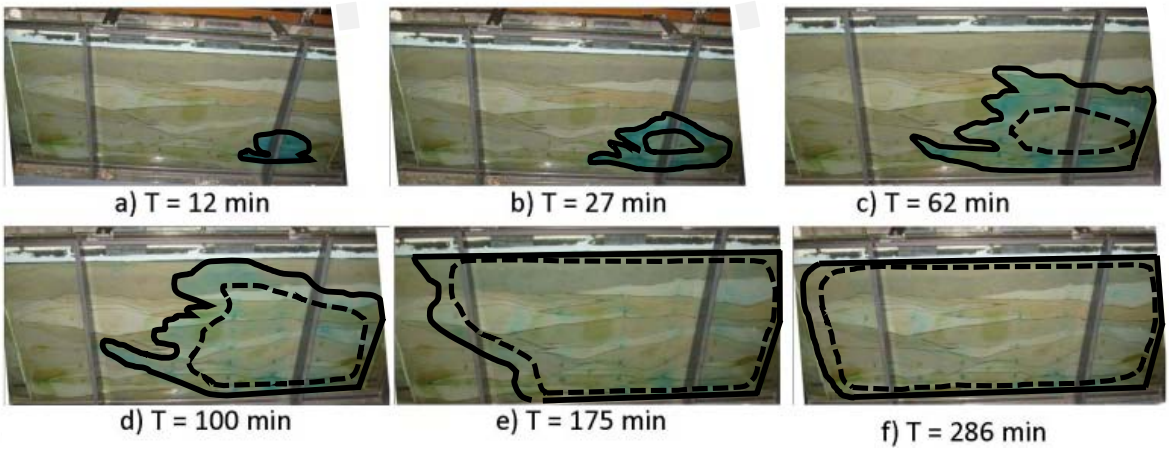


Figure 2.8. Photographs of sandbox during tracer test 1 at various times: a) t = 12 min; b) t = 27 min; c) t = 62 min; d) t = 100 min; e) t = 175 min; and f) t = 286 min. The solid line represents the edges of the dyed tracer plume. The dashed lines are approximations of the extent of the tracer plume. Note: at t = 175 and t = 286, most of the dye has left the system, however, some dye is trapped around the wells even after the tracer has been flush through the system giving the tank a blue-green tinge at these times.

Next, all three model domains were used to simulate the tracer test using the finite-element code MMOC3 [*Yeh et al.*, 1993]. For these simulations, Case 1 was assigned a longitudinal dispersivity of 50 cm based on the *Gelhar and Axness* [1983] effective parameter approach, and cases 2 and 3 were assigned macrodispersivity values of zero. Additional details to the calculation of the effective macrodispersivity are provided in Appendix B. Figures 2.9 a, b, and c show the simulated tracer results for the effective parameter field (case 1), the kriged field (case 2), and the SSHT tomogram (case 3) respectively.

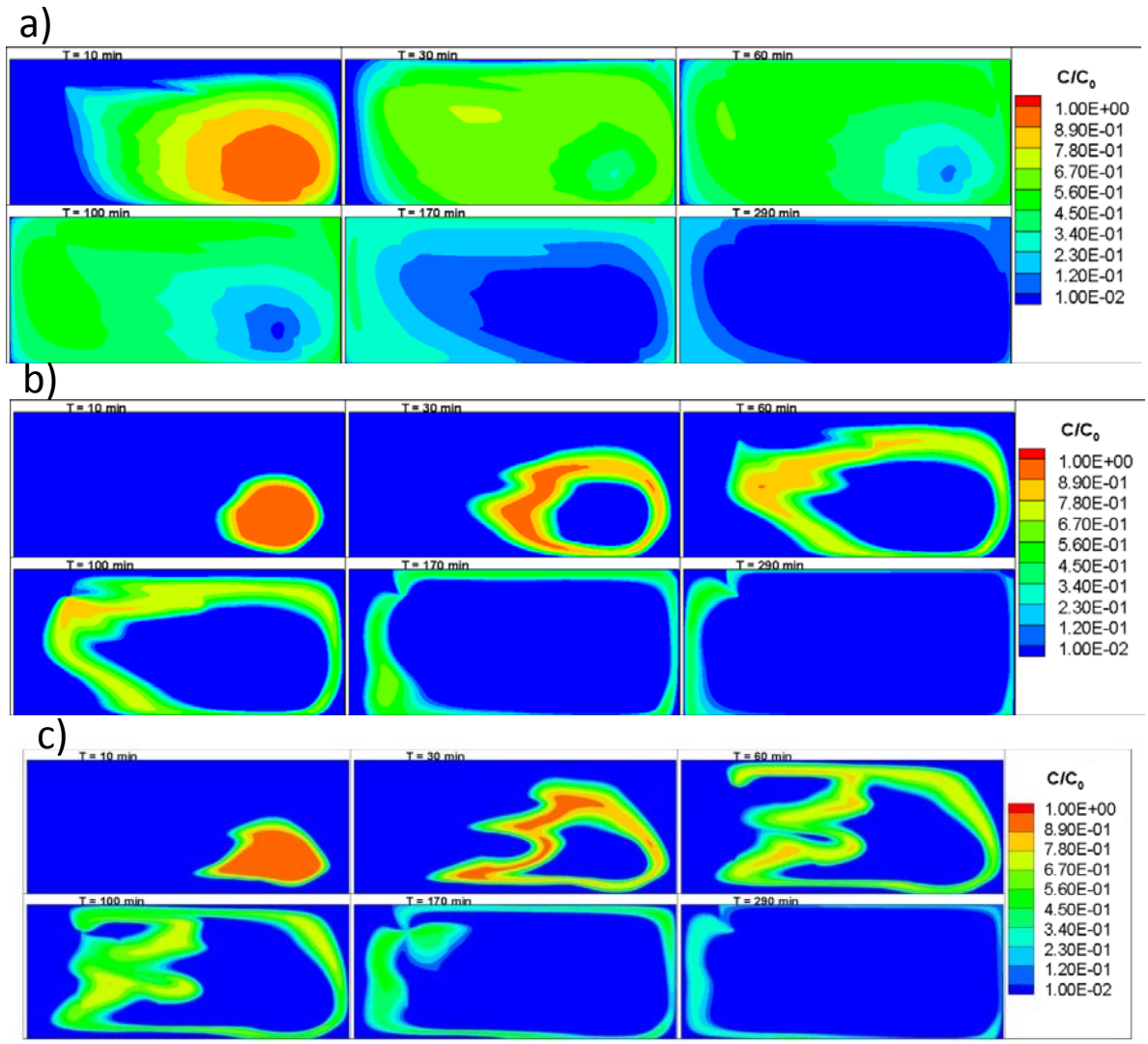


Figure 2.9. Concentration distributions from tracer transport simulation with: a) K_{eff} and A_{11} computed using *Gelhar and Axness* [1983] solution with statistics of permeameter analysis of core samples (case 1); b) kriged *K* field (case 2); and c) SSHT *K*-tomogram (case 3).

A direct comparison of the BTCs obtained from the numerical simulations and the sampled tracer data is shown in Figure 2.10. The matches of the BTCs obtained from ports (30, 35, and 47) close to the injection port are quite good for all cases. In particular, the arrival times, peak

concentrations, and time for tracer concentrations to reach background levels are quite consistent for all cases.

Differences in the quality of matches begin to emerge when the BTCs from the intermediate distance ports (10, 12, 22, 27, 37, and 39) are examined. Case 1 (effective parameters) consistently under-predicts the peak concentration at all ports in this category and shows an earlier arrival of tracers. In contrast, cases 2 (kriging) and 3 (*K*-tomogram) show a marked improvement in the quality of the fits with case 3 on average performing the best based on a visual comparison of the BTCs.

Examination of the ports furthest from the injection port (2, 14, 19) reveals mixed results. In general, case 3 performs better than cases 1 and 2 in terms of better predicting the peak concentrations and arrival times of tracers. However, it is noted that the matches are far from perfect.

At the extraction port (7), none of the three approaches yields a accurate prediction, however, case 3 captures the arrival of the first 2 peaks. While the magnitude of the peak concentrations are different, the arrival time of maximum concentration for the first two peaks are similar. Case 1, on the other hand, because it is based on the effective parameter approach, cannot capture this multi-peak behaviour. Likewise, case 2 with the smoother *K* field does not capture this behaviour.

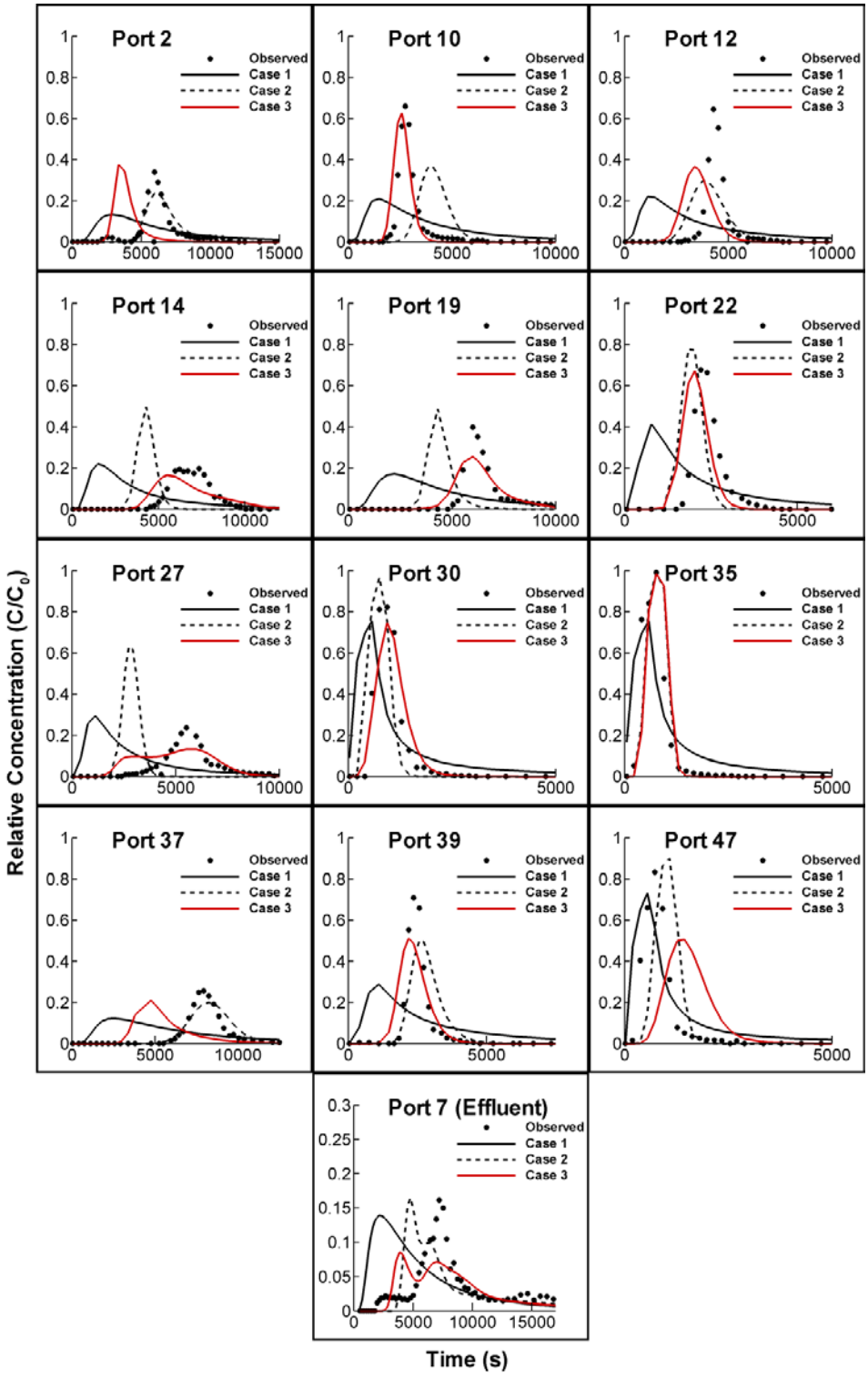


Figure 2.10. BTCs from the dipole tracer test and corresponding forward simulations.

The last comparison method to assess the relative performance of these cases for predicting the migration of the tracer plume was to perform a moment analysis on the BTCs to compare the total mass, center of mass with time, and the variance of the BTCs at each port for each case.

Figures 2.11 through 2.13 show the M_0 (total mass), μ (mean arrival time center of mass), and σ (variance of the breakthrough curve) computed from the temporal moment analyses of simulated and observed breakthrough curves. In particular, Figure 2.11 reveals the estimates of M_0 at these observation ports are, on average, higher for the simulated BTCs for case 1. Figures 2.11a and 2.11b indicate that when the heterogeneity in K is considered (i.e., Cases 2 and 3), the estimation of the total mass at a given sampling point improves. These results suggest that the heterogeneity estimated by kriging and HT is better able to capture the flow paths of the tracer reaching these sampling ports.

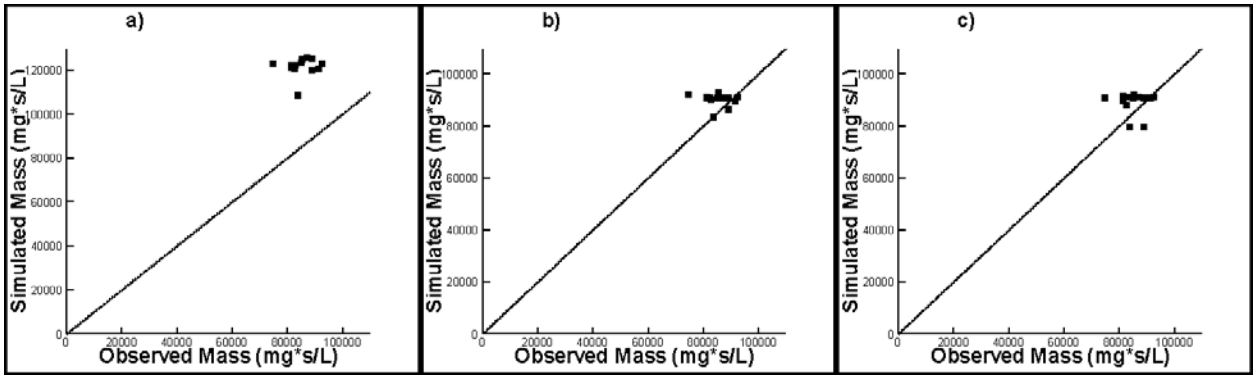
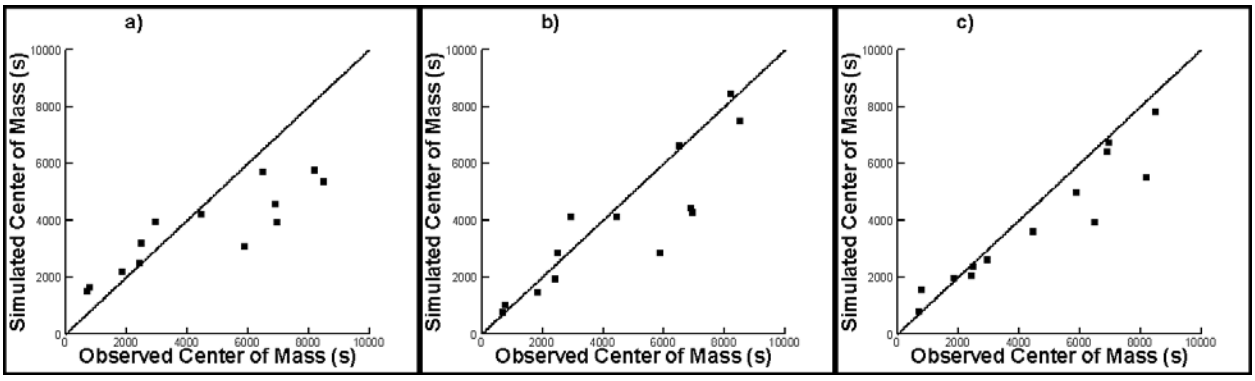


Figure 2.11. Total mass (M_0) computed by the zeroth moment for each sampling port: a) case 1; b) case 2; and c) case 3.

The arrival time of the center of mass (μ) for the 3 cases is plotted on Figure 2.12. Results of the three cases show that the arrival times of simulated breakthroughs for the ports close to the injection port are similar to the observed. At ports that are far away from the injection point, the discrepancy between the simulated and observed values increases. While the scatter is centered around the 1:1 line for all cases, case 3 shows considerably less scatter. There are two sampling points

(2 and 37) that cause the simulated versus observed relationship to be biased at later times for case 3. These two ports are at a large travel distance and near boundaries. If these two points are ignored, then the relationship between the simulated and observed values becomes quite similar. This suggests that the mapping of heterogeneity via hydraulic tomography yields more accurate estimates of the arrival times of the center of mass for this synthetic aquifer.



2.12. Mean arrival time of the center of mass (μ) computed for each sampling port: a) case 1, b) case 2, and c) case 3.

The variance of the BTCs (σ^2) was also calculated at each sampling point and plotted on Figure 2.13. The results suggest that the approach based on effective parameters over predicts the temporal spreading of the plume. The variance estimates at the extraction point (the data point with the largest variance value in Figure 2.13) for case 1 are, however, satisfactory, suggesting that an integrative behaviour (such as breakthrough at the extraction well) can be predicted by a low-spatial resolution approach. For case 2, the bias in the variance estimates diminishes at sampling intervals close to the injection well. In contrast, an improvement is seen in case 3 in which both the bias and scatter diminish. This suggests that hydraulic tomography can adequately map the heterogeneity, thus resulting in a higher-resolution prediction of tracer migration.

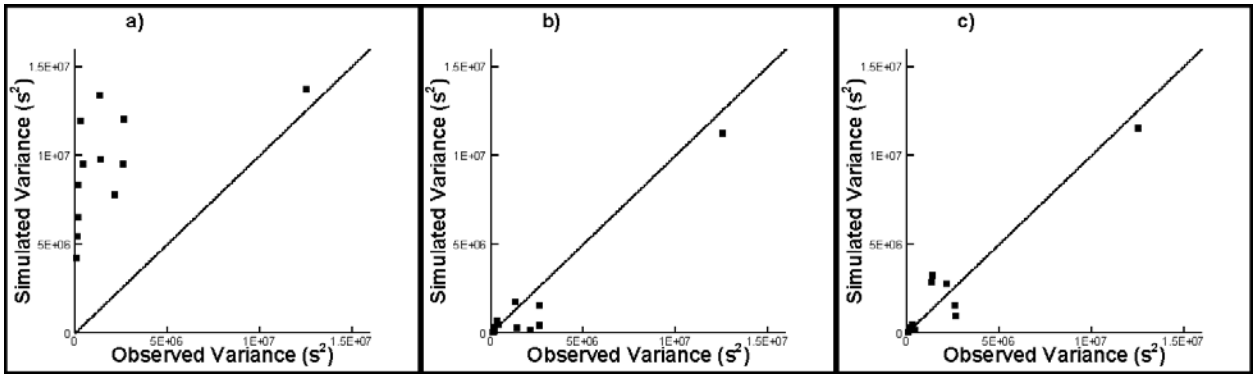


Figure 2.13. The variance of the breakthrough curve (σ^2) at each sampling port: a) case 1; b) case 2; and c) case 3.

Both qualitative and quantitative evaluations of the numerical tracer simulations showed that mapping detailed heterogeneity improved the predictions of the tracer migration, with the approach based on HT yielding a better result. Nevertheless, it must be emphasized that the *K* distribution obtained from hydraulic tomography was unable to capture all of the details of the tracer breakthrough. Some possible reasons for this are: 1) flow of the tracer along the boundaries where HT is unable to image due to an absence of observation ports; and 2) the assumption of a constant and uniform value of porosity.

This suggests that accurate predictions of solute transport may be possible if HT is used to map aquifer heterogeneity. With this approach, it is not necessary to collect large numbers of small scale samples to estimate effective parameters, or perform tracer tests for the purposes of estimating dispersivity, both of which are costly.

2.3 Flow to a Well in a Heterogeneous Unconfined Aquifer: Insights from an Intermediate Scale Sandbox

To date, a limited number of investigations have explicitly examined the effect of heterogeneity on pumping tests performed in unconfined aquifers [e.g., *Bunn et al*., 2010; *Mao et al*., *in press*]. In this study entitled "Flow to a well in a heterogeneous unconfined aquifer: Insights from an intermediate scale sandbox", by *Berg, S.J. and W. A. Illman* [*submitted manuscript*], hydraulic tomography is used to map the distribution of K and S_s in a fully saturated heterogeneous sandbox aquifer. These tomograms, and estimates of unsaturated parameters are then used to predict the response of the system to drainage. The manuscript is included as Appendix C.

The motivation for this study was the question "if the necessary parameters and forcing functions to populate a variably saturated flow model are known, is it possible to accurately model the responses to pumping in a heterogeneous unconfined aquifer in both the saturated and unsaturated zones?”

To answer this question, an intermediate-scale laboratory sandbox (see Figure 2.14) containing a heterogeneous aquifer was instrumented with pressure transducers, tensiometers, and water content sensors. The sandbox was 244 cm in length, 122 cm in height, and had a depth of 9.4 cm. While fully saturated, a number of pumping tests were performed for characterization purposes. These pumping tests were then used to estimate both homogeneous and heterogeneous K and S_s distributions. The homogeneous K and S_s estimates were obtained by coupling PEST [*Doherty*, 2005] with MMOC3 [*Yeh et al.*, 1993] and matching the simulated drawdowns to those observed during the pumping test at port 22. Heterogeneous K- and S_s-tomograms were estimated by analyzing 8 pumping tests using THT (see Figure 2.15).

Figure 2.14. Photograph of the sandbox showing all sensor locations (● = pressure transducers; ● without pressure transducers;○ = tensiometers; and x = water content sensors) and the various layers that were packed. Sensor locations are approximate. Numbers indicate what material type each layer is composed of.

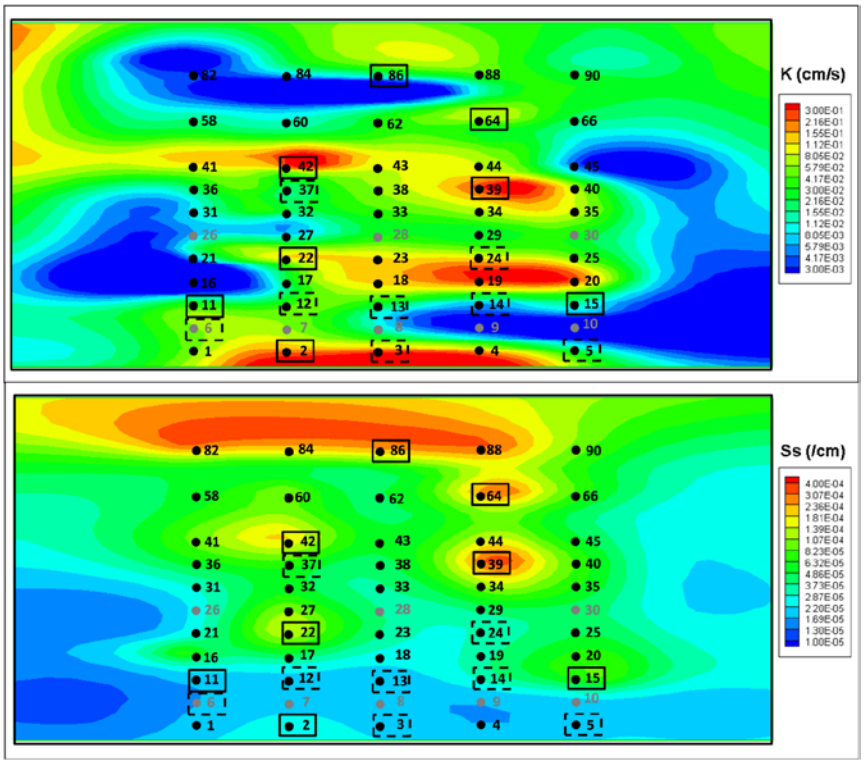


Figure 2.15. a) *K* and b) S_s tomograms computed using the transient hydraulic tomography algorithm of *Zhu and Yeh* [2005] SSLE with 8 pumping tests. Solid squares are pumping locations used for transient hydraulic tomography dashed squares are pumping locations for the validation of the results from transient hydraulic tomography.

Unsaturated parameters for the different material types in the sandbox were measured using the hanging water column method [*Stephens*, 1995]. The primary drainage curve was measured for five of the materials packed in the sandbox (F35, F45, F110, Sil-co-Sil 53, and Sil-co-Sil 106). Note F35-F110 are medium to fine sands, and the Sil-co-Sil material are various silt grades. These materials were selected because they were located in the upper portion of the sandbox and were likely to experience negative pressure heads (and as a result a possible release of water) when the water table was lowered during drainage or an unconfined pumping test. Figure 2.16 shows, the *van*

Genuchten [1980] match for the measured drainage curves and fitting parameters (α and n) for each material type (Figure 2.16 a-e).

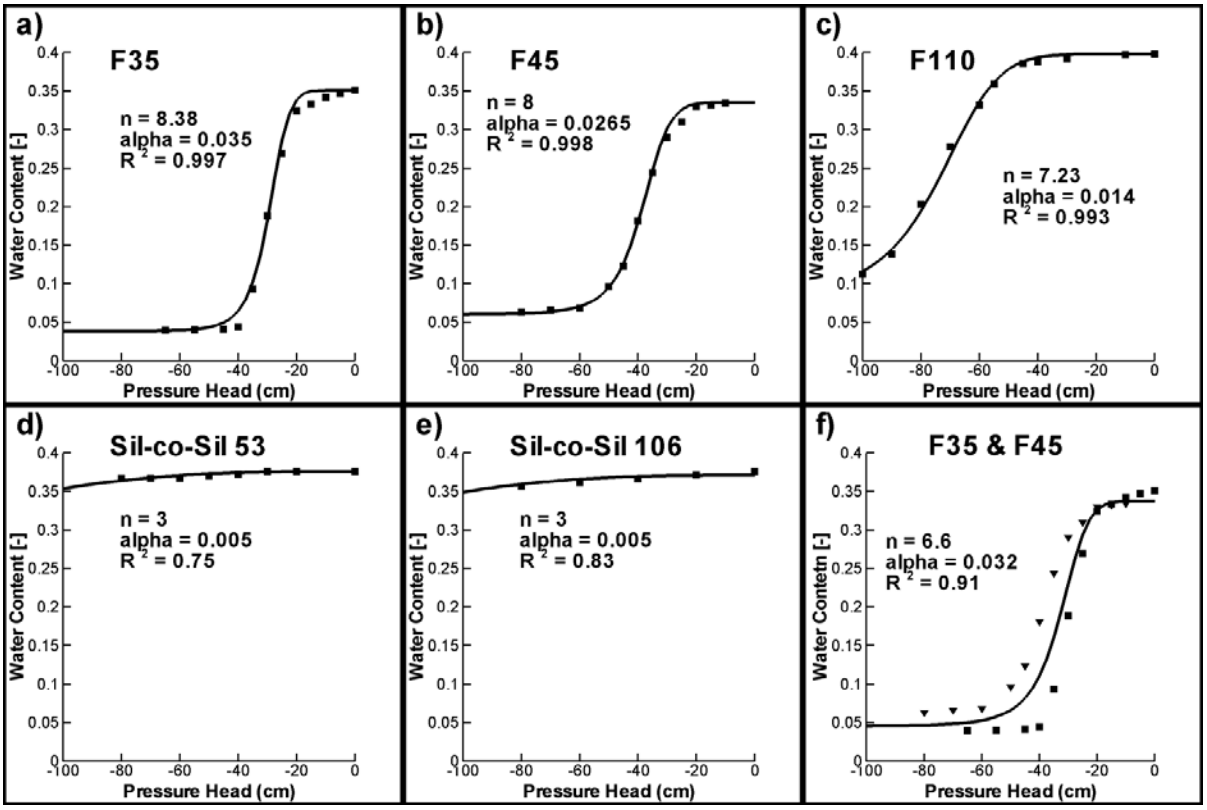


Figure 2.16. Moisture characteristic curves determined through the hanging column method of: a) F-35; b) F-45; c) F110; d) Sil-co-Sil 53; e) Sil-co-Sil 106; and f) F35 and F45 matched simultaneously. The squares represent F35 and gradient symbols represent F45.

To obtain an effective α and n representing the entire aquifer domain, the *van Genuchten* [1980] model was fit to the primary drainage curves for both F35 and F45 simultaneously (Figure 2.16f). These two sands were selected because they comprised the majority of the material in the unsaturated zone and were expected to control the drainage response. The other materials present in the unsaturated zone (F110, Sil-co-Sil 53, and Sil-co-Sil 106) were not included in this simultaneous match as they were only present in minor portions of the unsaturated zone and were present as

discontinuous lenses. Since they were present as discontinuous lenses, it is possible that if these coarser F35 and F45 materials were to drain, these lenses will become disconnected from the main aquifer and will be unable to effectively release their water. Thus, only the two dominant material types in the upper portion of the sandbox were selected for this effective parameter estimation.

Using these estimated parameter values, three representations of the aquifer sandbox were created: Case 1 - homogeneous saturated parameters (based on the PEST estimated K and S_s) and unsaturated parameters (based on Figure 2.16f); Case 2 - heterogeneous saturated parameters (based on THT) and homogeneous unsaturated (based on Figure 2.16f); and, Case 3 - heterogeneous saturated (based on THT) and unsaturated (based on Figures 2.16a-e) zone parameters.

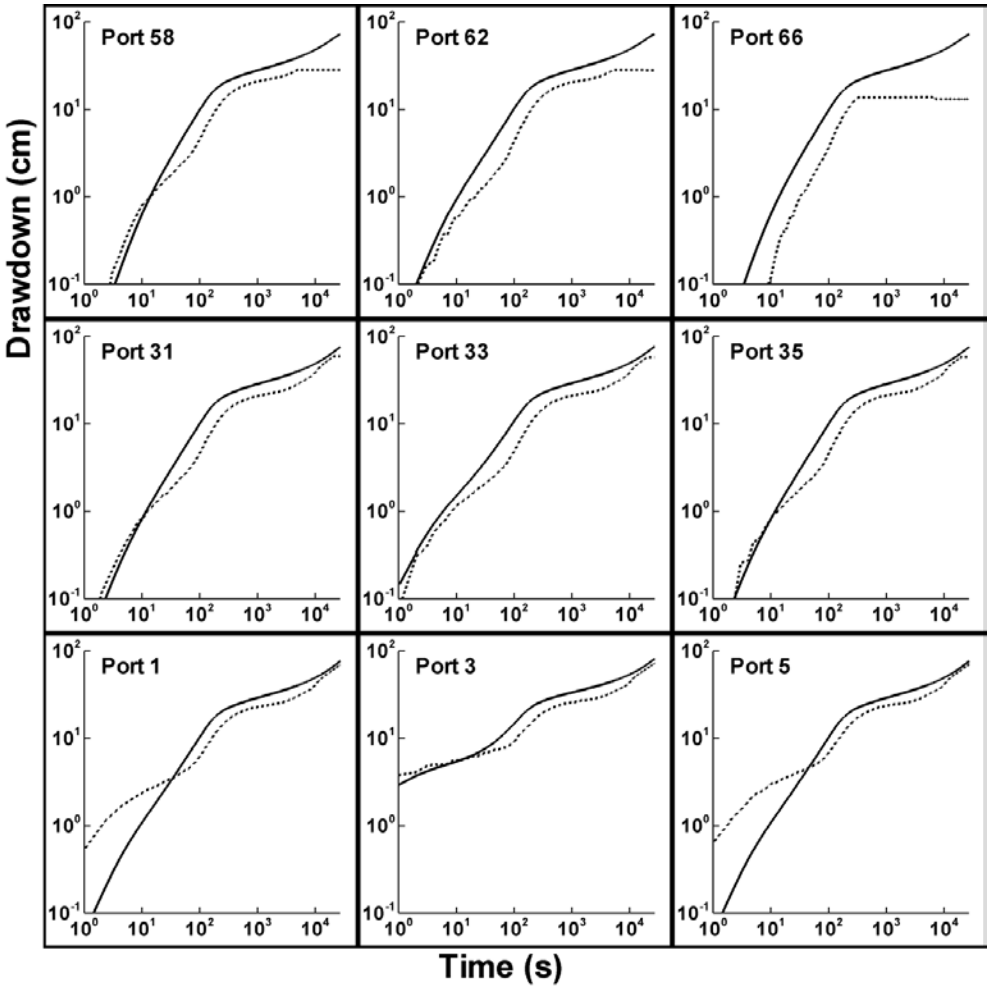
Following the characterization of the saturated (K and S_s) and unsaturated (saturated water content, residual water content, α, and n) parameters (summarized in Table 3), another pumping test was performed. For this test, the water table was allowed to fall while the drawdown in the sandbox was monitored at multiple locations in both the saturated (pressure transducers) and unsaturated zones (tensiometers equipped with pressure transducers and water content sensors). After this, forward simulations using MMOC3 [*Yeh et al*., 1993], a variably saturated flow and transport model were conducted to evaluate the ability of the different homogeneous and heterogeneous representations (cases 1-3) of the sandbox aquifer to reproduce the observed drawdown responses in both the saturated and unsaturated zones. By comparing these various cases, it is possible to determine which parameters have the greatest impacts on predicting the responses in the saturated and unsaturated zones.

Table 3. Unsaturated parameters for materials in the upper portion of the sandbox.

	θ_s	θ_r	α	n
F35	0.351	0.038	0.035	8.38
F45	0.335	0.06	0.0265	8
F110	0.398	0.08	0.014	7.23
Sil-co-Sil 53	0.376	0.08	0.005	3
Sil-co-Sil 106	0.371	0.08	0.005	3
F35 & F45	0.3375	0.045	0.032	6.6

Figures 2.17 and 2.18 show the simulated and observed drawdowns at selected sensors for Cases 1 and 3 respectively. Figures for case 2 are available in Appendix C (Note: the simulated results for cases 2 and 3 are very similar).

a)


Drawdown (cm)
Port 58
Port 62
Port 66
Port 31
Port 33
Port 35
Port 1
Port 3
Port 5
Time (s)

b)

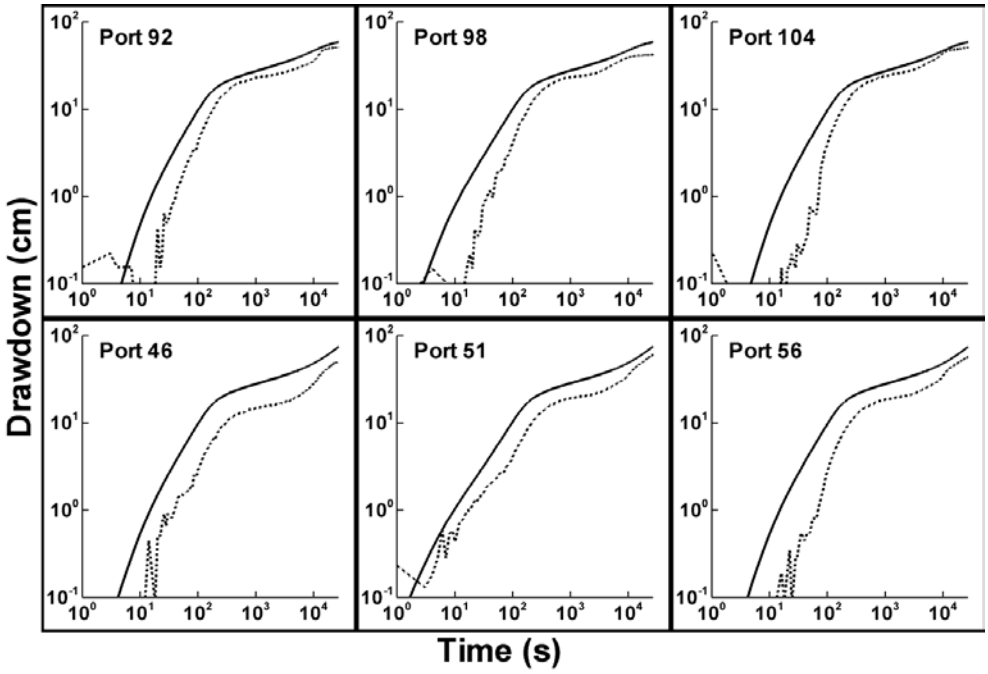


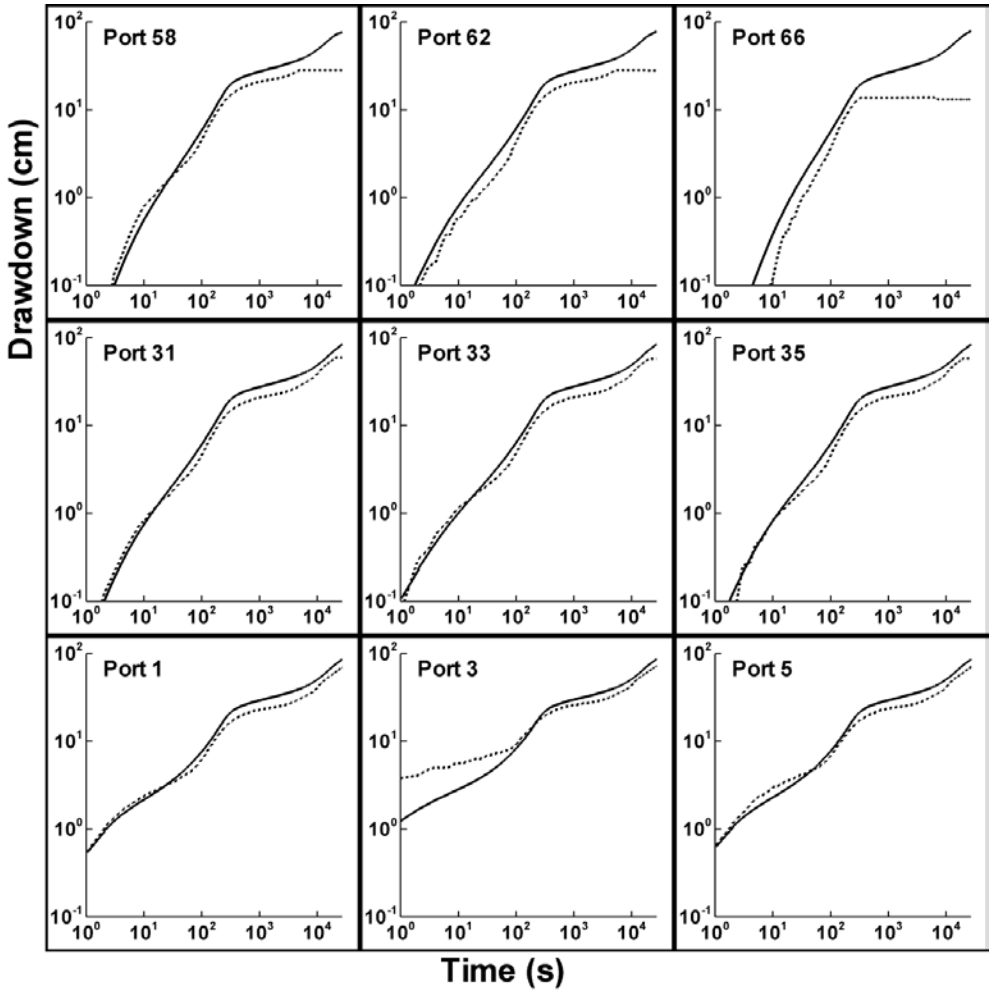
Figure 2.17. Simulated (solid line) and observed (dashed line) drawdown from selected: a) pressure transducers; and, b) tensiometers during the unconfined aquifer pumping test (Case 1). A comparison at all ports is available in Appendix C.

Figure 2.17 reveals that the comparison between the simulated and observed drawdowns in both the saturated (Figure 2.17a) and unsaturated (Figure 2.17b) zones were poor for Case 1. The simulated drawdown was generally greater than the observed, especially at intermediate and late time. Early time drawdown at pressure transducers were overestimated at some ports and underestimated at others. This is attributed to the homogeneous K and S_s field which captures the heterogeneity in an average sense, and thus, was not able to accurately simulate the response at all ports.

For case 2 (not shown here) K and S_s were heterogeneous and the unsaturated zone was identical to case 1. Forward simulation results showed a marked improvement in the prediction of drawdowns at all times with the variably saturated flow model. On the other hand, the drawdown

predictions in the unsaturated zone were slightly improved, but still not very good, with the simulated values generally exceeding the observed values, especially at early time.

a)



b)

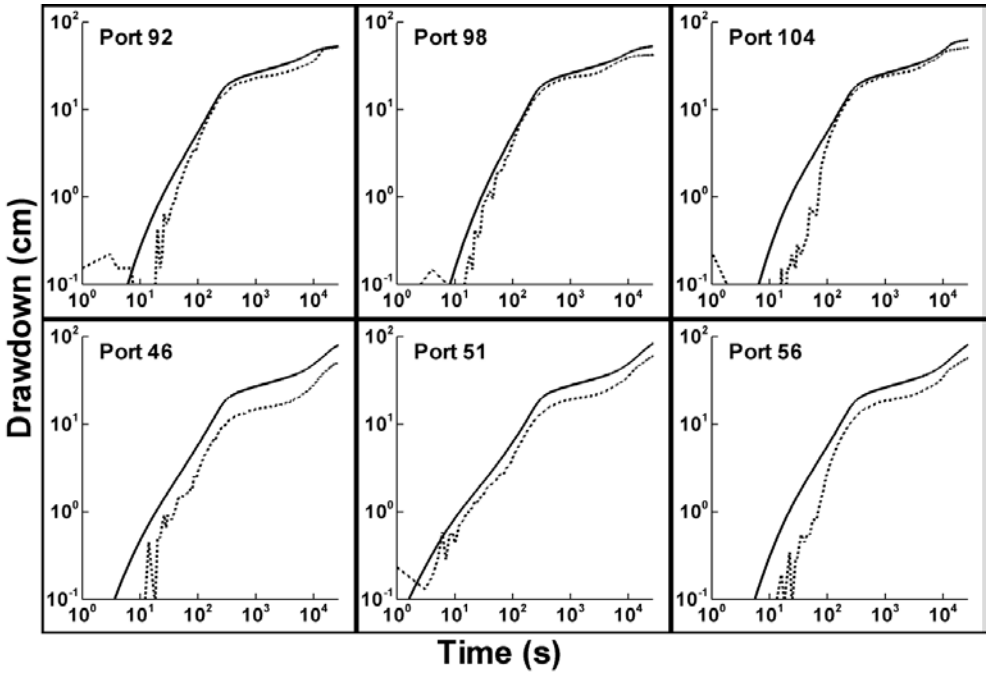


Figure 2.18 Simulated (solid line) and observed (dashed line) drawdown from selected: a) pressure transducers; and, b) tensiometers during the unconfined aquifer pumping test (Case 3). A comparison at all ports is available in Appendix C.

Case 3 considers the variability of the unsaturated zone parameters by assigning values for each material type. Figure 2.18a reveals that the quality of matches has virtually not changed in comparison to Case 2, perhaps suggesting that heterogeneity in unsaturated zone parameters may not be that critical to responses in the saturated zone. This observation is in agreement with the findings from a numerical study conducted by *Mao et al.* [*in press*] who concluded that drawdown is most sensitive to heterogeneity in the saturated parameters (K and S_s). Predictions of drawdown in the unsaturated zone show only a slight improvement, suggesting that as long as the K and S_s are accurately known and the unsaturated parameters are represented accurately (in an average sense), it is possible to predict the drawdowns induced by unconfined pumping tests using a variably saturated flow model.

In summary, these results show that mean or effective K and S_s values (Case 1) could not yield accurate predictions of drawdown responses in the saturated and unsaturated zones. Predictions of drawdown responses in the saturated zone improved dramatically when heterogeneous K and S_s distributions were used in the forward simulation of the unconfined aquifer pumping test (cases 2 and 3). Because the pressure changes in the unsaturated zone were monitored during this experiment, attempts were also made to evaluate the predictions of pressure responses in the unsaturated zone. Results showed that slightly improved predictions of drawdowns in the unsaturated zone could be achieved if heterogeneous zones were assigned to the model. However, the predictions of drawdowns in the saturated zone were relatively insensitive to the unsaturated zone, whether it was represented as homogeneous or heterogeneous supporting the conclusion reached by *Mao et al*. [*in press*] using numerical simulations.

Based on this study it is possible to accurately predict the response of a heterogeneous unconfined aquifer to pumping as long as the saturated parameters (K and S_s) are accurately characterized and accurate effective values of the unsaturated parameters are known. This study was unique in that the heterogeneity pattern (for the unsaturated zone parameters) was exactly known and it was possible to select effective unsaturated parameters by assessing which material would dominate the drainage response. However, in the field, this is not as straight forward and the selection of unsaturated parameters could pose a challenge. Thus, interpreting pumping tests in heterogeneous aquifers will benefit from inverse modeling of both the unsaturated and saturated zones.

2.4 Hydraulic/Partitioning Tracer Tomography for DNAPL Source Zone Characterization: Small-Scale Sandbox Experiments

The work presented in this manuscript uses the pumping tests and partitioning tracer tests conducted by *Massi* [2007] at the IIHR-Hydroscience & Engineering of the University of Iowa to estimate: i) the heterogeneous distribution of *K* using SSHT, and ii) the distribution of trichloroethene (TCE) spilled in a laboratory sandbox using HPTT. The results of this study are published as "*Hydraulic/Partitioning Tracer Tomography for DNAPL Source Zone Characterization: Small-Scale Sandbox Experiments"* by *Illman, W.A., S.J. Berg, X. Liu, and A. Massi* [2010b] in *Environmental Science and Technology*. This manuscript is provided in Appendix D. The authors contribution to this study was the interpretation and inverse modeling of data collected by *Massi* [2007].

Hydraulic/Partitioning Tracer Tomography (HPTT) was developed to characterize DNAPL source zones [*Yeh and Zhu*, 2007] by fusing data from hydraulic tomography (HT) and partitioning tracer tomography (PTT). PTT is similar in concept to HT, but the source is a partitioning tracer that is introduced at an injection point and the signals are the corresponding BTCs obtained at multiple sampling points. Additional partitioning tracer tests are then conducted sequentially by injecting tracers at other locations and monitoring the corresponding BTCs at a number of locations. Because the tracers sweep through the DNAPL source zone from several directions, PTT yields additional information on the distribution of NAPL saturation (S_N) that cannot be obtained from a single partitioning tracer test, or through its interpretation using a temporal moment method. The goal of this study is to validate the HPTT method of *Zhu and Yeh* [2007] through the tomographic analysis of the four partitioning tracer tests performed by *Massi* [2007].

Massi [2007] spilled 125 mL of TCE dyed with Sudan IV in a heterogeneous laboratory sandbox aquifer. Prior to spilling the TCE, the sandbox aquifer was characterized using 9 pumping tests that were analyzed using SSHT. The estimated *K*-tomogram is presented as Figure 2.19 and the distribution of TCE within the sandbox is presented as Figure 2.20.

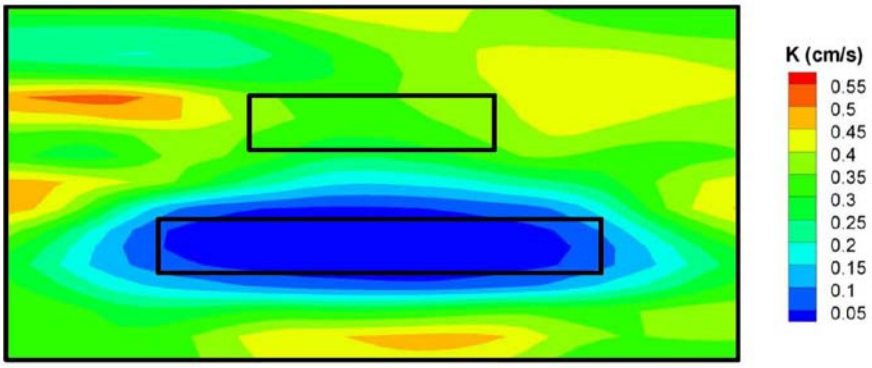


Figure 2.19. *K* tomogram generated using drawdown data from 9 dipole tests [after *Illman et al.*, 2010b].

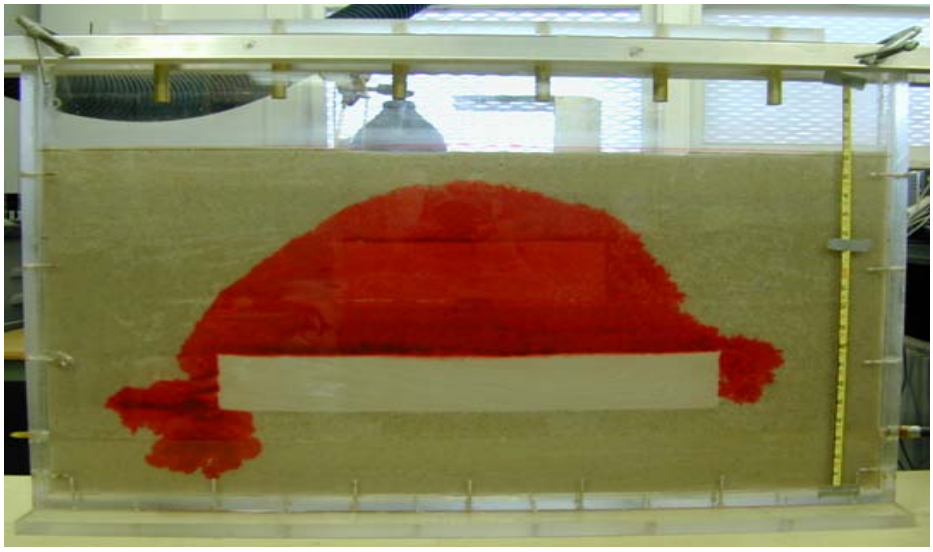


Figure 2.20. Photograph of TCE distribution within the sandbox [after *Massi,* 2007].

Upon settlement of the TCE in the source area, *Massi* [2007] conducted 4 partitioning tracer tests. For each tracer test, a dipole flow field was created by injecting tap water at one port, extracting at another port, and sampling at 12 ports in between. Once steady state flow conditions were achieved, a solution containing a partitioning tracer and a conservative tracer was injected for 10 minutes. This was followed by approximately 2 hours of tap water. Over the duration of the experiment, both the conservative and partitioning tracers were sampled for at each of the 12 sampling ports. Figure 2.21 a-d shows the injection, extraction, and sampling locations for each partitioning tracer test.

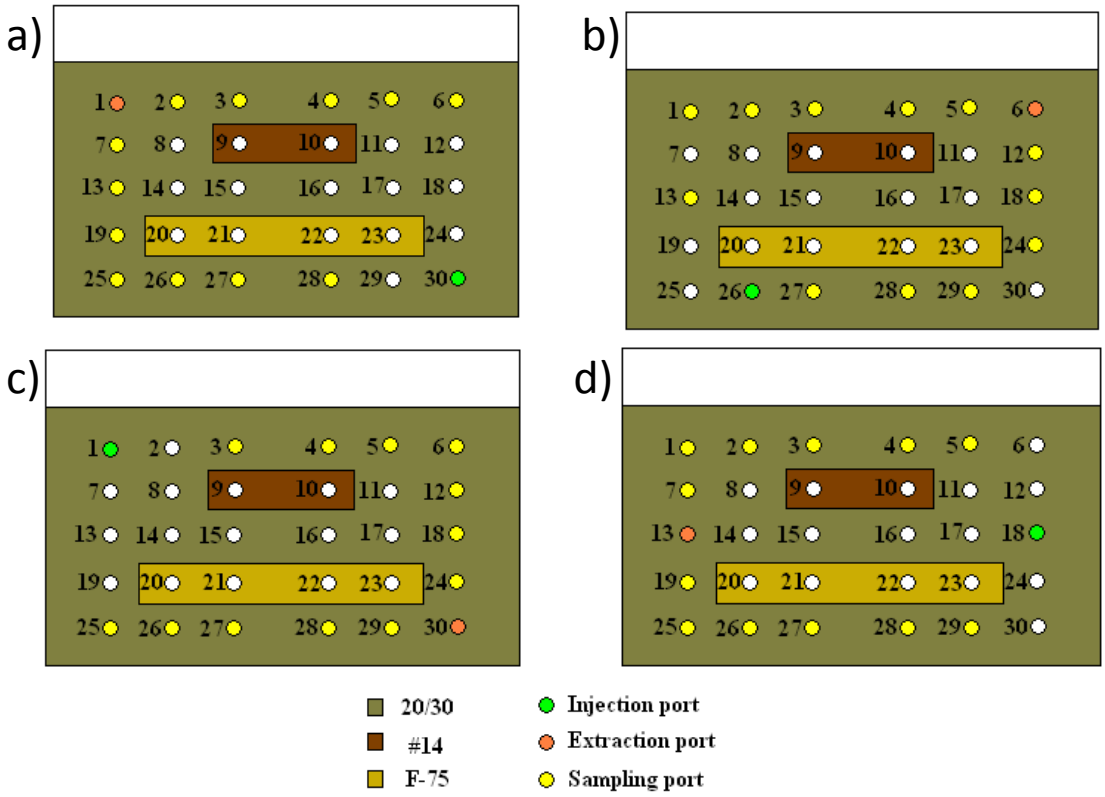


Figure 2.21. Injection, extraction, and sampling locations during partitioning tracer test: a) no. 1; b) no. 2; c) no. 3; and d) no. 4 [after *Massi*, 2007].

Next, the inverse procedure of *Yeh and Zhu* [2007] was used to estimate the distribution of S_N (S_N-tomogram) using the *K*-tomogram (Figure 2.19) as the *K* field for the tracer inversion. Figure 2.22 a-f shows results of the inversion by analyzing the partitioning tracer tests in several different ways. In particular, Figures 2.22 a–d shows inversions of individual partitioning tracer tests 1 through 4, respectively. These results show that with a single partitioning tracer test, the existence of TCE in the source area is confirmed although the estimated S_N distribution is not very realistic compared to the photograph of the TCE distribution. This suggests that perhaps one tracer test with a limited number of sampling ports (Test 1: Ports 2, 3, 6, 13, 25, 26, 27, 28; test 2 - Ports 4, 5, 12, 13, 18, 24, 27, 28, 29; test 3 - Ports 3, 4, 5, 6, 12, 24, 25, 26, 27; and test 4 - Ports 1, 5, 25, 26, 27, 28, 29) may not yield

accurate S_N distributions. Interestingly, each S_N tomogram shows a different S_N distribution and mean S_N (see Table 4) suggesting that the inversion of a single partitioning tracer test can yield different mean S_N values. For example, the inversion of test 1 alone yields a mean S_N of 3.7% while test 4 yields 1.9%. Recalling that 125 ml of TCE was injected into the sandbox, the mean S_N of the sandbox is 5%. A possible explanation for these differences in the mean S_N is the fact that each tracer test samples a different part of the aquifer. Next, the S_N tomogram is estimated by looping through the 4 tests sequentially [*Zhu and Yeh*, 2005] (Figure 2.22e for 1 loop and Figure 2.22f for two loops). This sequential analysis (and looping) improves the location of the estimated NAPL, indicating that while the analysis of each individual tracer test had difficulties identifying the location of the NAPL, the analysis of multiple tests may result in an improved estimation of NAPL location.

Table 4. Mean S_N and volume estimated through the inversion of various partitioning tracer tests using the *K* tomogram from steady state hydraulic tomography and a known *K* distribution.

	K tomogram from SSHT		Known *K* distribution	
Test	Mean TCE saturation	V_{TCE} (mL)	Mean TCE saturation	V_{TCE} (mL)
1	0.0372	98.10	0.0277	71.99
2	0.0496	128.55	0.0476	126.09
3	0.0317	81.39	0.0309	81.09
4	0.0189	46.53	0.0202	51.48
1-4	0.0218	55.46	0.0491	128.73
1-4 (L)	0.0473	130.06	0.0593	156.94
4-1	0.0360	97.09	0.0521	139.60
4-1 (L)	0.0475	128.12	0.0699	189.58

Despite the encouraging results, it is natural for one to question whether the S_N distribution can be further improved if the *K* heterogeneity were better represented. For the next case (Figure 2.23), values of *K* were deterministically assigned to the 3 sand types based on *K* measurements from permeameter tests. This *K* field was then used for the inversion of the partitioning tracer tests. Since the *K*-tomogram computed earlier was based on geostatistical methods, the variability in the heterogeneity is relatively smooth with a limited number of pressure measurement points. Assigning *K* values as was done here is tantamount to knowing precisely the location of the stratification and its morphology as well as the locations of sharp boundaries between zones of different *K* values.

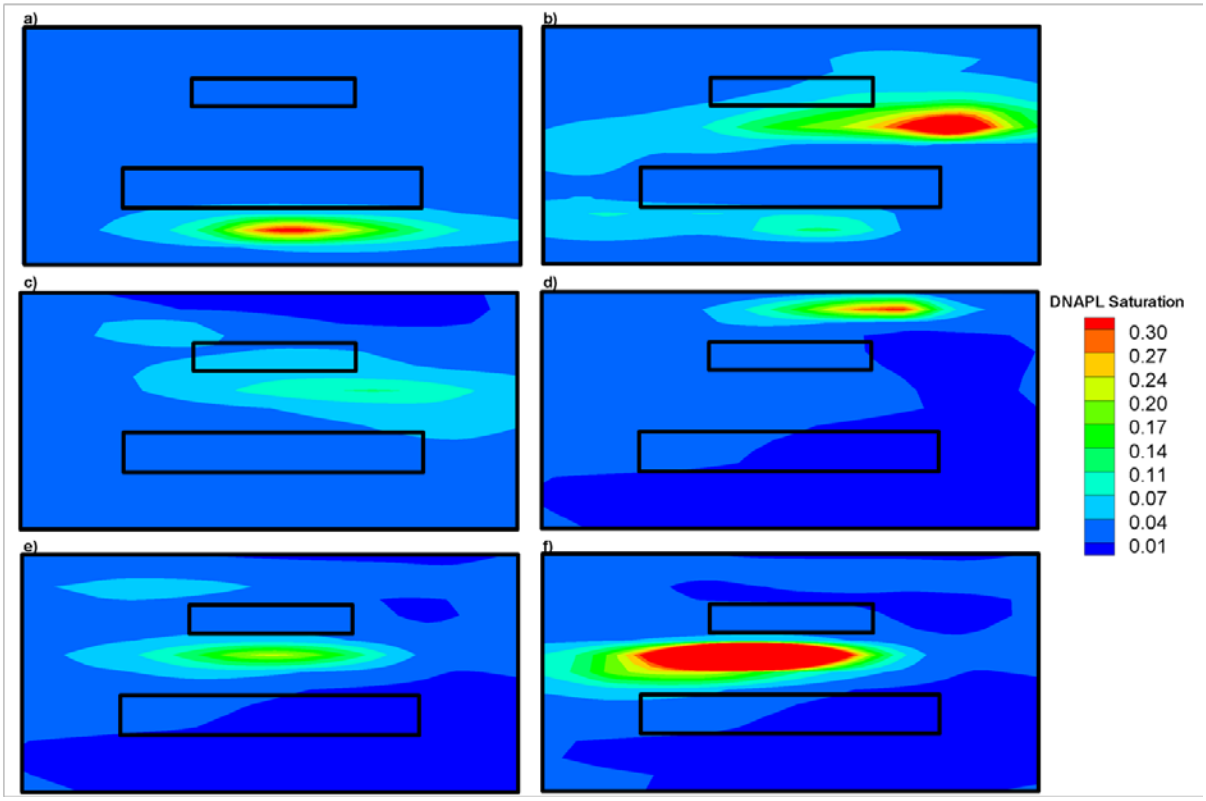


Figure 2.22. Estimated TCE content tomogram with inversion of: a) test 1 only; b) test 2 only; c) test 3 only; d) test 4 only; e) tests 1 - 4; and f) tests 1 - 4 with loop iteration. In this case, the *K*-tomogram estimated using steady state hydraulic tomography (Figure 2.19) was used for partitioning tracer tomography [after *Illman et al.*, 2010b].

Figures 2.23a through 2.23f are the estimated S_N-tomograms for the various cases that were examined before. As in Figure 2.22a-d, the inversion of individual partitioning tracer tests shows that the mean S_N estimates were not that much better than the case when the K-tomogram was used. However, as Figures 2.23e and 2.23f show the position of the high S_N zone is considerably better constrained when the positions of the strata are known. This suggests that improved knowledge of K heterogeneity and in particular, precise knowledge of boundaries between strata can be critical in imaging S_N in heterogeneous aquifers. Undoubtedly, such accurate knowledge of K heterogeneity is difficult to obtain. While this study shows the effectiveness of HPTT to image K and DNAPL S_N in a 2D sandbox, there is still much work required to upscale this approach for field-scale application.

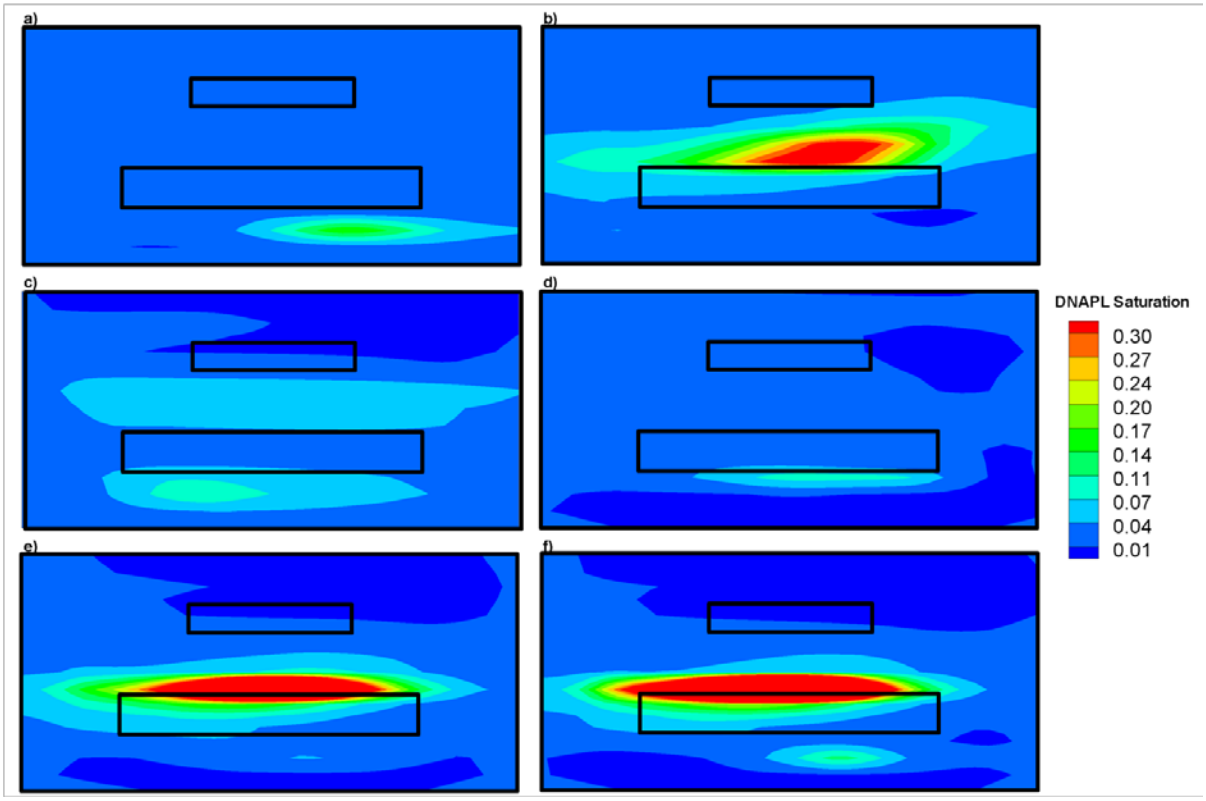


Figure 2.23. Estimated TCE saturation tomogram with inversion of: a) test 1 only; b) test 2 only; c) test 3 only; d) test 4 only; e) tests 1 - 4; and f) tests 1 - 4 with loop iteration. In this case, the K values of the 3 sand types were obtained from permeameter measurements and deterministically assigned for partitioning tracer tomography [after *Illman et al.*, 2010b].

.

Chapter 3

Field Studies of Hydraulic Tomography

3.1 North Campus Research Site

3.1.1 Site Description

The site used for all of the field studies in this thesis was the North Campus Research Site (NCRS) located at the University of Waterloo (UW) in Waterloo, Ontario, Canada. The surficial geology of the local area is glaciofluvial in origin and is highly heterogeneous. The field site is located within the Waterloo Moraine, which is an interlobate feature composed of kettle and kame deposits within alternating layers of till and glaciofluvial material [*Karrow*, 1993]. Site specific geology has been described by *Sebol* [2000] and *Alexander et al.* [2011]. The main feature of the site is an 'aquifer zone' located approximately 8 to 13 m below ground surface (mbgs). This zone consists of two high *K* units that are separated by a discontinuous low *K* layer. The upper aquifer is composed of sand to sandy silt, and the lower aquifer is composed of sandy gravel. The low *K* unit separating the two aquifers is discontinuous and provides a hydraulic connection [*Alexander et al.*, 2011]. In addition, despite being interpreted as layers, none of the units are continuous across the entire study site. Situated above and below the aquifer zone are low *K* silts and clays. At approximately 18 mbgs is the dense Catfish Creek Till which acts as a hydraulic barrier [*Alexander et al.*, 2011] and is taken to represent the bottom boundary of this study.

The aquifer system in the Waterloo Moraine is generally confined by a laterally extensive upper aquitard layer, however, this aquitard is known to contain stratigraphic windows in some areas [*Martin and Frind*, 1998]. Based on previous pumping tests performed at the site [*Alexander et al.*, 2011], the aquifer at the NCRS behaves as a confined to semi-confined system. None of the

drawdown responses observed during these previous pumping tests suggest that the main aquifer zone behaves in an unconfined manner, which might indicate the presence of a stratigraphic window. Water levels collected in the vicinity of the site indicate that groundwater flow is towards the south-east.

Alexander et al. [2011] also performed a detailed description of soil cores collected during the installation of 5 wells at the NCRS. In addition, they performed 471 permeameter tests, and 270 grain size analyses to produce a detailed *K* profile along these boreholes. Furthermore, slug tests and pumping tests were performed to estimate *K* along these boreholes. Figure 3.1 shows *K* estimates along 5 boreholes using these various approaches. This figure highlights the considerable heterogeneity at the site, with large changes in *K* over very short vertical distances. The estimated vertical correlation length was 15 cm, and *K* ranged over 5 orders of magnitude [*Alexander*, 2009; *Alexander et al.*, 2011].

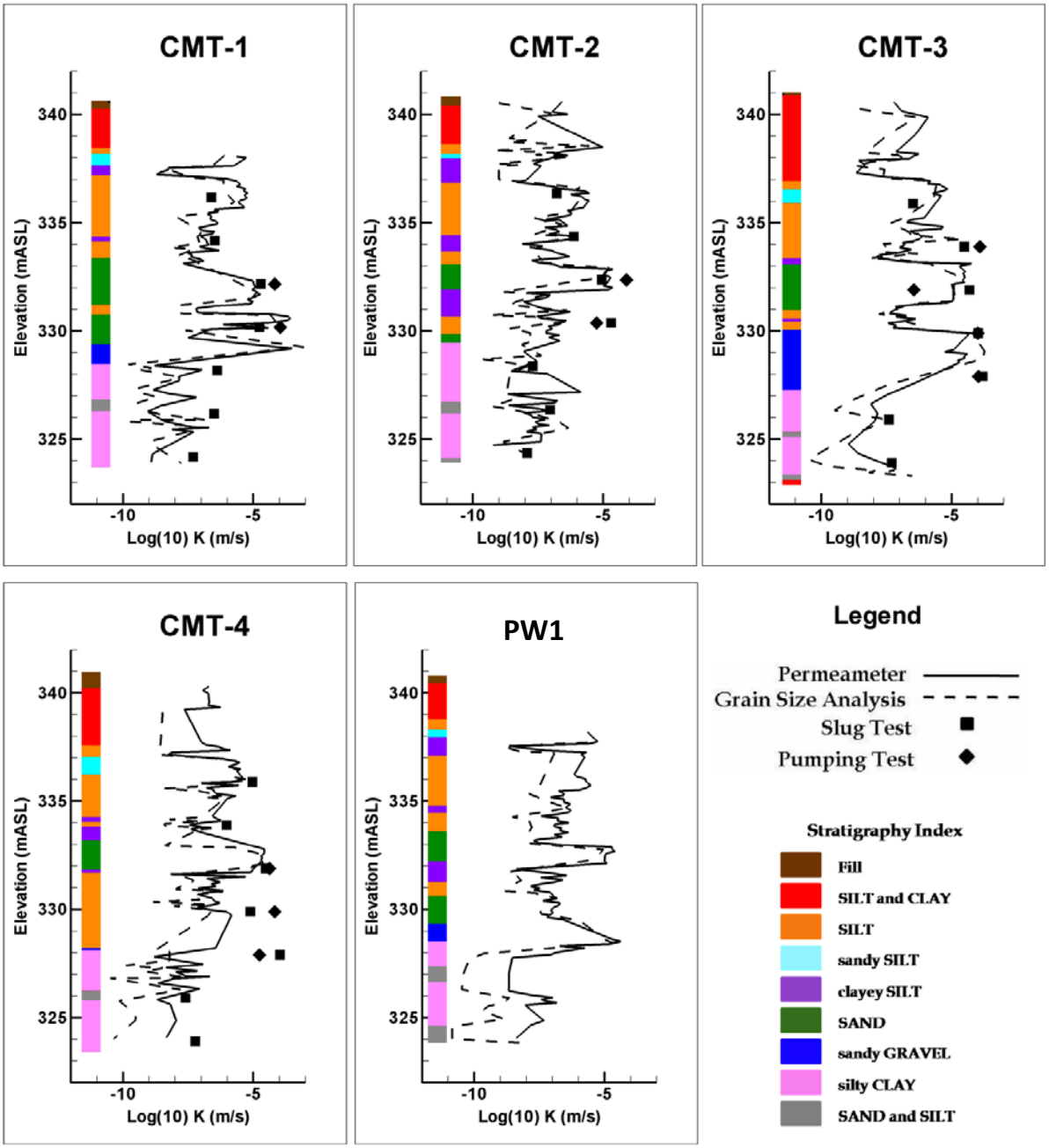


Figure 3.1. *K* estimates along 5 boreholes at the NCRS [after *Alexander et al.*, 2011].

3.1.2 Well Installation and Design

The NCRS was instrumented with a total of 9 wells in a square pattern (1 at each corner, 1 at the center of each face, and 1 in the center) measuring 15 m × 15 m. Figure 3.2 is a schematic layout of the wells. Continuous multichannel tubing (CMT) wells (*Solinst Canada Ltd.*), containing 7 channels each (7 screened intervals), were used strictly as observation wells and were installed in between the 4 corners of the square pattern (Figure 3.2). The screened intervals of the CMT wells were spaced 2 m apart with the upper screens (ports) located between 4.5 to 5.5 mbgs, and the deepest ports 16.5 to 17.5 mbgs. The remaining 5 wells were pumping wells, 3 of which were multi-screen wells (PW1 contained 8 screens, PW3 and PW5 contained 5 screens). These multi-screen wells were 10 cm in diameter, and contained screens 1 m in length. Each screen was separated from adjacent screens by 1 m of solid pipe. PW1 extended to approximately 18 mbgs; and, PW3 and PW5 to 12 mbgs. PW2 and PW4 were well nests consisting of 3 separate wells each (5 cm diameter). Each well in the nest had one screen that was 1 m long. Screen elevations for PW2 were 4, 7, and 8 mbgs; and, screen elevations for PW4 were 5, 8.5, and 11.5 mbgs.

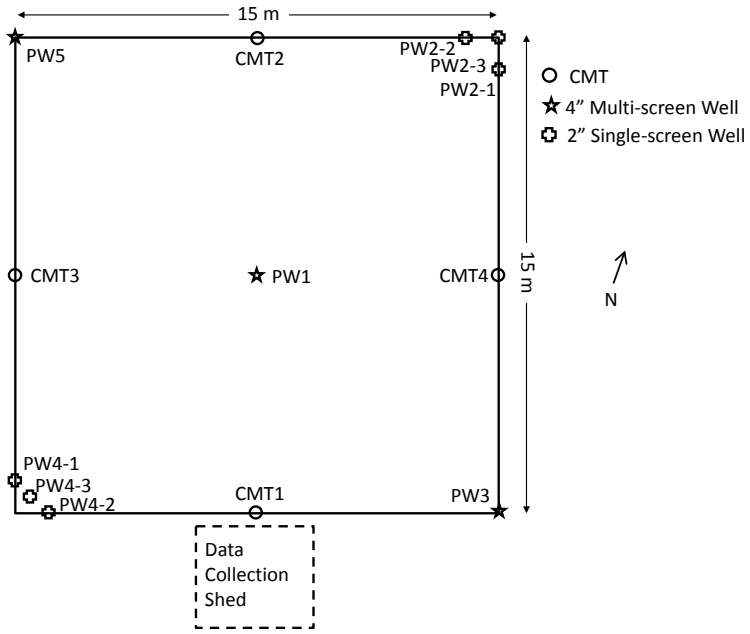


Figure 3.2. Two-dimensional plan view showing well locations.

Since these wells were installed for the purposes of HT, it was important that vertically adjacent screens be hydraulically isolated. Thus, the sand packs around the well screens were separated from each other by a layer of bentonite. The exceptions to this were the well nests (PW2 and PW4) which were installed in direct contact with the native formation. Figure 3.3 shows the three-dimensional perspective view of the wells, the various screens as well as bentonite seal elevations at the NCRS.

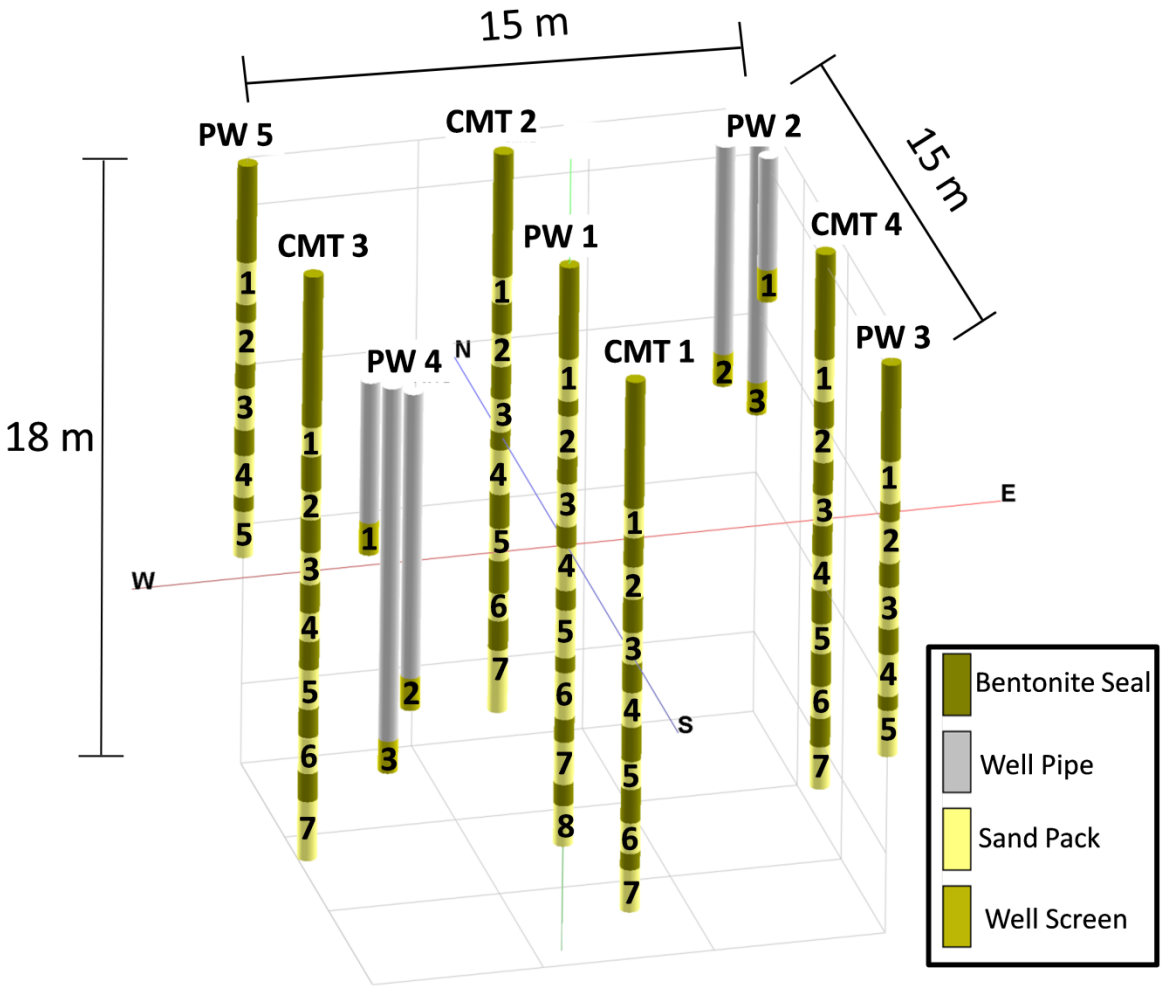


Figure 3.3 Three-dimensional perspective view of various wells and corresponding intervals at the NCRS.

3.1.3 Sensors and Data Collection

A network of pressure transducers was installed at the NCRS to allow for data collection during pumping tests. Depending on the particular pumping tests, up to 44 observation ports were monitored. The CMT ports were monitored with 0 - 15 psig (model MP100: *Micron Systems*) pressure transducers. Since the center channel of the CMT wells was too small to allow for both a pressure

transducer and a water level tape (for verification purposes), these central ports were monitored manually with a water level tape. These ports were found to respond very slowly, thus, it was still possible to obtain a complete record for these points. The well nests (PW2 and PW4) were monitored with 0 - 5 or 0 - 10 psig pressure transducers (model 3001 LT Leveloggers Junior, *Solinst Canada Ltd.*).

If one of the multi-screen wells (PW1, PW3, or PW5) was not being used for pumping, a FLUTe (*FLUTe Ltd.*) liner was installed in the well to prevent hydraulic short circuiting between adjacent screens within the well. Two of these liners contained 5 vented pressure transducers (Level Troll, *In Situ Inc.*), at elevations that corresponded with the screens of the PW wells. PW1, PW3, and PW5 were similar enough in construction to allow for the liners to be moved between wells depending on the pumping configuration. When PW1, PW3, or PW5 were used for pumping, the FLUTe systems were installed in the 2 unpumped wells. When PW2 or PW4 was pumped, the FLUTe systems were installed in PW3 and PW5, and a blank FLUTe liner (no transducers) was installed in PW1.

For some data sets, particularly those with little drawdown, sensor noise was evident in the recorded pressure signal. For these cases a 10 point centrally weighted moving average was applied to smooth the signal. Since the data was collected at a high frequency (~ 4 Hz at early time, 1 Hz at late time) the application of this filter did not change the shape of the drawdown curve.

3.1.4 Pumping Test Description

A total of 9 pumping tests were performed in the well network at the NCRS. These tests ranged in duration from 6.5 hrs to 22 hrs. Table 5 provides a summary of the pumping tests. For the multi-screen wells (PW1, PW3, and PW5) pumping was performed with a submersible pump (Model

SQE05, *Grundfos Canada Inc.*) located between two packers that were inflated to isolate the target screen. Pumping in PW4 was performed using a surface lift pump.

The locations for the pumping tests were selected for their ability to stress the aquifer system (i.e. flow rate), as well as their spatial coverage of the site. The pumped locations were; PW1-3, PW1-4, PW1-5, PW3-3, PW3-4, PW4-3, PW5-3, PW5-4 and PW5-5. PW4-3 had the highest flow rate at 30.2 L/min, followed by PW1-3 at 10.5 L/min. PW3-3 and PW3-4 had the lowest flow rates at 2.1 and 1.5 L/min respectively. The flow rates for the 5 remaining pumping tests were between 4.4 and 8.8 L/min.

Table 5. Summary of pumping tests performed at the NCRS.

Pumped Location	Flow Rate (L/min)	Duration (hrs)	Maximum Observed Drawdown (m)	Port of Maximum Drawdown
PW1-3	10.5	6	1.10	CMT 4-3
PW1-4	6.3	8.5	0.98	CMT 2-4
PW1-5	4.4	22.5	0.18	CMT 2-4
PW3-3	2.1	22	0.57	CMT 1-3
PW3-4	1.5	22	0.08	CMT 1-5
PW4-3	30.2	22.5	0.88	CMT 3-4
PW5-3	7.8	22	0.92	CMT 2-3
PW5-4	7.8	8.5	0.45	CMT2-4
PW5-5	8.1	22	0.32	CMT2-3

For most of the pumping tests, the pumped interval was hydraulically isolated, as indicated by pressure transducer readings from above and below the packed interval. Two exceptions to this were PW1-5 and PW5-5, both of which experienced pronounced drawdown above the packed-off interval. For these tests, the packers were operating properly, thus the short circuiting occurred either through the bentonite seal between the sand packs, or directly through the aquifer.

Figures 3.4 and 3.5 show the drawdown response recorded at the CMT ports and the PW wells during the pumping test at PW1-3. Solid lines represent transducer recorded data, and the dots are water levels manually recorded with a water level tape. During this test, the drawdown response was most pronounced at ports CMT2-3, CMT3-3, and CMT4-3. The deepest ports (ports 6 and 7 in each CMT well) generally experienced very little to no drawdown.

In some cases, negative drawdown was observed. This response, known as the Noordbergum effect was observed in deep ports during some of the pumping tests at the NCRS and was attributed to deformation-induced head change. The findings of *Berg et al.* [2011] indicate that when this behaviour manifests in pumping test data, it is still possible to estimate hydraulic parameters if late time data (after the water level draws down below static) is analyzed (this effect is discussed in greater detail in section 3.2). Thus, late time drawdown data for these ports is included if available, if not, the negative drawdown is treated as zero drawdown. Similar plots for the remaining 8 pumping tests are provided in Appendix E.

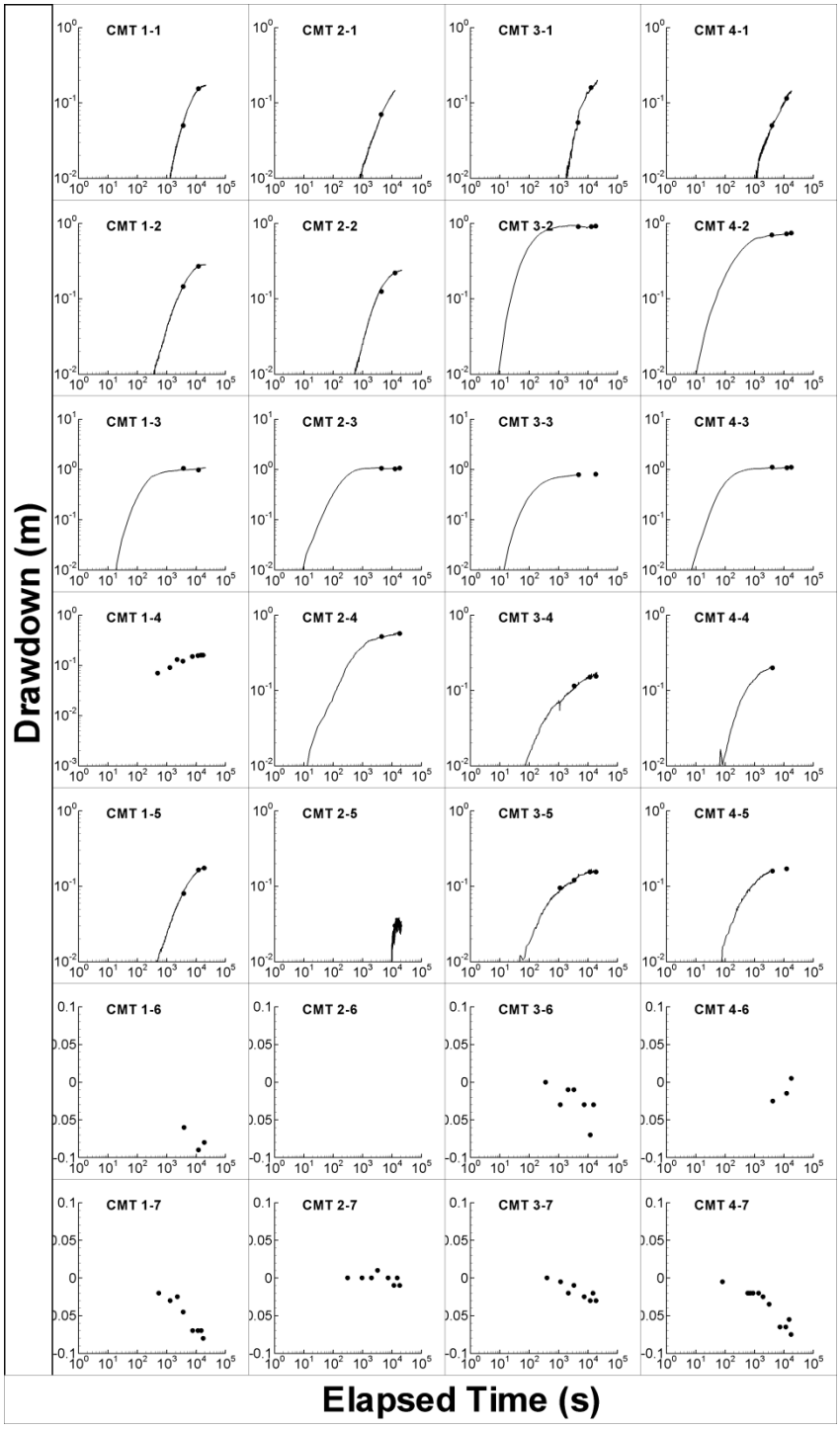
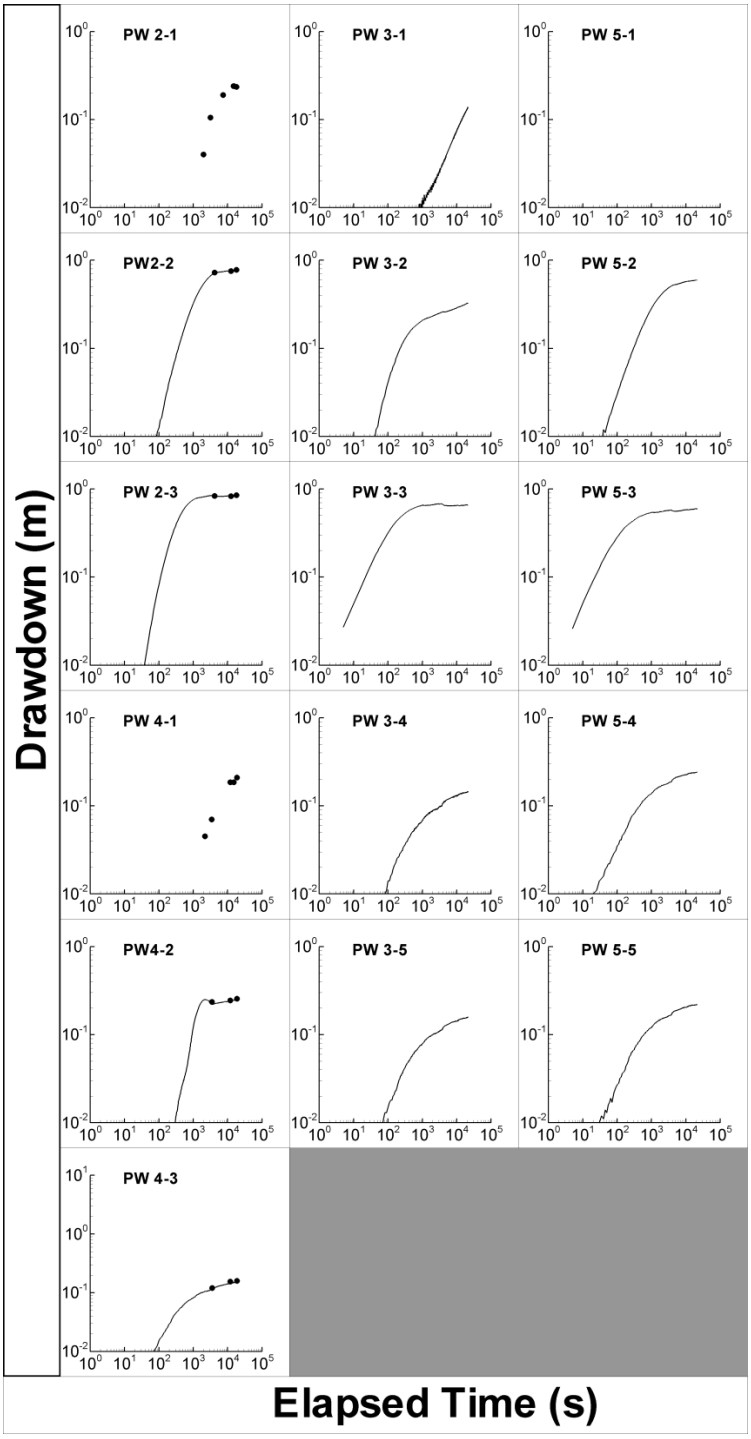


Figure 3.4. Observed drawdown at the CMT wells during the pumping test at PW1-3. The solid line represents transducer recorded data, and the dots are manual water level measurements collected with a water level tape.



3.5. Observed drawdown at the PW wells during the pumping test at PW1-3. The solid line represents transducer recorded data, and the dots are manual water level measurements collected with a water level tape.

3.2 Estimating Hydraulic Parameters When Poroelastic Effects are Significant

During several pumping tests at the NCRS negative drawdown was observed at several deep and shallow observation ports. This effect was determined to be a deformation-induced response caused by deformation of the pumped unit, also known as the Noordbergum effect. This strain results in a stress in the adjacent aquitards that can manifest as either an increase or decrease in water level depending on the geometry of the system. Figure 3.6 shows an example of the head response observed at PW4-1 during pumping at PW4-3 (approximately 6.5 m deeper than PW4-1). This response, though small (approximately 1 cm) at this port, was repeatable.

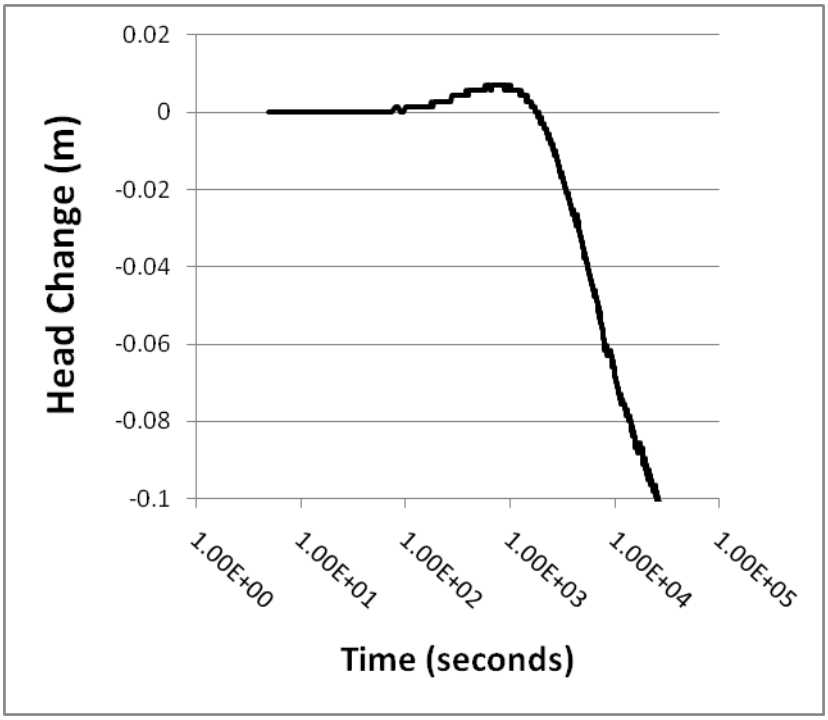


Figure 3.6. Water level response at PW4-1 during pumping at PW4-3. Positive head change is negative drawdown, and negative head change is drawdown.

Prior to this study, there was very little guidance in the literature on how to interpret pumping tests displaying this uncommon early time behaviour. However, in order to incorporate this data into

an inverse model, such as SSLE, it was necessary to understand the impact this effect could have on the estimation of hydraulic parameters. A numerical study was conducted to investigate the impact of poroelastic effects on the estimation of K and S_s. A summary of the manuscript "Estimating hydraulic parameters when poroelastic effects are significant" by *Berg S.J., P.A. Hsieh, and W.A. Illman* [2011] published in *Ground Water* is provided here, with the full manuscript in Appendix E.

Deformation-induced head changes, which are most conspicuous in confining layers overlying or underlying a pumped aquifer, are often unobserved because confining layers are often not monitored during pumping tests. However, even when deformation-induced head changes are recorded, they are often overlooked, unrecognized, or ignored. Possible reasons for this are: 1) the effect is often small and therefore missed; 2) if observed, it may be attributed to other transient effects (diurnal fluctuations, sensor drift or signal noise); and 3) data may not be collected at a frequency high enough to record the effect often seen during a relatively brief period after pumping begins. However, the impact of ignoring the poroelastic effects on the estimation of aquifer parameters has not been addressed in the literature to date.

For this study, three aquifer systems were investigated to assess the impact of these deformation-induced head changes on the estimation of hydraulic parameters. Figure 3.7 shows the various aquifer systems. Each of these model domains were axisymmetric and the pumped well (located in the 'pumped aquifer') was screened across the entire thickness of the aquifer (100 m), had a radius of 0.1 m, was located at the inner boundary of the model, and was pumped at a rate of 50 L/s for a period of 50 days. The time step was initially 15 seconds, and increased by a multiple of 1.2 per time step until model completion (a total of 60 time steps). A summary of the hydraulic properties assigned to each unit for the simulations is included in Table 6.

The following boundary conditions were assigned for all simulations. 1) The inner radial boundary was a specified-flux boundary and experienced no radial displacement and no change in the

vertical component of boundary traction. 2) The outer radial boundary (located at a distance of 10km) was a no-flow boundary with no change in applied stress. 3) The top boundary was a water table and was represented by a constant-head boundary. This assumes that any drawdown that would have occurred at the water table was negligible. This boundary was free of applied stress and was free to deform. 4) The bottom boundary was a no-flow boundary and a no-displacement boundary.

The dots in Figure 3.7 represent observation locations which were divided into five columns of six points each, mimicking a set of multilevel observation wells. Within each group, the observation points were located at different depths below ground surface: 10 m, 30 m, 50 m, 70 m, 90 m, and 150 m for Cases 1 and 2, and 60 m, 80 m, 100 m, 120 m, 140 m, and 200 m for Case 3. This positioning means that for all cases, observation points were located at identical points within each system relative to the vertical center of the pumped aquifer. This allows for a direct comparison of the synthetic water-level data for the different cases. The upper 5 observation points were all contained within the overlying aquitard, while the 6th (deepest) observation point was located in the vertical center of the pumped aquifer. The five multilevel observation wells were located at increasing radial distances from the pumped well: 10 m, 25 m, 50 m, 100 m, and 200 m.

Table 6. Parameters used for model simulations [after *Hsieh*, 1996].

Property	Aquifer Value	Aquitard Value
[1] S_s, specific storage (m^{-1})	1.2×10^{-5}	1.1×10^{-4}
K, hydraulic conductivity (m/s)	1×10^{-4}	1×10^{-7} (1×10^{-9})*
n, porosity (dimensionless)	0.3	0.4
B_w, fluid compressibility (m^2/N)	4.4×10^{-10}	4.4×10^{-10}
[2] G, shear modulus (N/m^2)	3×10^{8}	3×10^{7}
[2] v, drained Poisson's Ratio (dimensionless)	0.25	0.25

[1] Parameters used for simulation without poroelastic effects
[2] Parameters used for simulation with poroelastic effects
* Value of lower aquitard for Case 1

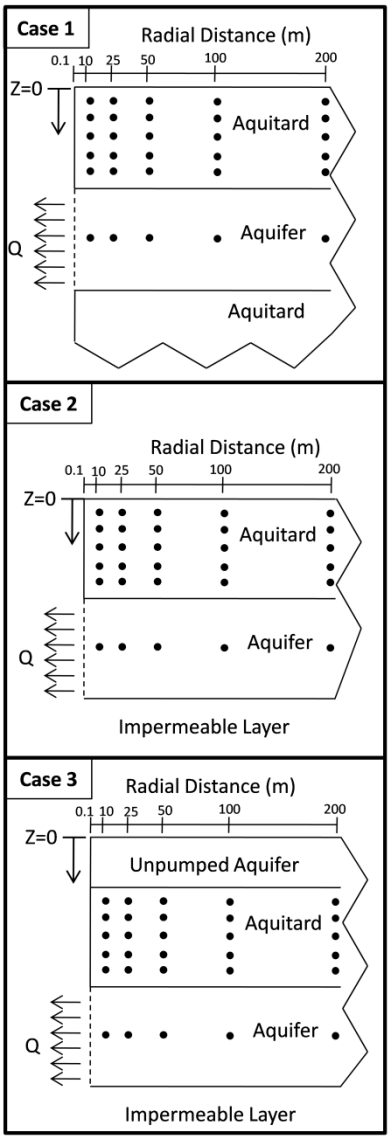


Figure 3.7. Schematic diagram of the 2 dimensional axisymmetric domain for Case 1 after (Hsieh 1996), Case 2, and Case 3. (Dots represent observations points) [after *Berg et al.*, 2011].

For comparison purposes, only Cases 1 and 3 are presented in depth here (Note: Cases 1 and 2 behave in a similar fashion). Figures 3.8 and 3.9 show the simulated head response for the Cases 1 and 3 respectively. The solid line represents the observed response for the simulation with poroelastic

effects, and the dashed line represents the simulation without poroelastic effects (i.e. traditional groundwater model).

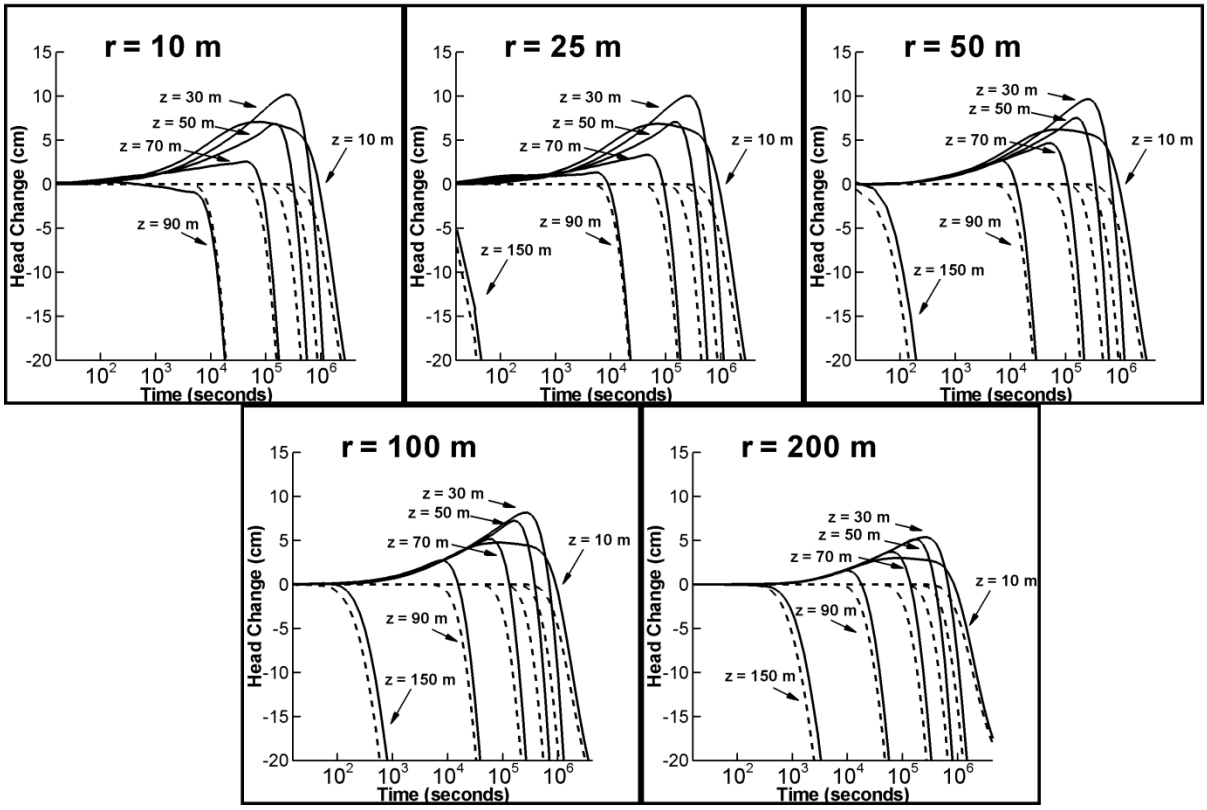


Figure 3.8. Change in hydraulic head vs. time for Case 1. Solid line represents the simulation with poroelastic effects. Dashed line represents the simulation without poroelastic effects. Positive head change indicates head rise. Negative head change indicates head drop [after *Berg et al.*, 2011].

The simulated head changes in Case 1 (Figure 3.8) show that, with one exception, deformation-induced head rise occurs in all observation points in the aquitard. The magnitude of the head rise depends on vertical position within the aquitard and radial distance from the pumped well, with those observation points closest to the pumped well (r = 10 m) and in the upper part of the aquitard (z = 30 m) experiencing the greatest head rise. Near the top of the aquitard (z = 10 m), the head rise is lessened due to proximity to the constant-head boundary. The exception to head rise

occurs at the observation point at r = 10 m and z = 90 m. At this observation point, hydraulic head starts to fall earlier than the start of drawdown in the corresponding non-poroelastic simulation. This behavior is consistent with the finding by *Hsieh* [1996, p. 1085]—deformation caused by pumping can induce a small region of pore volume increase, and therefore head drop, in the aquitard near the pump interval.

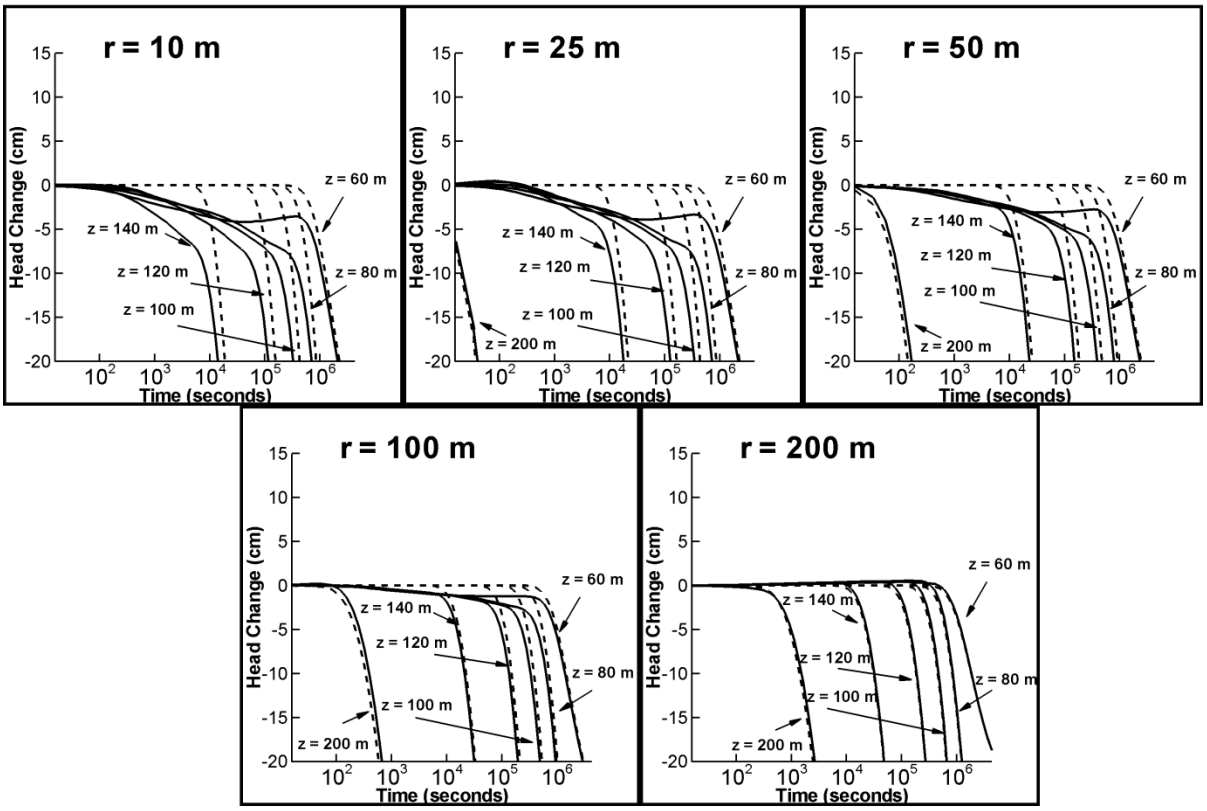


Figure 3.9. Change in hydraulic head vs. time for Case 3. Solid line represents the simulation with poroelastic effects. Dashed line represents the simulation without poroelastic effects. Positive head change indicates head rise. Negative head change indicates head drop [after *Berg et al.*, 2011].

The simulated head changes for Case 3 (Figure 3.9) are significantly different from those of Cases 1 and 2. In Case 3, there is no head rise at the aquitard observation points. Instead, the onset of drawdown in the aquitard occurs much earlier in the poroelastic case than in the non-poroelastic case.

This behavior is more prominent at small radial distances and diminishes with distance from the pumped well (Figure 3.9). The absence of a 'reverse' water-level response in Case 3 seems to be the result of including the overlying aquifer and the change in position of the top (traction-free) boundary. For Cases 1 and 2 (Figures 3.8 for Case 1), the analyst would likely recognize that the rising portion of the head data cannot be matched by traditional type curves and possibly exclude it from the analysis. However, for Case 3, if the deformation-induced head drop is interpreted as hydraulic propagation from the pumped aquifer into the aquitard, this could severely overestimate the aquitard hydraulic diffusivity (K'/S_s').

Using the poroelastic drawdown response for these cases, hydraulic parameters were estimated for the pumped aquifer (K and S_s) and the overlying aquitard (K' and S_s') using; a) the *Neuman-Witherspoon* [1969] solution, and b) matching the data by coupling a parameter estimation program (PEST) to a traditional groundwater model that does not account for poroelastic effects. The PEST analysis was performed in two ways; i) the entire head change response (both positive and negative drawdown) was incorporated into the parameter estimation (PEST-1); and ii) only the data after the drawdown was below static was included in the parameter estimation (PEST-2).

When examining the aquifer parameters obtained by type-curve analysis and parameter estimation, it is straightforward to compare the hydraulic conductivities (K and K') estimated from the type-curve analysis with the corresponding hydraulic conductivities used in the poroelastic simulation. The latter conductivities are referred to as the “true” K and K'. By contrast, there is no rigorous way to evaluate the specific storages (S_s and S_s') estimated from the type-curve analysis and parameter estimation, as specific storage is not used in the poroelastic simulation. Nonetheless, for the purpose of comparison, specific storage values are computed for the poroelasticity simulation as follows. First, the parameters G (shear modulus of the skeletal matrix) and v (drained Poisson's ratio)

from the poroelastic simulation are used to calculate the vertical matrix compressibility β_p, according to (Hsieh 1996 Equation 3):

$$\beta_p = \frac{1-2\nu}{2G(1-\nu)} \tag{4}$$

where β_p is the vertical matrix compressibility.

Then, the specific storage is computed using [*Domenico and Schwartz*, 1997 Equation 4.30]:

$$S_s = \rho_f g\left(\beta_p + n\beta_f\right) \tag{5}$$

where β_f is the fluid compressibility.

Equation 5 gives the standard specific storage in groundwater theory. In poroelasticity theory, this quantity is sometimes referred to as "uniaxial specific storage" [*Wang*, 2000].

The results of the type-curve and PEST estimations are presented on Figure 3.10 and compared to the 'true values' (those that were used in the original poroelastic simulations) and are presented in Table 6. In the overlying aquitard, the K' estimates obtained by PEST using data near the pumped well are significantly closer to the true K' than the K' estimates obtained by type-curve matching. This is because the numerical model coupled with PEST accounts for radial flow in the aquitard, whereas the *Neuman-Witherspoon* [1969] model does not. For the same reason, the S_s' estimates obtained by PEST using data near the pumped well are significantly closer to the specific storage value computed with Equation 5 than the S_s' estimates obtained by type-curve matching. When matching to data at greater distances from the pumped well, both PEST and type-curve matching yield similar estimates for K' and for S_s'.

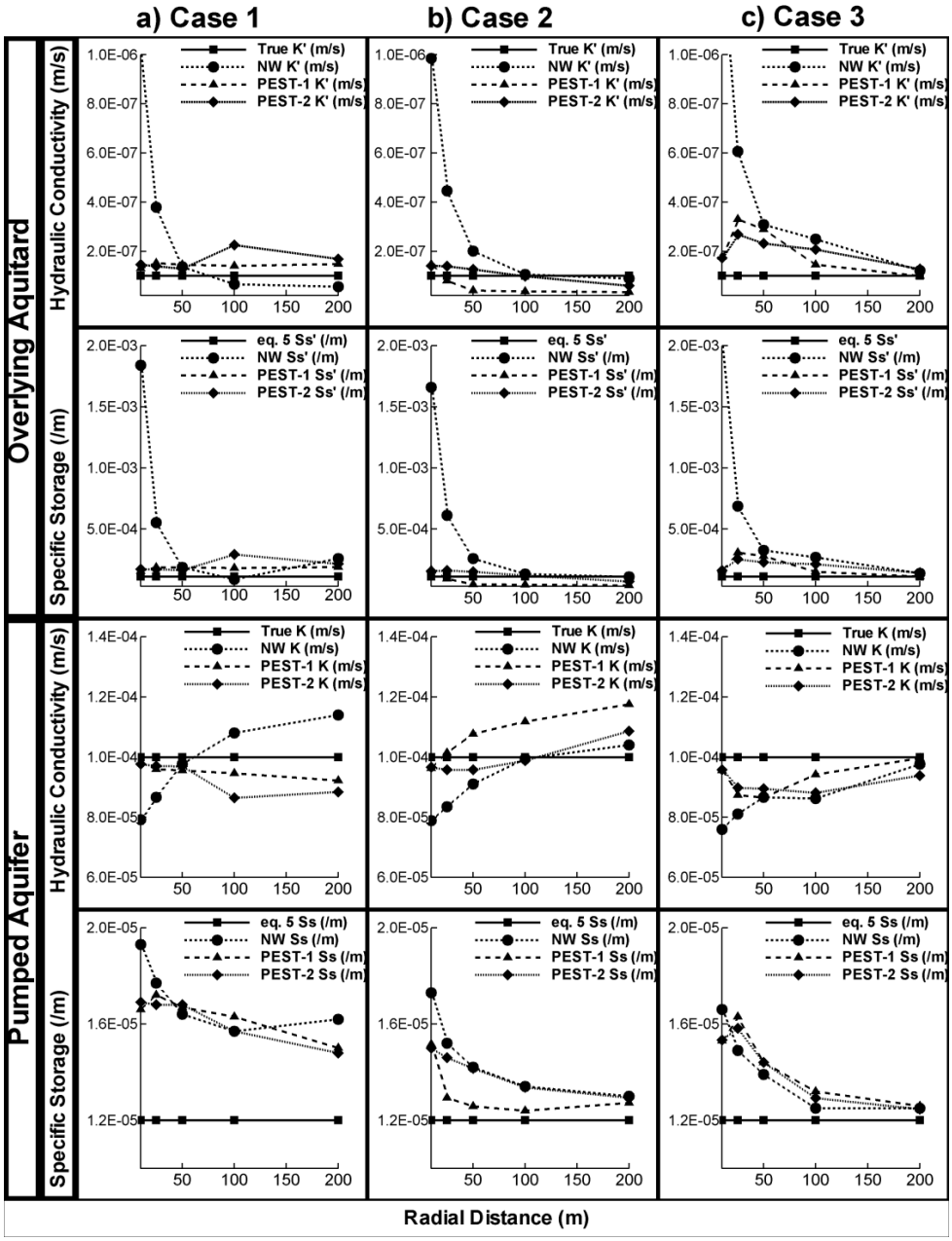


Figure 3.10. Estimated hydraulic parameters for Cases 1, 2 and 3 with poroelasticity effects, using data from varying radial distances. The solid line with squares represents the true value for K and K', and the value calculated from Equation 5 for S_s and S_s'. The dashed line with circles represents a parameter estimated by the *Neuman-Witherspoon* [1969] type-curve method (NW). The dashed line with triangles represents a parameter estimated using PEST and including all synthetic data (PEST-1). The dotted line with diamonds represents a parameter estimated using PEST and excluding synthetic data that are influenced by deformation-induced effects (PEST-2) [after *Berg et al.*, 2011].

For the pumped aquifer, the K estimates obtained by PEST and by type-curve matching are all within ±20% of the true K. However, the S_s estimates obtained by PEST and by type-curve matching are generally higher than the S_s value computed using Equation 5. This difference is particularly noticeable in Case 1. This observation suggests that the volume of fluid released by a unit volume of porous medium per unit decline in hydraulic head in a poroelasticity simulation (with three-dimensional deformation) is somewhat larger than the fluid released in a traditional groundwater simulation using S_s computed by Equation 5. For the aquifer-aquitard systems examined in this study, the S_s estimated by well test analysis can be 10% to 60% higher than the S_s computed by Equation 5. However, these differences are comparable to uncertainties associated with parameters estimated from typical pumping tests.

Figure 3.10 also shows that little is gained by the PEST-2 runs, which exclude head data that are influenced by deformation-induced effects, compared to the PEST-1 runs, which include all head data. This is likely due to the relatively small magnitude of the deformation-induced effects when compared to the increasing magnitude of drawdown as the pumping test progresses. The relatively large drawdowns during the latter part of the test dominate the estimation of the hydraulic parameters.

For the most part, both the Neuman-Witherspoon type-curve matching method and the PEST estimation method yield reasonable estimates for hydraulic parameters, as long as pumping is sufficiently long so that deformation-induced effects have largely dissipated. The exception occurs when applying the curve-matching method to estimate aquitard parameters using data collected in the aquitard near the pumped well. As noted above, the assumption of vertical flow in the aquitard can cause significant errors in these estimates. Excluding head data that exhibit deformation-induced effects does not yield significant improvements in the estimated parameters.

These findings suggest that for drawdown data exhibiting deformation-induced head changes it should be possible to ignore the portion of the curves affected by this behaviour and just analyze the portion of the curve that has drawn down below the static water level.

3.3 Three-Dimensional Transient Hydraulic Tomography in a Highly Heterogeneous Glaciofluvial Aquifer-Aquitard System

This section demonstrates THT at the field scale in a highly heterogeneous glaciofluvial aquifer-aquitard sequence. This manuscript "Three-Dimensional Transient Hydraulic Tomography in a Highly Heterogeneous Glaciofluvial Aquifer-Aquitard System" by *Berg S.J. and W.A. Illman* [*submitted manuscript*] is summarized here and included as Appendix F.

Objectives of the study included: 1) performing THT at the field scale in 3-dimensions; 2) examining the effect of including *a priori* information on the inversions; 3) assessing the performance of K- and S_s-tomograms by comparing the results to borehole data, and through the simulation of independent pumping tests.

To accomplish these objectives, 9 pumping tests were performed at the North Campus Research Site (NCRS), located on the University of Waterloo (UW) campus, in Waterloo, Ontario Canada. Details of the site and instrumentation can be found in section 3.1. The data from these tests were then analyzed using the stochastic inverse model, Successive Sequential Linear Estimator (SSLE) [*Zhu and Yeh*, 2005], to perform THT. First a stochastic inversion of each individual pumping test was performed, followed by sequentially inverting 4 pumping tests (THT). For the purposes of this summary, only the results of THT are presented (see Appendix G for the stochastic inversion of individual pumping tests). These THT inversions were performed using three different types of *a priori* information: 1) effective K and S_s fields estimated from pumping test data (case 1); 2) 3-layer K and S_s fields with each layer being homogeneous based on parameter estimates from type-curve analysis based on the *Hantush* [1960] solution (case 2); and 3) homogeneous K and S_s fields with the inverse model populated with local-scale K obtained via permeameter analysis of core samples along 5 boreholes (case 3). The tomograms (K and S_s fields) generated from the single- and multi-test

inversions were then used to simulate the 9 original pumping tests. The simulation of pumping tests used to generate the tomograms are considered calibration tests and give an indication of how well the tomogram reproduces the input data; and, tests not included in the generation of the tomograms are considered 'validation' tests, and reflect the predictive ability of the tomograms.

Due to the significant computational effort of running the THT inversions for the full duration of the pumping tests (6.5 to 24 hrs), only the first hour of each pumping test was included in the inversion (currently additional simulations are being completed to investigate the inclusion of intermediate and late time data, so as to capture the entire drawdown curve). During the first hour of each test, many ports show a strong response, and some were observed to reach steady state. For each drawdown curve, 3 to 5 points were selected to define the curve (between 112 and 130 data points were included for each test). Unresponsive ports were also included in the inversion as they provide information on connectivity between the pumped and observation ports.

For each case, the tests with pumping taking place at intervals PW4-3, PW1-3, PW5-3, and PW3-3 were included in the inversion (in the order listed). These tests, and the order they were included in the inverse model were selected based on the degree to which they stressed the aquifer (PW4-3, and PW1-3 had the highest flow rates of the 9 pumping tests) and their spatial coverage of the site (PW5-3, and PW3-3). This is consistent with the recommendations of *Illman et al.* [2008]. After all 4 pumping tests were included in the inversion, the process started over at the first test. This loop was performed 3 times or until the convergence criteria were met. Convergence criteria were met when either the change in the estimated fields is below a minimum threshold (e.g. variance changes less than 0.01) or the maximum number of iterations has been achieved.

The model domain used for all inversions and forward simulations measured 45 m × 45 m × 15 m and was composed of variably-sized rectangular elements. The smallest elements, measuring 0.5 m × 0.5 m × 0.5 m were located in the central portion of the model domain and correspond to the

location of the 15 m x 15 m test plot. Outside of the test plot, the element size increased with increasing distance from the center of the model domain. In total, the model domain consists of 29,791 elements (31 × 31 × 31) and 32,768 nodes (32 × 32 × 32). The boundary conditions were no flow for top and bottom boundaries and constant head for the remaining outer boundaries.

Estimates of K and S_s to populate the initial homogeneous field (case 1) were obtained by coupling PEST [*Doherty*, 2005] with HGS [*Therrien* et al., 2005] and individually matching the drawdown response at each observation port for pumping tests performed at PW1-3. The model domain used for the parameter estimation was the same as the one used for the THT inversions. The geometric means of these individual estimates were rounded and then incorporated into the starting model ($K = 8.0 \times 10^{-6}$ m/s and $S_s = 1.0 \times 10^{-4}$ m^{-1}). Figures 3.11a and 3.11c show the final estimated K- and S_s-tomograms for case 1 respectively.

In general, this case was able to capture the heterogeneity evident in the upper portion of the aquifer. However, the inversion did assign high K material to the lower portion of the domain, which, based on core logs and permeameter data (see Figure 3.1) is known to be low K (silt/clay). Examination of the variance maps (Figures 3.11b and 3.11d) indicates that the estimates in this region have a high degree of uncertainty, a reflection of the very small, to absent, drawdown responses observed in the lower ports, even at very late time (>20 hrs).

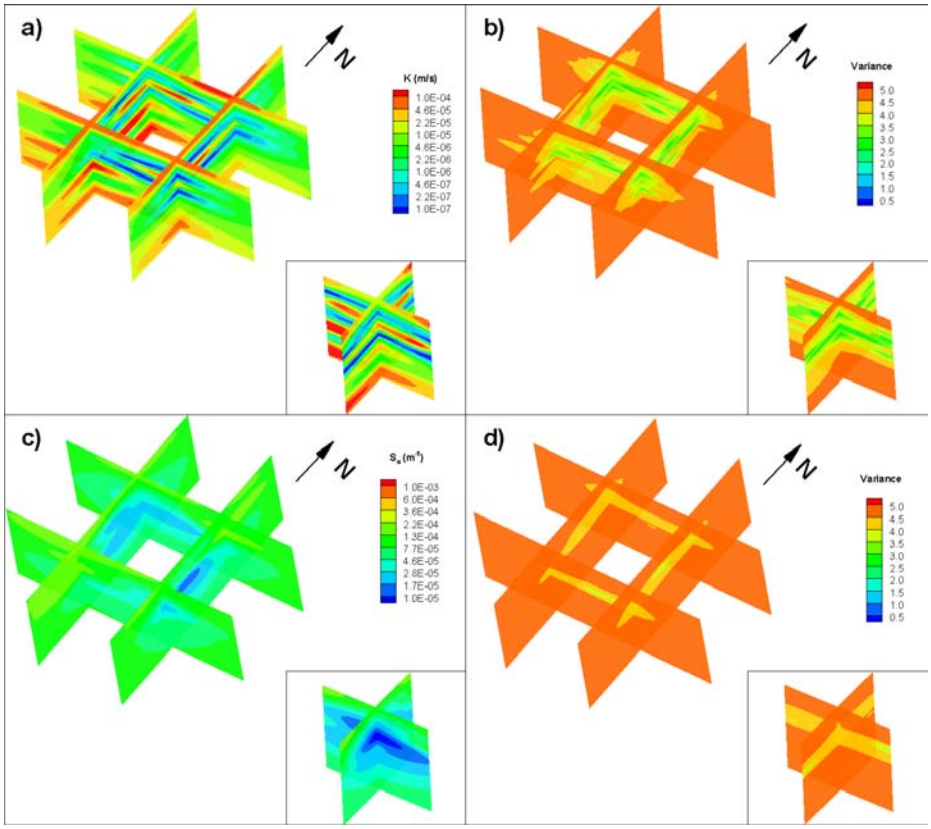


Figure 3.11. Case 1 THT *K*- and S_s-tomograms after the inclusion of 4 pumping tests (PW4-3, PW1-3, PW5-3, and PW3-3) : a) *K*-tomogram; b) variance associated with the estimated *K* tomogram; c) S_s-tomogram; and d) variance associated with the estimated S_s-tomogram. Note that the square formed by the slices in the main figure correspond to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

For case 2, the aquifer system was conceptualized as a main aquifer system bounded above and below by aquitards (each 5 m thick). Initial values of *K* and S_s for these layers were estimated by performing a *Hantush* [1960] type-curve analysis on the observation points located within the central aquifer unit during pumping tests performed at PW1-3, PW1-4, and PW 1-5. Each drawdown response was individually matched using *Aqtesolv Pro 4.5* [*Duffield*, 2007] to estimate *K* and S_s for the 'aquifer zone' as well as leakage parameters for the adjacent aquitards. The leakage parameters can

then be used to calculate estimates of K and S_s for the aquitards. However, since this solution does not account for observation ports situated within the aquitard units, there is much more uncertainty in the estimates of these parameters. Thus, the geometric mean of the aquitard estimates was taken and assigned to both aquitards. Additionally the geometric mean of all K and S_s estimates made for the aquifer zone was used to represent the aquifer. Values assigned to the aquifer as starting conditions for THT were 8.5×10^{-6} m/s and 2.2×10^{-5} m^{-1} for K and S_s respectively; and 4.5×10^{-7} m/s and 1.0×10^{-4} m^{-1} for the aquitard K and S_s respectively.

Figure 3.12 shows the results of the THT analysis starting with a simple layered model. While the upper and lower portions of the model domain appear significantly different than Case 1, the pattern in the central aquifer zone is quite similar (see inset in Figures 3.12a and 3.12b).

SSLE [*Zhu and Yeh*, 2005] allows for the inversion to begin with local-scale estimates of K and S_s when they are available. For case 3, the THT inversion started with homogeneous K and S_s fields (same values as case 1), and had 127 values of K specified along boreholes PW1, CMT1, CMT2, CMT3, and CMT4. The K estimates were based on the permeameter measurements of *Alexander et al.* [2011]. Since the vertical element size was 0.5 m and *Alexander et al.* [2011] sampled at intervals as small as 0.1 m, the geometric mean of multiple samples that fell within the same model element was used.

Illman et al. [2008] found that starting the inverse model with poor point data can have a detrimental effect on the resulting tomogram. Since the estimates of S_s for the NCRS span a very wide range and have a high degree of uncertainty, these values are omitted from Case 3.

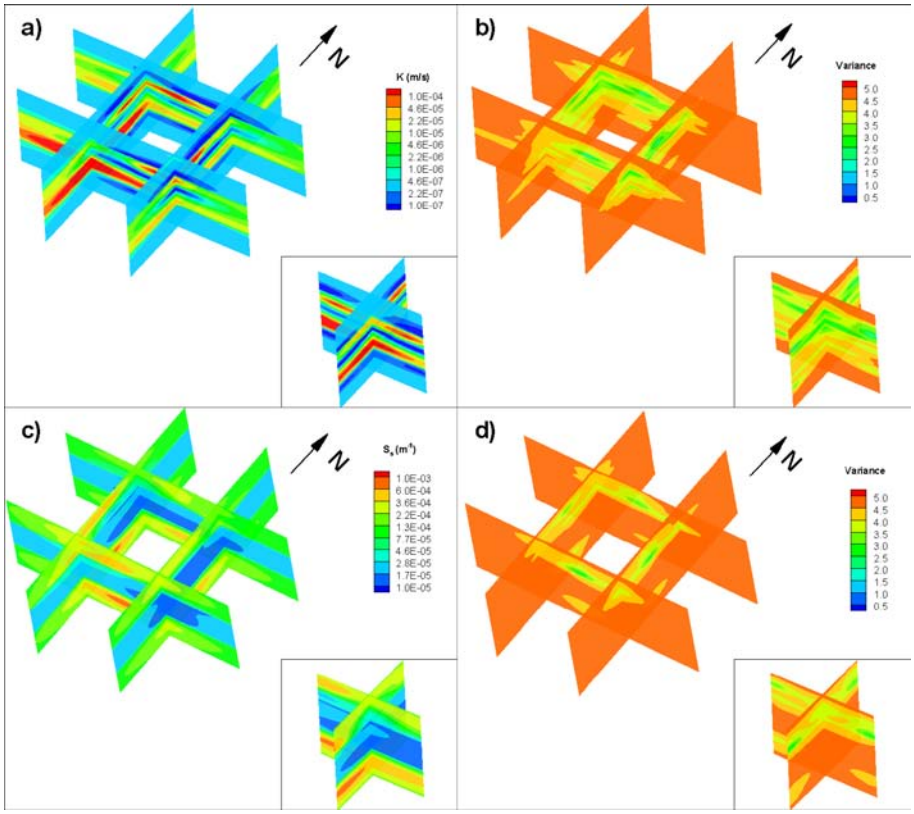


Figure 3.12. Case 2 THT *K*- and S_s-tomograms after the inclusion of four pumping tests (PW4-3, PW1-3, PW5-3, and PW3-3): a) *K*-tomogram; b) variance associated with the estimated *K*-tomogram; c) S_s-tomogram; and d) variance associated with the estimated S_s-tomogram. Note that the square formed by the slices in the main figure correspond to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

The results of THT that include permeameter *K* data in the starting model are presented in Figure 3.13 (case 3). Initially, the tomograms appear quite different than those presented previously. However, the difference mainly occurs in the lower region of the aquifer where there is greater uncertainty in the estimated values. In the middle to upper portions of the model domain where the main aquifer zones exists, a very similar pattern of an aquifer zone separated by a low *K* zone is evident.

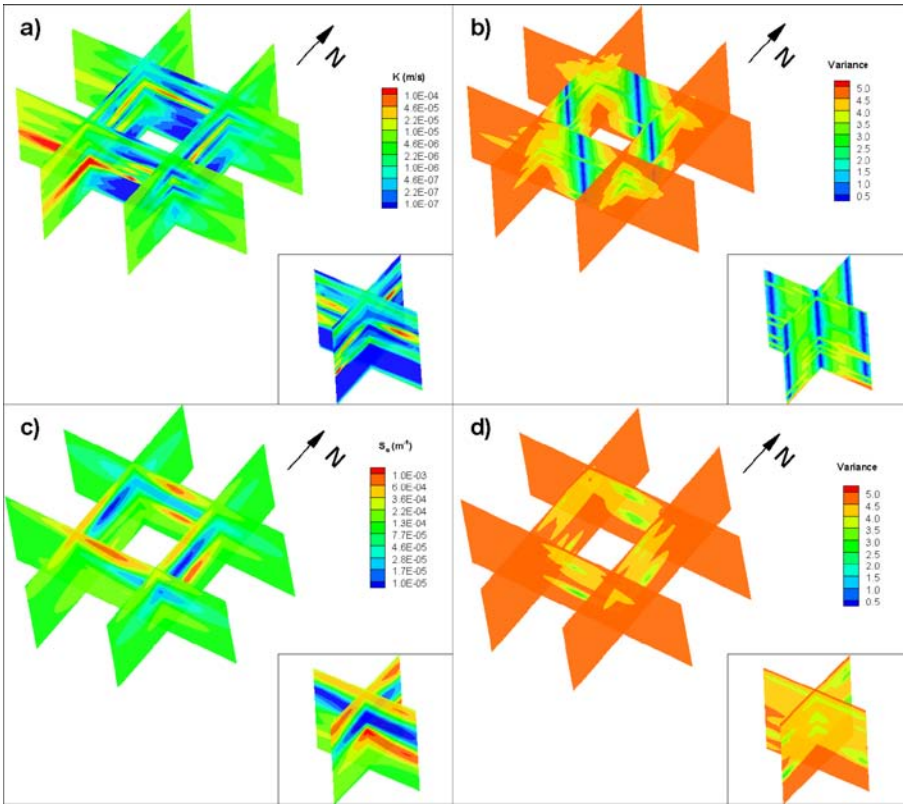
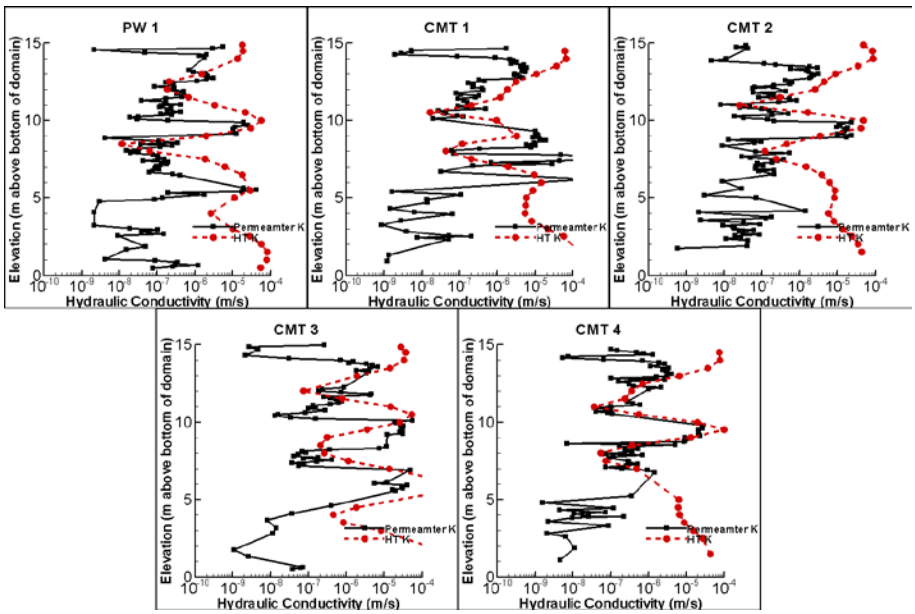


Figure 3.13. Case 3 THT *K*- and S_s-tomograms after the inclusion of four pumping tests (PW4-3, PW1-3, PW5-3, and PW3-3): a) *K*-tomogram; b) variance associated with the estimated *K*-tomogram; c) S_s-tomogram; and d) variance associated with the estimated S_s-tomogram. Note that the square formed by the slices in the main figure correspond to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

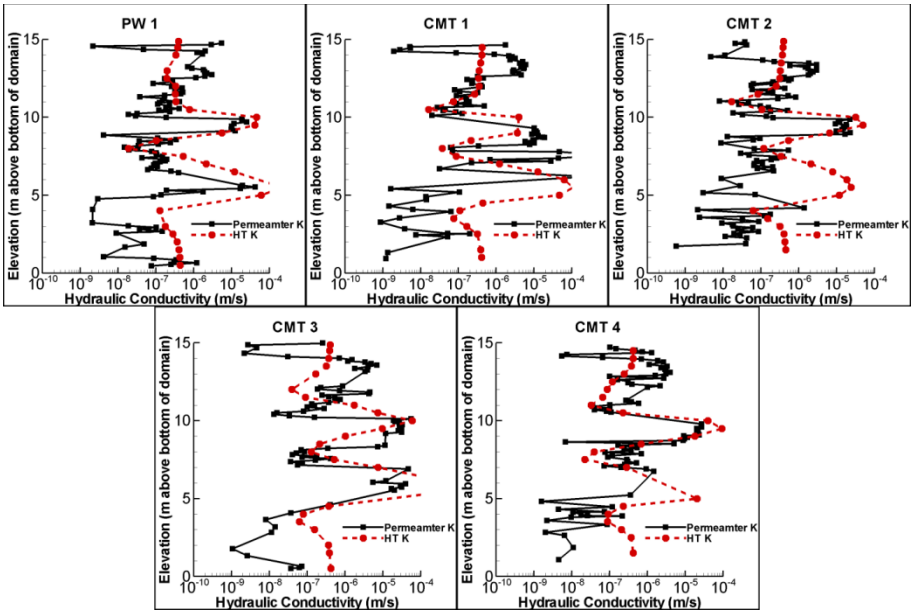
The accuracy of the *K*- and S_s-tomograms generated by three THT cases were assessed: 1) visually (i.e., do the tomograms reflect what is known of the site based on core data?); 2) by comparing *K* profiles along boreholes to permeameter values; and, 3) simulating all 9 pumping tests using the *K* and S_s tomograms from each case and comparing the simulated drawdown to the observed drawdown. For the purposes of this summary, only points 2 and 3 will be presented.

The permeameter *K* data obtained by *Alexander et al.* [2011] is compared to the tomograms from three THT cases. The comparison is made along 5 boreholes (PW1, CMT1, CMT2, CMT3, and CMT4). Figures 3.14a, 3.14b, and 3.14c, show the comparison for THT cases 1, 2, and 3 respectively. This permeameter data is the same as that presented in Figure 3.1.

a)



b)



c)

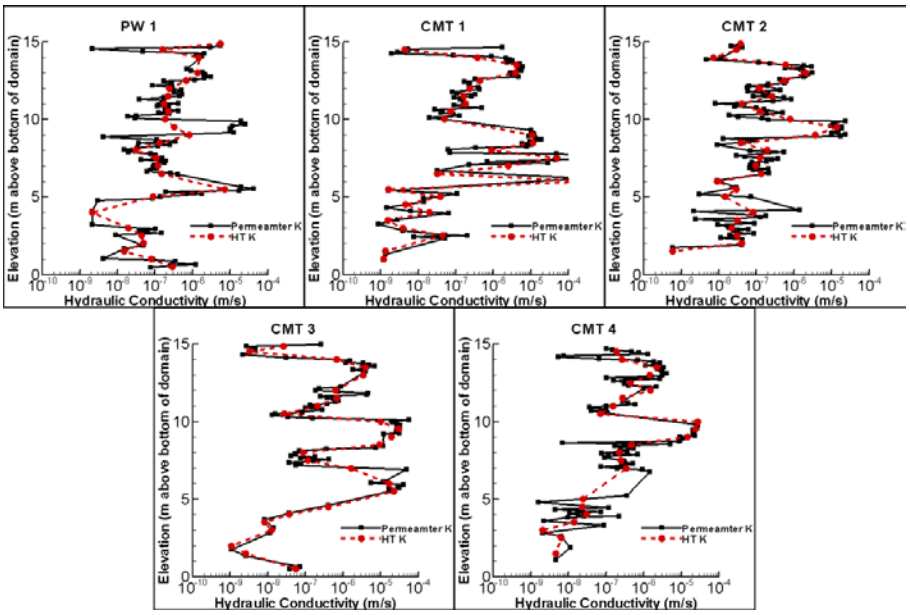


Figure 3.14. Vertical *K* profile along boreholes for the three THT cases, 1 (a), 2 (b), and 3 (c). The solid black line with squares is the permeameter *K* data, and the dashed red line with circles are the values estimated by the HT analysis.

Case 1 captured the main high K unit (located approximately 10 m above the base of the model domain) and the adjacent low K zones. This was particularly evident for CMT4. However, the K profile as estimated by THT for this case deviates significantly from the permeameter data near the upper and lower boundaries of the model domain.

Case 2 was able to quite accurately capture both aquifers located at 10 m and 5 m above the base of the model domain. In addition, the upper and lower portions of the model domain were better represented for this case in comparison to case 1. This suggests that starting the model with K and S_s fields that approximate the true case, in a general sense, can lead to more realistic tomograms.

Despite being allowed to vary after the estimation of the first pumping test was complete, estimates of K along the 5 boreholes for Case 3 were still quite close to the original data incorporated in the THT inversion. This could suggest that permeameter K data, even with the availability of a large number of estimates, may not be entirely suitable for representing large scale groundwater flow within an aquifer-aquitard sequence as heterogeneous as that at this site. This once again reinforces the argument of *Illman et al.* [2008] that the data included as 'true' must be selected carefully.

For each of the THT cases, all 9 pumping tests were simulated and the observed and simulated results were compared using scatterplots (see Figures 3.15, 3.16, and 3.17 for cases 1, 2, and 3 respectively). Note that THT was performed using the 1st hour of data recorded for each pumping test, however, the entire test duration was simulated for each pumping test (6.5 - 22.5 hrs).

All 3 cases appear to perform similarly, with quite good matches for the pumping tests used for calibration purposes (PW1-3, PW4-3, PW5-3, and PW3-3). Poorer matches were seen for pumping tests not included in the inversion, in particular, matches for PW1-4 and PW5-4 were quite poor for all 3 cases. Interestingly, these ports were adjacent to pumping tests included in the THT inversion. Intermediate quality matches were observed for PW1-5, PW3-4, and PW5-5. Overall, Case

2 appears to perform slightly better than cases 1 and 3, particularly for independent pumping tests performed at PW1-5, PW3-4, and PW5-5.

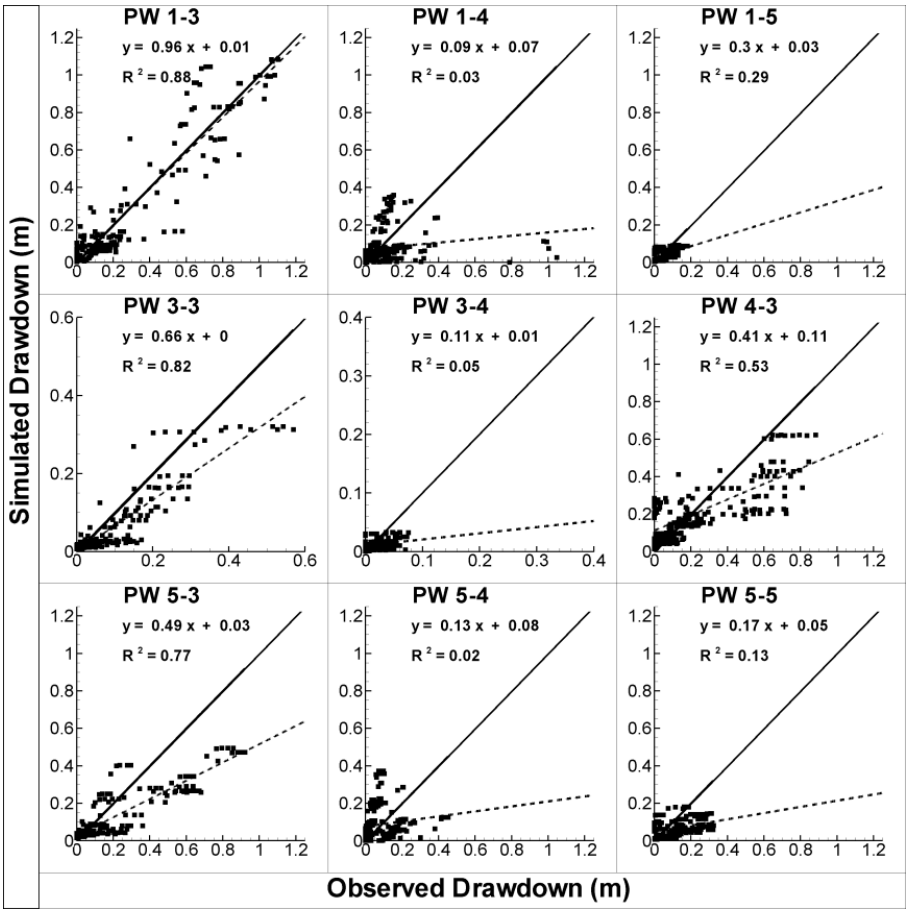
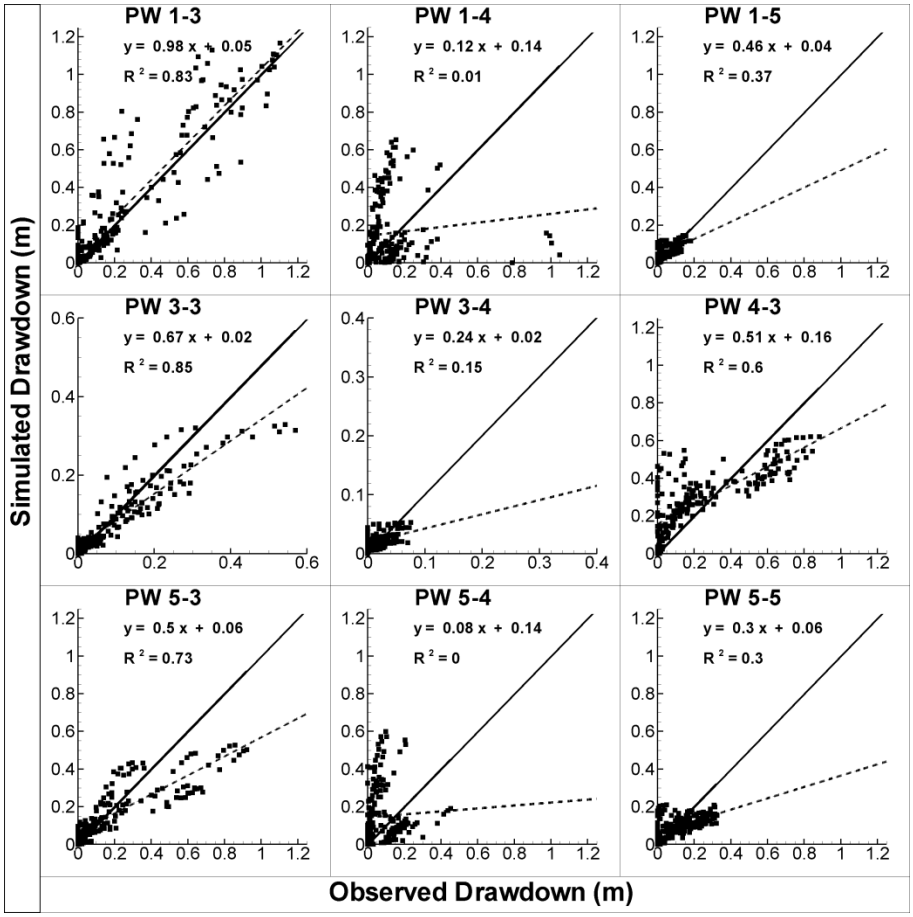


Figure 3.15. Scatterplot for all 9 pumping tests using the tomograms generated from THT Case 1. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.



Figures 3.16. Scatterplot for all 9 pumping tests using the tomograms generated from THT Case 2. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

Table 7. Summary of L2 norms (simulated and observed) for the three THT cases.

	PW1-3	PW1-4	PW1-5	PW3-3	PW3-4	PW4-3	PW5-3	PW5-4	PW5-5
Case 1	0.01	0.03	0.002	0.004	0.0004	0.03	0.02	0.01	0.01
Case 2	0.02	0.06	0.001	0.003	0.0003	0.03	0.02	0.03	0.01
Case 3	0.06	0.04	0.01	0.003	0.001	0.05	0.02	0.05	0.01
Max	0.06								
Min	0.0003								

All 3 cases produced fairly similar representations of the central portion of the aquifer where a strong response was recorded, however, the most noticeable difference between these cases was for the lower portion of the domain, where very little to no drawdown was recorded. The lack of drawdown in this lower portion of the system provides a challenge to the inversion resulting in higher uncertainties in this region. Cases 2 and 3 preserved the low *K* material in the bottom portion of the aquifer, indicating, that for unresponsive regions (or regions with very little drawdown) the inclusion of additional information, either with a heterogeneous starting model, or the point estimates of hydraulic parameters may improve estimates for these regions. In terms of predicting drawdown, both cases 2 and 3 are very similar, however, case 2 better reproduces the observed drawdowns for the pumping tests at PW1-3 and PW4-3. Thus, case 2 is identified as the best performing THT inversion.

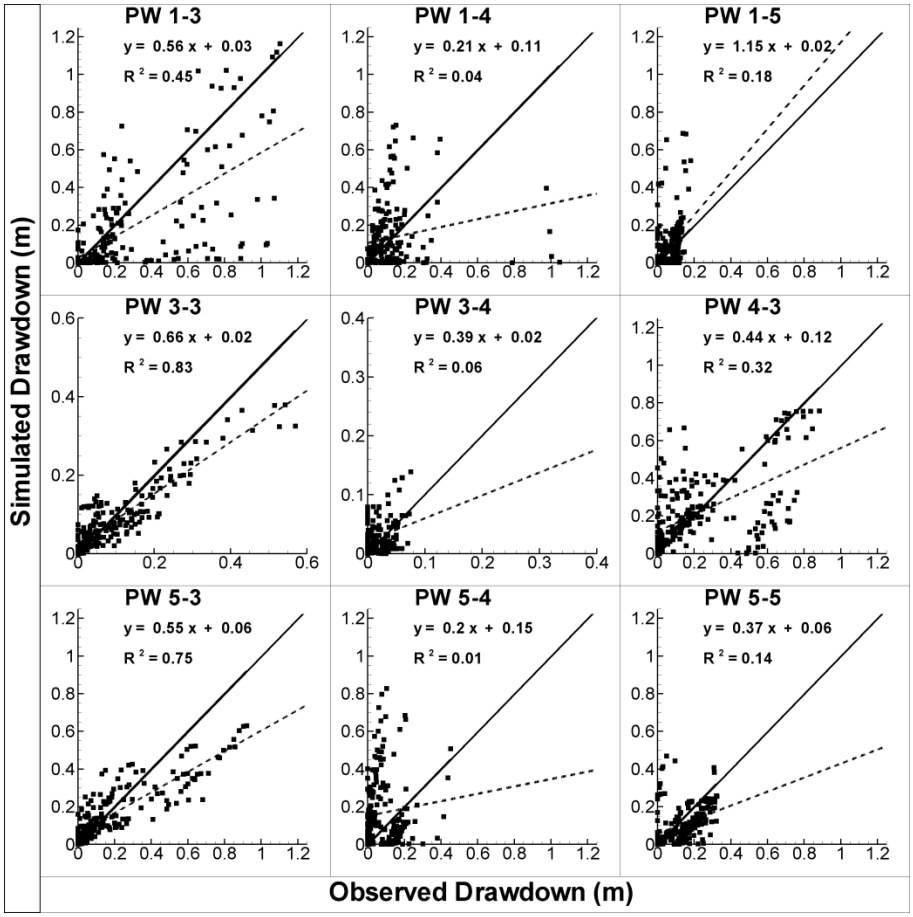


Figure 3.17. Scatterplot for all 9 pumping tests using the tomograms generated from THT Case 3. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

3.4 Comparison of Subsurface Heterogeneity Modeling Approaches Through Transient Groundwater Flow Modeling in a Highly Heterogeneous Glaciofluvial Aquifer-Aquitard Sequence

This manuscript compares *K*- and S_s-tomograms estimated by THT to other characterization/modeling approaches. The manuscript, "Comparison of Subsurface Heterogeneity Modeling Approaches Through Transient Groundwater Flow Modeling in a Highly Heterogeneous Glaciofluvial Aquifer-Aquitard Sequence" by *Berg S.J. and W.A. Illman* [*in prep*] is summarized here and included as Appendix G.

This study is similar to the THT laboratory validation presented in section 2.1 except the comparison is made at the field scale. For the purposes of this comparison, 5 heterogeneous modeling/characterization approaches are applied to the highly heterogeneous glaciofluvial aquifer at the NCRS. These include: 1) calculation of effective parameters through model calibration of a pumping test; 2) ordinary kriging of *K* values obtained by performing permeameter tests on soil cores (based on the interpretation of *Alexander et al.* [2011]); 3) conditional simulations based on the Transition Probability/Markov Chain method using soil core logs; 4) construction of a stratigraphic model and its calibration to a single pumping test; and 5) transient hydraulic tomography analysis of 4 pumping tests (similar to case 2 of section 3.3).

While other studies have examined and compared various inverse modeling strategies [*Zimmerman et al*., 1998; *Maji, 2005; Maji and Sudicky*, 2008; *Franssen et al*., 2009], these studies relied on synthetic true fields, thus allowing the results of the various strategies to be compared to the true fields. However, for this study, the true parameter fields are unknown, thus, we evaluate the various approaches by: 1) comparing the *K* obtained by these various cases to permeameter *K* values

measured along 5 boreholes; and 2) simulating 9 pumping tests performed at the field site and comparing the simulated drawdowns to the observed.

Since permeameter tests were only performed on the soil cores collected along 5 boreholes (PW1 and CMT1 through 4) the characterization/modeling efforts for this comparison only use data from these boreholes. For example, the THT inversion, as applied in this manuscript, only analyzed drawdown recorded in these 5 wells. The data collected at the four corner wells, however, was still used for validation purposes.

The parameter fields for heterogeneous cases 2 through 5 are shown in Figures 3.18 to 3.21 respectively. The kriged permeameter field (case 1) was based on the analysis of 471 permeameter tests performed on soil cores collected during the installation of PW1, and CMTs 1 through 4. *Alexander et al.* [2011] determined that the *K* data were both log-normally distributed and that the mean and variance within the investigation area were stationary. *Alexander et al.* [2011] kriged the permeameter *K* data by fitting an exponential omni-directional variogram to the experimental variogram for 5 boreholes (Figure 3.18).

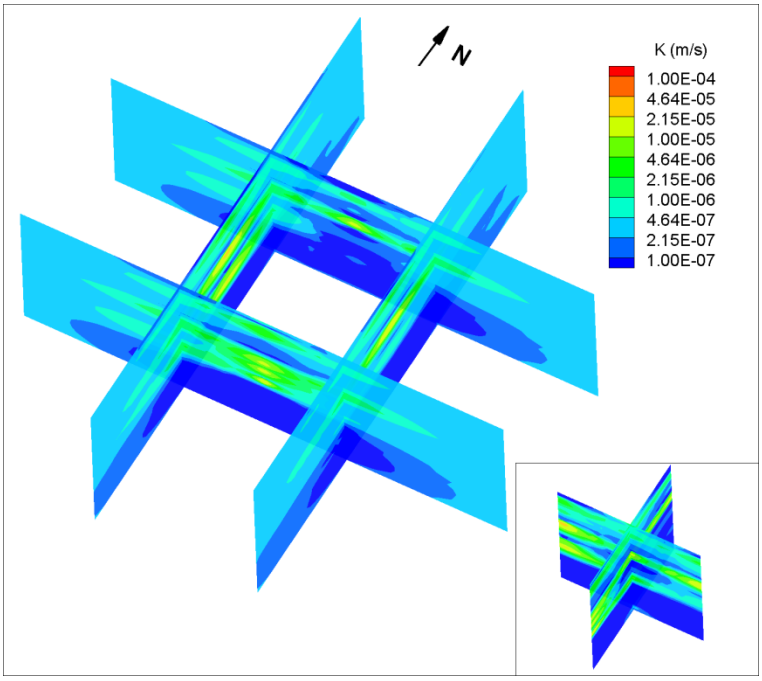


Figure 3.18. Three-dimensional perspective view of the kriged permeameter *K* field. The square created by the slices represents the outer edge of the 15 m x 15 m plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

For case 3 (Figure 3.19) the 8 material types identified by *Alexander et al.* [2011] (sandy Gravel, Sand, Sand and Silt, sandy Silt, Silt, clayey Silt, Silt and Clay, and silty Clay) were simplified to 4 (sandy gravel, sand, silt, and silty clay). These material types were then used with the transition probability/Markov Chain geostatistical approach to create conditional realizations of the site [*Carle*, 1999; *Carle and Fogg*, 1997; *Weissmann et al.*, 1999]. The reduction of categories was performed based on grouping categories with similar *K* (e.g. clayey Silt, Silt and Clay, and silty Clay). It was also found that reducing the number of categories improved the simulated to observed matches of the transition probabilities. Using the data along boreholes PW1, CMT1, CMT2, CMT3, and CMT4 the bi-directional (horizontal and vertical) experimental transition probabilities were computed and

matched automatically with a Markov chain model using TPROGs [*Carle*, 1999]. A sample realization generated by TPROGs [*Carle*, 1999] is presented in Figure 3.19. K values were assigned by taking the geometric mean of permeameter samples that corresponded with each material type. The material types were assigned K values of 1.0×10^{-8} m/s, 8.7×10^{-7} m/s, 8.8×10^{-6} m/s, and 4.6×10^{-5} m/s for silty clay, silt, sand and sandy gravel respectively. Because of the difficulty in estimating S_s from small scale samples, estimates of S_s for the different material types were assigned based on the individual estimates obtained by using PEST to estimate effective S_s values by matching the drawdown observed at each port. Each observation port was categorized based on the predominant material surrounding it, and the geometric mean of multiple S_s values for each material type was taken. The assigned S_s values were 1.4×10^{-4} m/s, 2.2×10^{-4} m/s, 1.1×10^{-4} m/s, and 1.0×10^{-4} m/s for silty clay, silt, sand and sandy gravel respectively.

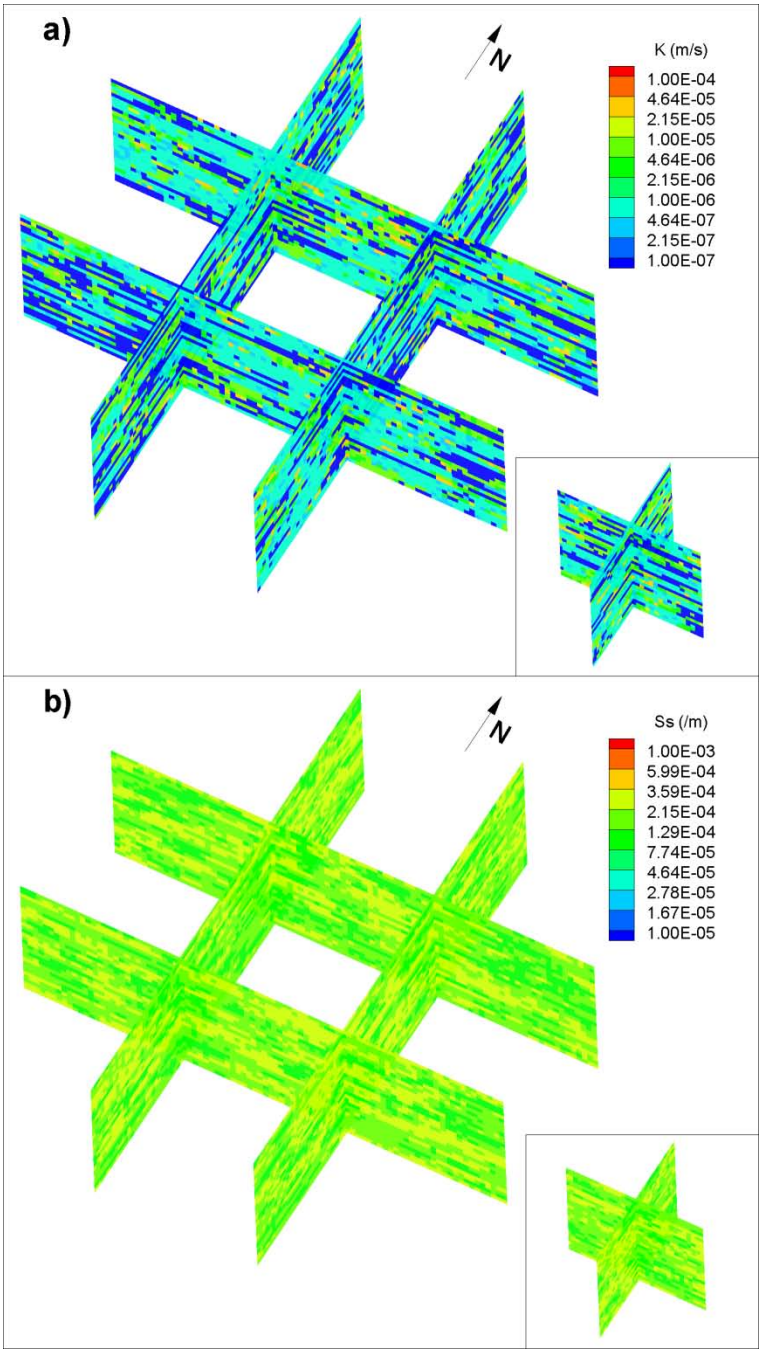


Figure 3.19. Three-dimensional perspective view of a single realization of K and S_s generated using the TPROGS code. The square created by the slices represents the outer edge of the 15 m x 15 m plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

Figure 3.20 shows the calibrated geological model (case 4), and Figure 3.21 shows the results of THT (case 5) starting as a simple layered model (similar to case 2 of section 3.3). For the stratigraphic model, 20 individual layers were assigned based on the material classifications of *Alexander et al.* [2011]. These layers were then calibrated for K and S_s by matching the simulated drawdown to that observed during the pumping test performed at PW1-3. Initial estimates of K were based on permeameter K data, and initial estimates of S_s were based on estimates made at individual ports by matching the drawdown response while assuming the domain to be homogeneous. The calibrated K field, (Figure 3.18a) shows the presence of the double aquifer system separated by an aquitard.

The THT K- and S_s-tomograms are quite similar to those presented in section 3.3 for case 2. The tomograms capture the presence of the double aquifer system, and the variance fields indicate that estimates K and S_s are best constrained in the central portion of the domain. Of note is that the variance is higher at the corners of the plot (the intersection of the slices) as no drawdown data was included from the wells at these locations (PW2, PW3, PW4, and PW5).

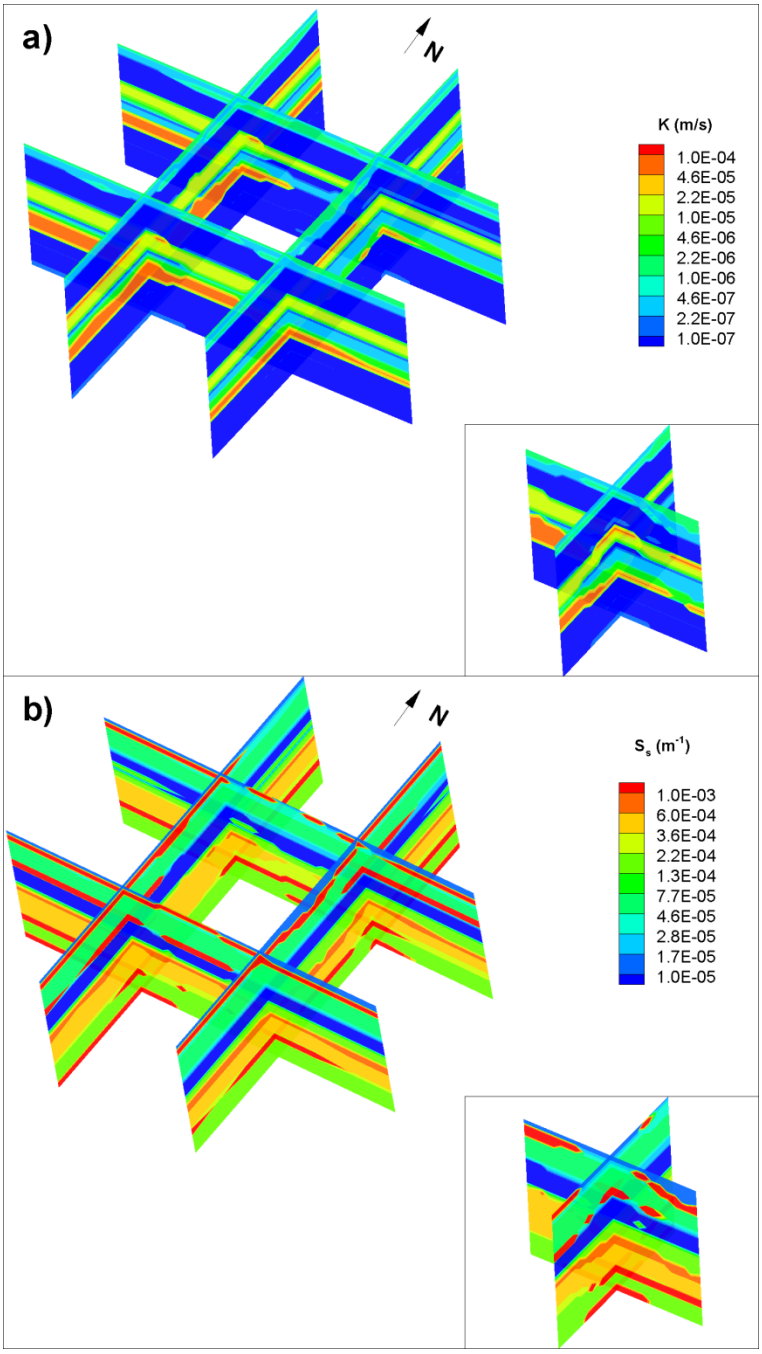


Figure 3.20. Calibrated *K* (a) and S_s (b) fields for the stratigraphic model. Note that the square formed by the slices in the main figure corresponds to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

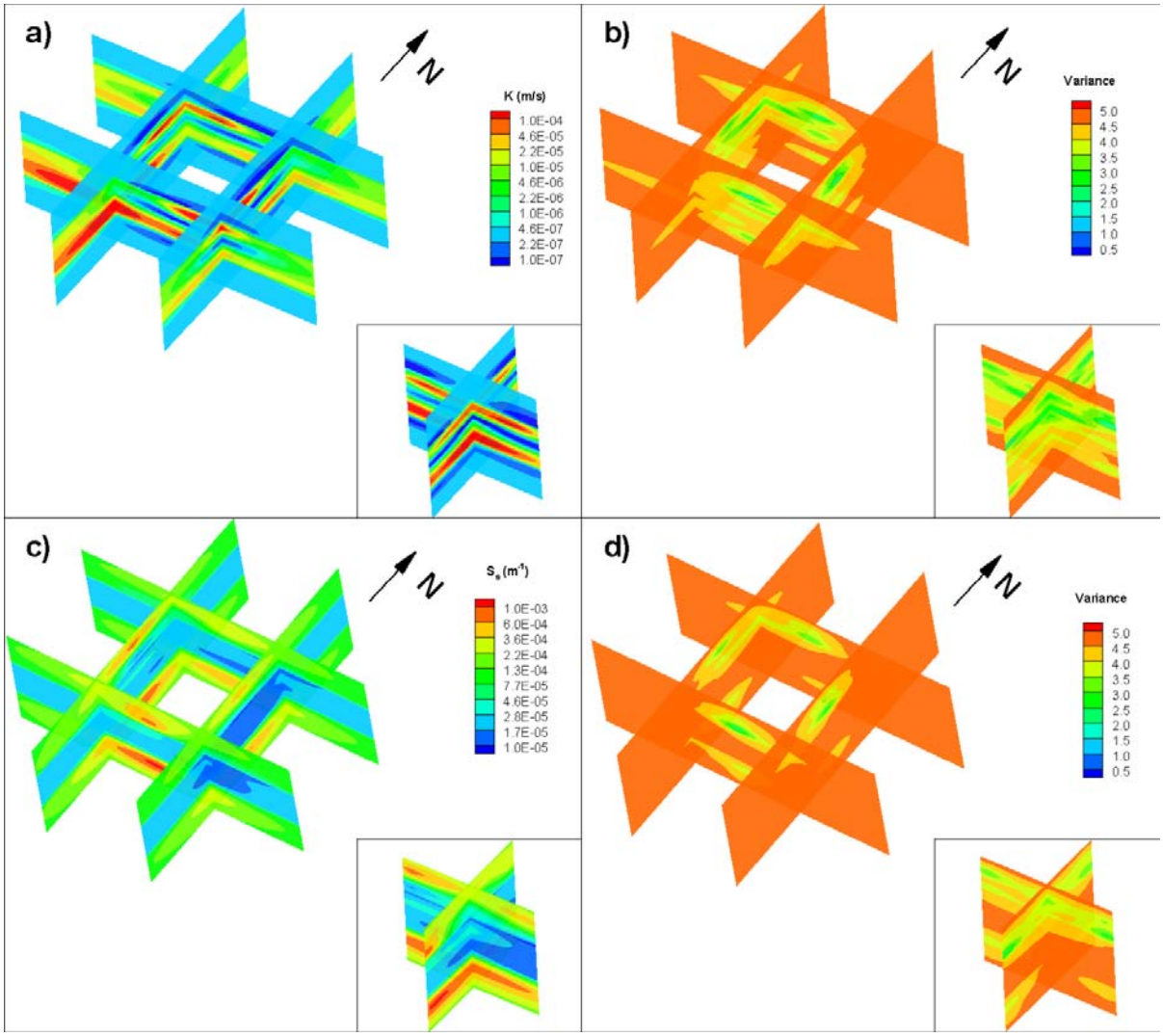
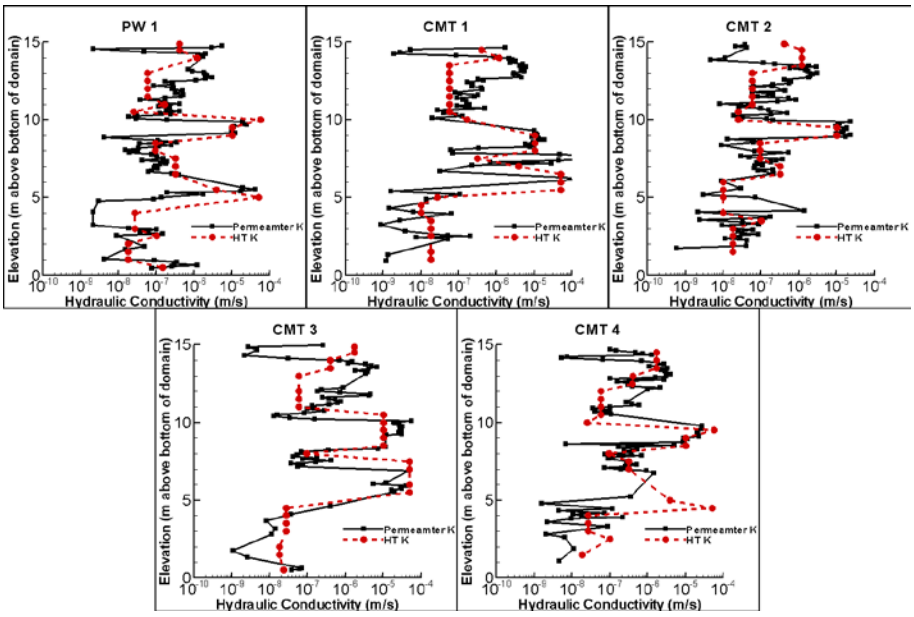


Figure 3.21. Case 4 THT *K*- and S_s-tomograms after the inclusion of 4 pumping tests (PW4-3, PW1-3, PW5-3, and PW3-3): a) *K*-tomogram; b) variance associated with the estimated *K* tomogram; c) S_s-tomogram; and d) variance associated with the estimated S_s-tomogram. Note that the square formed by the slices in the main figure corresponds to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

One method for assessing the performance of heterogeneous modeling approaches is to compare the modeled fields to known permeameter data. To do this, *K* values estimated along 5 boreholes (PW1 and CMTs 1-4) are compared to the permeameter data collected by *Alexander et al.* [2011]. This comparison is only made for cases 4 and 5 as cases 1 to 3 are either homogeneous, or are directly based on the permeameter data. Figures 3.22a and 3.22b show this comparison for cases 4 and 5, respectively. For this comparison, the estimated *K* values (red dashed line in Figure 3.22) are compared against measured permeameter values obtained by *Alexander et al.* [2011] along 5 boreholes (PW1, CMT1, CMT2, CMT3 and CMT4).

a)



b)

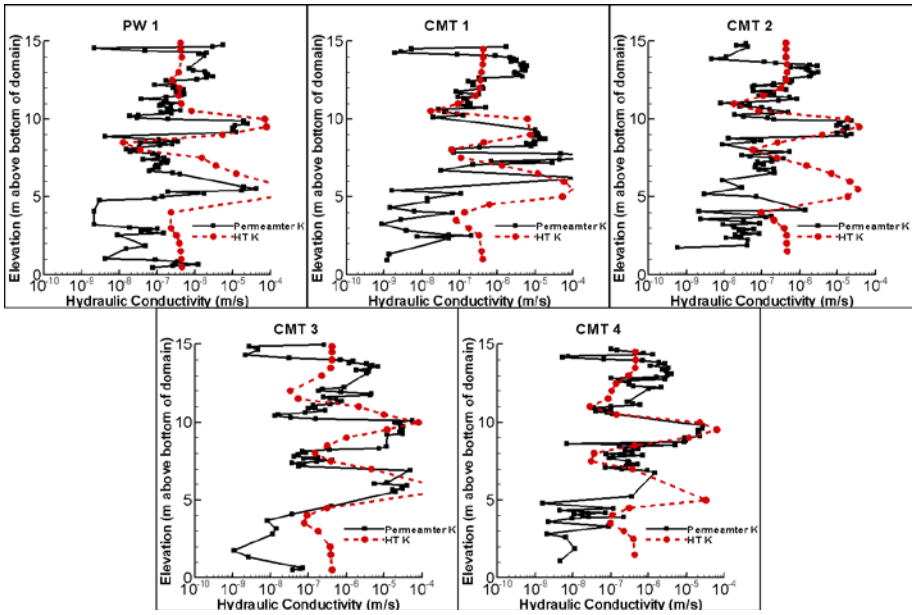


Figure 3.22. Vertical *K* profile along boreholes for the stochastic inversion of the pumping test at cases 4 (a) and 5 (b).

Examination of Figure 3.22a shows that the agreement between the *K*'s estimated by case 4 and the permeameter data are quite good. It should be reiterated at this point that the permeameter *K* data was used to provide the initial estimates of *K* for the parameter estimation procedure for the stratigraphic model. This helps to constrain the estimation procedure, particularly in the lower portion of the model domain where very little hydraulic data is available, and thus, parameter estimates for these layers would be uncertain. Figure 3.22b shows the comparison for case 5 (THT).

For this case, the *K*-tomogram captures the presence of the double aquifer zones located at approximately 5 and 10 m above the base of the model domain. However the *K* estimates made by case 5 for the aquifers is noticeably higher than the permeameter *K*.

Simulating pumping tests not used in the inversion procedure is an approach for validating heterogeneous models that has been used in other HT studies [*Illman et al.*, 2010; *Berg and Illman, in review (a);* and *Berg and Illman, in review (b)*]. For this comparison, all 9 pumping tests performed

at the NCRS were simulated using the heterogeneous models created for each case. The forward simulations were conducted using HGS [*Therrien et al.,* 2005] and the model domains described for each case (details in Appendix G).

For each case, all 9 pumping tests were simulated and the observed and simulated results were compared using scatterplots. For the purposes of this summary all cases are discussed, however, only the scatterplots for the two best performing cases (cases 4 and 5) are presented (as Figures 3.23 and 3.24 respectively). These scatterplots compare observed and simulated drawdowns at 5 times throughout the duration of the pumping tests (these times are different for each test, and range from early time to late time). The solid line is the 1:1 line indicating a perfect match between observed and simulated drawdown, the dashed line is a linear model fit to the data (the equation for this line is presented on each graph) and, R^2 is the coefficient of determination.

For case 1 (effective homogenous K and S_s) there was a significant amount of scatter in the plots, suggesting that these homogeneous fields, which do not honour the heterogeneous nature of the site, were poor representations of the true field, even in an average sense.

The scatterplots for cases 2 and 3 are quite similar to each other with the simulated drawdown generally being greater than the observed. This suggests that the geometric mean K of these model domains was too low. Since both of these model domains were populated with K data obtained from the permeameter analysis of soil cores, the most likely explanation is that the aquifer material was under represented in the sampling process. In fact, during the collection of soil cores, *Alexander et al.* [2011] reported that recovery averaged 80% for 4 of the wells (PW1, CMT1, CMT2, and CMT4), and 69% for CMT3. The lost material was often associated with the aquifer units [*Alexander et al.*, 2011], thus resulting in a biased estimate of K.

The drawdown scatterplots for the calibrated stratigraphic model (case 4) show a significant improvement in comparison to previous cases. The best matches are for the pumping tests at PW1-3

(the test to which the stratigraphic model was calibrated), and PW1-5. For the remaining 7 pumping tests, matches are improved in comparison to the previous cases. This indicates that for a layered system such as this, calibration of a stratigraphic model to a single pumping test may be able to reasonably predict the response of a system to pumping at other pumping locations especially when the stratigraphy (i.e., geology) is accurately captured.

Case 5 shows a slight improvement over case 4, with the best matches occurring for the pumping tests to which it was calibrated (PW4-3, PW1-3, PW3-3, and PW5-3). Poorer matches are seen for pumping tests not included in the inversion, in particular, PW1-4 and PW5-4 show considerable scatter.

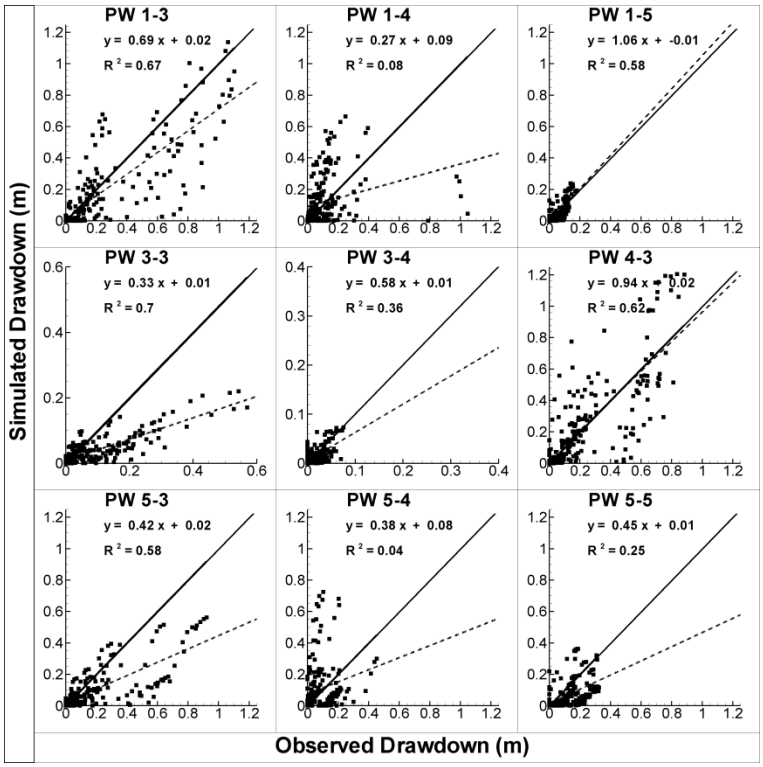


Figure 3.23. Scatterplot for all 9 pumping tests using the tomograms generated from Case 4 (calibrated stratigraphic model). The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

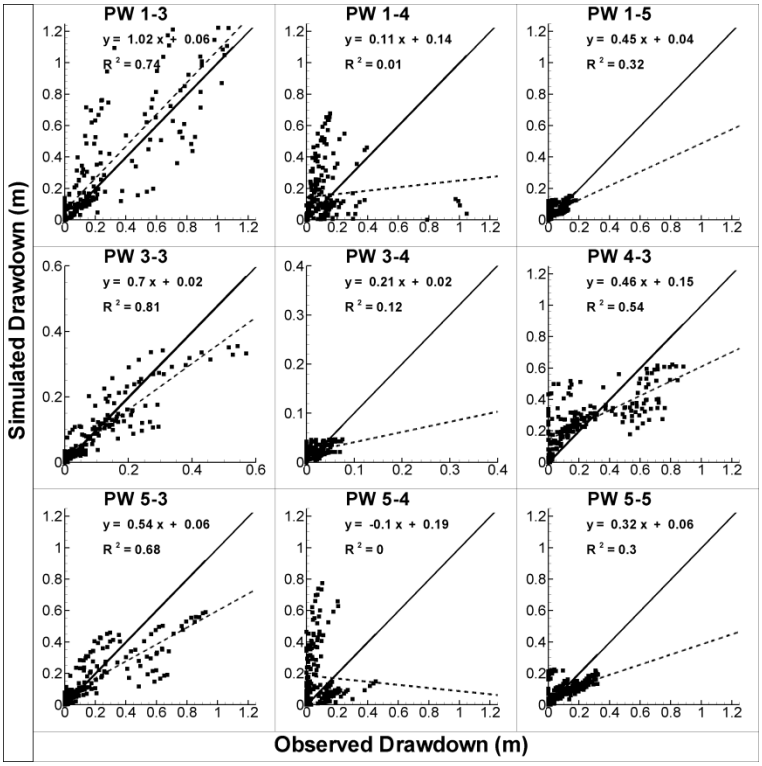


Figure 3.24. Scatterplot for all 9 pumping tests using the tomograms generated from THT Case 5. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

The main purpose of this study was to compare commonly used approaches for modeling subsurface heterogeneity in K and S_s fields. In particular five transient groundwater flow models consisting of different K and S_s fields that incorporate heterogeneity in various degrees of detail were created to represent the heterogeneous aquifer-aquitard system at the North Campus Research Site situated on the University of Waterloo. Methods compared include: 1) effective parameters; 2) ordinary kriging; 3) conditional simulation based on the Transition Probability/Markov Chain Geostatistical method; 4) geological modeling, and 5) transient hydraulic tomography. Site characterization data used to construct the various parameter fields included 471 estimates of K from the permeameter analyses of core samples and 4 pumping tests conducted in a tomographic manner.

The performance of these cases were then assessed by comparing the estimated hydraulic parameter fields to known geology (permeameter data) and through the numerical simulation of 9 pumping tests performed at the site. Drawdown values from these simulated pumping tests were then compared to the observed drawdowns using scatterplots. For each of the scatterplots, a linear model was fit through the data and the coefficient of determination (R^2) was calculated. These parameters give an indication of the goodness-of-fit between the observed and simulated drawdowns (R^2) and an indication of bias (the slope and intercept of the best fit line).

This study leads to the following key findings:

1) While estimating effective parameters (case 1) for relatively homogeneous systems may be useful and simpler to implement, the effective parameter approach when applied to the highly heterogeneous NCRS was unable to predict the response of the system to pumping. This point is particularly important for contaminated sites where an accurate understanding of heterogeneity is crucial in the prediction of contaminant migration or the design of remediation systems;

2) The interpolation of point data (case 2) through the use of ordinary kriging is a useful way of identifying key aquifer features and obtaining a 3D understanding of the aquifer system. However, when using this modeling approach for groundwater simulations, the data must be complete and accurate to avoid significant bias and error. For this particular study, the lack of samples associated with the key aquifer zones resulted in biased *K* estimates, and thus, caused the simulated drawdowns to be significantly greater than the observed. Additionally, estimates of specific storage are routinely not obtained from core samples thus making the prediction of transient head changes using a groundwater model difficult;

3) When a large amount of data are available, the use of geostatistical interpolators, such as ordinary kriging, is an attractive option. For this site, despite the availability of 471 *K* samples, the

approach was unsuccessful. This is likely a result of only having data along 5 boreholes, thus, the vertical characterization was relatively accurate, however, the horizontal extent of the heterogeneity was not well defined. This is a problem common to many sites, where a large number of boreholes are rarely available, resulting in a poor understanding of the lateral extent of aquifer materials;

4) While ordinary kriging honours the observed data, the transition between data points is smooth and may not accurately reflect the geology. Categorical interpolators such as the transition probability/Markov Chain approach can produce abrupt material changes. However, this approach produces many realizations and selecting the best case can pose a challenge. In addition, these categories must be assigned parameter values, which, if not accurate (e.g., obtained from incomplete core logs) may result in biased simulation results;

5) If parameter values estimated from soil cores are questionable, a stratigraphic model can be created and the layers calibrated to observed drawdowns. For case 4, this was done by creating a stratigraphic model based on the geological information fromt the site, and calibrating each layer for K and S_s to the observed drawdown during pumping at PW1-3. The resulting parameter fields were able to better predict pumping tests at the site than the previous cases. However, assigning layers in a model can be restrictive to the estimation process as the layer is assumed to be homogeneous, which may not be an accurate reflection of reality; and

6) For this study, the data requirements of THT were significantly less than cases 2-4 which required detailed core data (cases 2 and 3), as well as hydraulic data for calibration purposes (case 4). The data requirements for THT were a basic understanding of the geology and the hydraulic data from 4 pumping tests. In terms of data collection, THT requires significantly less effort. Additionally, since THT is calibrated to more pumping tests than the calibrated stratigraphic model, it is better able to represent the heterogeneity of the system. However, for this particular system, THT was unable to accurately predict all of the independent pumping tests, unlike previous findings by *Illman et al.*

[2010] and *Berg and Illman* [*in review* (a)] which suggests that, although hydraulic tomography is a powerful approach for imaging subsurface heterogeneity, additional research into its application in highly heterogeneous systems is required.

Chapter 4

Discussion

As long as aquifers have been investigated, efforts have been made to understand the importance of heterogeneity and how it is most effectively characterized/modeled. In many cases, simplifying assumptions are made that allow a heterogeneous system to be treated as homogeneous. These are commonly made so that analytical solutions can be used to interpret pumping test data to obtain effective values of K and S_s (e.g., *Theis* [1935]; *Hantush* [1960], and *Neuman* [1972]), or to allow solute transport to be simulated using effective parameters. Recently, through the use of numerical studies, it has been demonstrated that these effective parameter estimates provide little insight into the heterogeneous distribution of aquifer properties. In particular, for pumping tests performed in synthetic heterogeneous K and S_s fields, *Wu et al.,* [2005] found that estimates of K and S_s obtained using the *Theis* [1935] type-curve method varied both in time (duration of the test analyzed) and with the location of the observation port. At late time, K was found to converge on the geometric mean of the aquifer and S_s was dominated by the material between the pumped well and the observation location [*Wu et al.,* 2005].

For solute transport (particularly over large scales), small-scale heterogeneities in K are often up-scaled to produce an effective estimate of K for an equivalent homogeneous medium. Since the system is represented as homogeneous, the effects of heterogeneity on solute spreading are lost, and, the macrodispersivity concept is applied to account for this lost heterogeneity [e.g., *Gelhar and Axness,* 1983]. While this approach has been demonstrated to capture average plume behaviour over long flow paths (sufficient for the heterogeneity to be sampled by the plume) [e.g., *Sudicky,* 1986*; Sudicky et al.,* 2010], heterogeneity plays an important role at smaller scales [e.g., *Yeh et al.,* 1995*; McCarthy et al.,* 1996; *Zheng and Gorelick,* 2003]. In addition, to accurately estimate

macrodispersion requires intensive sampling of either small scale *K* data, or conducting a large scale tracer test. Both methods are often prohibitive with respect to time and money. As a consequence, new techniques are required to effectively and efficiently characterize/model aquifer heterogeneity. Recent advances in high resolution characterization (direct push techniques) and modeling (e.g., HT) techniques may allow for these smaller scale transport problems to be handled more effectively without the need for collecting a large number of samples.

Hydraulic tomography as a tool for characterizing and modeling the distribution of K and S_s in heterogeneous aquifers has been in development for over 20 years, with many of the laboratory and field demonstrations being published in the last 5 to 10 years. Despite the considerable interest in hydraulic tomography from the research community, many questions still remain. In particular, only one study has rigorously validated the tomogram estimated from a HT survey [*Illman et al.*, 2010]. The laboratory studies presented in this thesis extended the validation of HT from SSHT [*Illman et al.*, 2010] to THT (Section 2.1 and Appendix A). Additionally, alternative validation approaches were investigated, such as predicting solute transport (Section 2.2 and Appendix B) and predicting the flow of water to a well in a heterogeneous unconfined aquifer (Section 2.3 and Appendix C). These validations, prior to field scale studies, were critical to assess the ability of HT to map heterogeneity when all the forcing functions were well known and the sources of error were minimal. These laboratory studies demonstrated that when compared to other characterization/modeling efforts, HT is superior at predicting independent pumping tests, solute transport, and flow to a well in a heterogeneous unconfined aquifer.

Characterizing and modeling heterogeneous systems at the field scale is much more difficult than lab scale characterization as there are more sources of error, boundary conditions are poorly defined, and it is difficult to instrument a field site as densely as a sandbox aquifer. However, based on the laboratory studies, if the forcing functions are accurately known it should be possible to obtain

reasonable K and S_s estimates using THT. The highly heterogeneous nature of the glaciofluvial aquifer system at the NCRS makes heterogeneous characterization and modeling particularly difficult. As described by *Alexander* [2009] and *Alexander et al.* [2011], the vertical correlation length is approximately 0.15 m, K ranges over 5 orders of magnitude, and the variance of ln K based on permeameter samples is approximately 6.5. Using the data collected from soil core analysis [*Alexander et al.*, 2011] and multiple pumping tests, various heterogeneous modeling approaches were applied. These included: estimating effective parameters; kriging point scale permeameter K data; interpreting borehole data using a transition probability/Markov Chain approach; creating and calibrating a stratigraphic model based on core data; and performing THT (3 cases). Of these methods, the calibrated stratigraphic model and THT performed similarly, with THT performing slightly better. This assessment was made on the basis of comparison to known geology (core data) and the ability of these heterogeneous models to reproduce drawdowns observed during 9 pumping tests. The nature of the drawdown response at the NCRS (very small response in deep ports even at late times) created a high degree of uncertainty in parameter estimates for lower portions of the model domain. However, since the inverse method (SSLE) produces estimates of uncertainty (variance in the estimates of ln K and ln S_s) the operator is aware of the regions of higher uncertainty and can take steps to improve the estimates in these regions.

These field experiments demonstrated that characterizing and modeling highly heterogeneous aquifer systems can be very costly and time consuming. Characterization of subsurface heterogeneity and modeling efforts can be roughly separated into two categories; 1) those that rely on detailed borehole data, and 2) those that rely on pumping test data. Based on the work performed in this study a brief cost comparison is performed for a similar scale project. Where possible, real costs from this study are used, and labour is estimated at approximately $100/hr to better reflect the costs of industry. Tables 8 and 9 summarize the costs for categories 1 and 2 respectively.

For category 1 it is assumed that continuous soils cores are collected during the installation of fully screened wells. For category 2 multi-screen wells are installed without the collection of soil cores. Based on the assumptions, drilling costs are estimated to be approximately equal. Both installation methods are slower than a traditional well installation. For category 1 the collection of continuous core is a slow process, whereas, for category 2 alternating between bentonite and sand pack during backfilling is a slow process. The total estimated cost for drilling is approximately $60,000 and is based on the cost of installing all 9 wells at the NCRS.

The bulk of the costs for category 1 result from the laboratory analysis of soil cores, particularly permeameter analysis which can be a very slow process for lower *K* material. For this case, the estimated duration for sample analysis is based on the experience of *Alexander* [2009] who performed the core analysis presented in the field studies in this thesis. For category 2, the costs are spread across the instrumentation required for monitoring a large number of wells, the man hours required to perform multiple pumping tests, data processing, and inverse modeling.

Table 8. Cost estimate for heterogeneity characterization relying on point data (Category 1).

Detailed Characterization	**Estimated Costs**
Drilling (with complete core collection)	**$60,000.00**
Grain Size Analysis (300 Samples @ 10 samples/day)	**$24,000.00**
Permeameter Analysis (500 Samples @ 1 sample/hr)	**$50,000.00**
Data Processing (2 weeks)	**$8,000.00**
Statistical Analysis (2 weeks)	**$8,000.00**
Flow Modeling (2 weeks)	**$8,000.00**
Total	**$158,000.00**

While these estimates are very approximate, they suggest that implementing HT may not be cost prohibitive, especially in comparison to analyzing hundreds of core samples for characterization purposes. A potential advantage of HT is that some of these costs are one time purchases. For example, the instrumentation required for conducting the pumping tests and monitoring drawdown can be used at other sites. The same can be said for the computer cluster used for running the inverse model. Once purchased, this cluster can be used for many sites.

Table 9. Cost estimate for performing HT (Category 2).

Transient Hydraulic Tomography	**Estimated Costs**
Drilling (with complex well installation)	**$60,000.00**
Instrumentation (transducers, pumps, etc)	
Transducers (28 CMT, 6 for 2" wells)	**$30,000.00**
FLUTe Liners (with 5 transducers each)	**$36,000.00**
Pump-Packer System	**$5,000.00**
Electronics for Data Acquisition	**$8,000.00**
Conducting 9 x 24 hr Pumping Tests	**$27,000.00**
Inverse and Forward Modeling	
Data Processing (2 weeks)	**$8,000.00**
Inverse Modeling (2 weeks)	**$8,000.00**
Forward Modeling (2 weeks)	**$8,000.00**
PC Cluster for Modeling	**$20,000.00**
Total	**$210,000.00**

An advantage of HT is that it does not require the collection and analysis of soil cores, but rather relies on the data collected during pumping tests to estimate aquifer properties. Thus, it also provides accurate estimates of S_s which are critical in applications involving transient groundwater flow. Characterizing sites with HT before implementing remediation or treatment systems may allow for more efficient and effective cleanup operations. Thus, the costs are spent upfront to accurately

understand the site, minimizing issues that could arise later due to poor site characterization. Aside from contaminated sites, HT may also be a useful tool for other applications where heterogeneity is important, such as, mapping heterogeneity for: 1) aquifer storage and recovery (ASR) operations; 2) in situ leaching of metals; 3) assessing nuclear waste disposal sites; 4) mapping regional aquifer systems; and 5) delineating capture zones for municipal wells. Recently, new tomographic approaches for mapping basin scale heterogeneity have been proposed that use natural stimuli, such as; lightning, earthquakes, storm events, barometric changes, and changes in river-stage [*Yeh et al.*, 2008]. To date, promising numerical simulation results for river-stage tomography have been published [*Yeh et al.*, 2009].

Chapter 5

Conclusions and Recommendations

5.1 Summary of Conclusions

The main conclusions of this thesis are grouped into two categories; those from the laboratory studies, and those from the field studies.

Conclusions from the laboratory studies:

- Provided all of the forcing functions are known, THT is able to accurately map aquifer heterogeneity, and it performs better than other characterization/modeling approaches for predicting the response to independent pumping tests;
- *K* tomograms generated from HT analysis can predict the migration of a conservative tracer with reasonable accuracy, and based on the cases tested, performs better than other characterization techniques. The use of HT for mapping aquifer heterogeneity for predicting solute transport would result in cost savings as it is not necessary to collect a large number of small-scale samples, or conduct tracer tests to estimate effective transport parameters;
- For unconfined aquifers, the heterogeneous distribution of saturated parameters has a greater impact on drawdown responses than unsaturated parameters, corroborating the findings of *Mao et al*., [*in press*]. This suggests that, in order to accurately predict the response of an unconfined aquifer to pumping, characterization efforts should focus on saturated parameters (K and S_s); and,
- For a laboratory sandbox, with a simple heterogeneity pattern, HPTT can identify the location of a DNAPL pool by analyzing multiple partitioning tracer tests in a tomographic fashion.

Conclusions from the field studies are:

- It is possible to interpret individual or multiple pumping tests in a highly heterogeneous geologic medium using the THT code SSLE [*Zhu and Yeh*, 2005]. The approach yields highly resolved maps of K and S_s heterogeneity as well as their uncertainties;
- THT analysis of transient drawdown data from 4 pumping tests showed that the most salient features of the glaciofluvial aquifer-aquitard system can be captured with reasonable fidelity. In particular, the THT analysis detected the presence of 2 aquifers as well as the discontinuous aquitard between them. However, the THT analysis without the knowledge of site geology (case 1 from section 3.3) gave unsatisfactory results for regions with little drawdown, at the lower portion of the model domain, that was known to contain low K material;
- The THT analyses which started with K and S_s fields that better reflected the stratigraphy (case 2 from section 3.3) or populating the model with K values (case 3 from section 3.3) did not result in significantly improved results in terms of the predictability of independent pumping tests over Case 1(from Section 3.3). However, the stratigraphy along boreholes was better represented. This suggests that including information about the geology for these regions may produce more accurate K- and S_s-tomograms;
- When compared to other characterization techniques, THT performed the best in terms of capturing key aquifer features and predicting pumping test results;
- The efforts required for data collection for HT were significantly less than those methods that relied on the detailed analysis of many point scale samples;

- While still in the early stages of field implementation, THT appears to be a promising approach for characterizing aquifer heterogeneity for small scale plots. However, more work needs to be done to validate THT at a larger scale.

5.2 Recommendations for Future Research

Over the past 2 decades a considerable amount of research and development has been performed on HT. Despite this there are still many questions that need to be answered. Below are some recommendations for future areas of research:

- Up-scaling of HT for the prediction of solute transport to a large sandbox or a small scale field plot;
- Applying HT at the field scale for the prediction of flow to an unconfined well in a heterogeneous aquifer;
- Development of an inverse method for HT that estimates the heterogeneous distribution of both saturated and unsaturated parameters through multiple pumping tests;
- Demonstration of HPTT in a large 3D sandbox or at the source zone scale in the field;
- Application of THT to larger scale field sites;
- Development of new high resolution pressure measurement sensors that allow for many (10's of points) vertical observation points to be monitored along a single borehole; and
- Assessment of pumping strategies for HT, e.g., dipole pumping tests versus single hole.

Appendix A

Capturing Heterogeneity in Groundwater Flow Parameters: Comparison of Approaches Through Controlled Sandbox Experiments

This manuscript by *Berg, S.J. and W.A. Illman* is currently under review.

EXECUTIVE SUMMARY

Groundwater modeling has become a vital component to water supply and contaminant transport investigations. These models require representative hydraulic conductivity (K) and specific storage (S_s) estimates, or a set of estimates representing subsurface heterogeneity. Currently, there are a number of approaches for characterizing and modeling K and S_s heterogeneity in varying degrees of detail, but there is a lack of consensus for an approach that results in the most robust groundwater models with the best predictive. The main goal of this study is to determine which heterogeneity modeling approach (e.g. effective parameters, geostatistics, geological model, and hydraulic tomography) performs the best when input into a forward groundwater model and used to predict 16 independent pumping tests. We first characterize a sandbox aquifer through single-hole and cross-hole pumping tests, then use this data to construct forward groundwater models of various complexities (both homogeneous and heterogeneous distributions). Two effective parameter models are constructed: 1) by taking the geometric mean of single-hole test K and S_s estimates; and 2) calibrating effective K and S_s estimates by simultaneously matching the response at all ports during a cross-hole test. Heterogeneous models consist of spatially variable K and S_s fields obtained via: 1) kriging single-hole data; 2) calibrating a geological model; and 3) conducting transient hydraulic

tomography [*Zhu and Yeh*, 2005]. The performance of these fields are then tested through the simulation of 16 independent cross-hole pumping tests. Our results convincingly show that transient hydraulic tomography produces smallest discrepancy between observed and simulated drawdowns.

A1. INTRODUCTION

A1.1 Characterization methods of subsurface heterogeneity in hydraulic parameters

Subsurface characterization for groundwater investigations relies on the determination of the distribution of hydraulic parameters such as hydraulic conductivity (K) and specific storage (S_s). These values are then used to build groundwater models of various complexities to obtain quantitative estimates of hydraulic heads, groundwater fluxes, and the distribution and concentration of contaminants. Commonly, hydraulic parameters are estimated by collecting cores and subjecting them to permeameter tests and grain size analysis in a lab, or conducting slug, single-hole and/or pumping tests *in situ*. These *in situ* methods rely on analytical solutions that treat the geological medium to be homogeneous. These simplified solutions and the resulting estimated parameters have been used in a variety of real world applications and academic studies, despite the fact that the subsurface is heterogeneous at multiple scales. In particular, the knowledge of detailed three-dimensional distributions of K is critical for the prediction of contaminant transport, delineation of well catchment zones, and quantification of groundwater fluxes including surface-water/groundwater exchange. Even though many studies treat S_s to not vary significantly, in some formations, where the aquifer compressibilities vary significantly from one material type to the next (e.g., sands versus clays), S_s could vary several orders of magnitude. The characterization of subsurface heterogeneity is fraught with difficulties as numerous samples are required to delineate the variability of hydraulic parameters as well as their spatial correlations and connectivity. Using soil cores to accurately characterize the K heterogeneity of a site requires a large number of samples to be tested in the laboratory using a

constant or a falling head permeameter [e.g., *Sudicky*, 1986; *Sudicky et al*., 2010]. Alternatively, these samples are sieved to obtain grain size distributions, which can then be analyzed using various empirical relations to estimate *K*.

Characterizing the heterogeneity in S_s is seldom done as the parameter is considered to be less variable than *K* [e.g., *Gelhar*, 1993 and others]. In fact, data in the literature suggests this to be the case in both porous and fractured geologic media [e.g., *Meier et al*., 1998; *Sanchez-Vila et al*., 1999; *Illman and Neuman*, 2001; *Vesselinov et al*. 2001; *Martinez-Landa and Carrera*, 2005; *Illman and Tartakovsky*, 2007; *Liu et al*., 2007; *Willmann et al*., 2007; *Illman et al*. 2009], although the estimated variance is known to be dependent on the estimation method. In particular, *Sanchez-Vila et al*. [1999] showed that when one uses the Jacob's method to infer transmissivity (*T*) and storage coefficient (*S*), one obtains some "apparent" *T* which is constant and an "apparent" *S* which is highly variable in an aquifer, but in reality, *T* is variable and *S* is constant. In a different study, *Wu et al*. [2005] using numerical experiments and a first-order correlation analysis showed that the effective *T* and *S* for an equivalent homogeneous aquifer of Gaussian random *T* and *S* fields vary with time as well as the principal directions of the effective *T*. These studies suggest that the traditional interpretation of pumping tests by treating the medium to be homogeneous could potentially lead to biased estimates of hydraulic parameters and new methods for capturing subsurface heterogeneity are therefore necessary.

A1.2 Methods for capturing spatial heterogeneity in hydraulic parameters

Common approaches when mapping *K* (and less so S_s) heterogeneity are to use geostatistical or stochastic estimation techniques or more sophisticated interpolation methods. In particular, these approaches are considered to be the *de facto* standards which assume that a user-specified covariance function is valid and hydrogeologic parameters are log-normal and stationary. These assumptions are

difficult to satisfy in many geologic settings. Because of these assumptions, and when data are not abundant, stochastic estimation techniques may provide a smooth image of the spatial heterogeneity and may not represent the true distribution accurately. Although a variety of stochastic simulation techniques [e.g., *Deutsch and Journel,* 1998] exist that can overcome this issue of smoothing, it still does not address the preservation of many geological features. This is due to the fact that traditional geostatistical methods are based on variograms computed using two-point statistics. To overcome this shortcoming, multiple point geostatistics [e.g., *Guardiano and Srivastava*, 1993; *Caers*, 2001; *Strebelle*, 2002; *de Vries et al*., 2009] have been developed through the use of more complex point configurations, whose statistics are retrieved from training images that represent the geological facies distributions obtained from outcrop mappings and/or geophysical imaging.

Other approaches used to model subsurface heterogeneity include the Transition Probability Markov Chain method [*Carle and Fogg*, 1997; *Carle*, 1999; *Weissmann et al*., 1999] and the indicator kriging approach [*Journel*, 1983; *Journel and Isaaks*, 1984; *Journel and Alabert*, 1990; and *Journel and Gomez-Hernandez*, 1993]. Both of these approaches allow one to construct discontinuous facies models. However, the Markov model is better able to account for spatial cross-correlation, such as juxtapositional relationships, including the fining-upward tendencies of different facies, than the indicator approach [*De Marsily et al*., 2005].

More recently, geostatistical and stochastic inversion methods have received increasing attention. The approach produces the first and second statistical moments of hydrogeologic variables, representing their most likely estimates and their uncertainty, respectively, conditioned on available observations. Cokriging relies on the classical linear predictor theory that considers spatial correlation structures of flow processes (such as hydraulic head and velocity) and the subsurface hydraulic property, and cross-correlation between the flow processes and the hydraulic property. In the past few decades, many researchers [e.g., *Kitanidis and Vomvoris*, 1983; *Hoeksema and Kitanidi*s, 1984

and 1989; *Rubin and Dagan*, 1987; *Gutjahr and Wilson*, 1989; *Sun and Yeh*, 1992; *Harvey and Gorelick*, 1995; *Yeh et al.,* 1995 and 1996] have demonstrated its ability to estimate K, head, velocity, as well as solute concentrations in heterogeneous aquifers.

Recently, hydraulic tomography has been developed to obtain information on subsurface heterogeneity of K and S_s through sequential pumping tests. To our knowledge, *Neuman* [1987] was the first to suggest the approach using geophysical tomography as an analogy. Since then, various inverse methods have been developed for hydraulic tomography, which utilize pumping test data simultaneously or sequentially [e.g., *Gottlieb and Dietrich*, 1995; *Yeh and Liu*, 2000; *Bohling et al.*, 2002; *Brauchler et al.,* 2003; *McDermott et al.,* 2003; *Zhu and Yeh*, 2005, 2006; *Li et al*., 2005; *Fienen et al*., 2008; *Castagna and Bellin*, 2009; *Xiang et al*., 2009; *Liu and Kitanidis*, 2011]. Numerous laboratory [e.g., *Liu et al*., 2002; *Illman et al*., 2007, 2008; *Liu et al*., 2007; *Yin and Illman*, 2009; *Illman et al*., 2010] and field experiments [e.g., *Bohling et al*., 2007; *Straface et al*., 2007; *Illman et al*., 2009; *Cardiff et al*., 2009] have been conducted to show the utility of hydraulic tomography, but a rigorous study which compares the results to other more traditional characterization methods is generally lacking.

In the laboratory, *Illman et al*. [2010] recently assessed the performance of various methods for characterizing K estimates by predicting the hydraulic response observed in cross-hole pumping tests in a synthetic heterogeneous aquifer, and total flow rates obtained via flow-through tests. Specifically, they characterized a synthetic heterogeneous sandbox aquifer using various techniques (permeameter analyses of core samples, single-hole, cross-hole, and flow-through testing). They then obtained mean K estimates through traditional analysis of test data by treating the medium to be homogeneous. Heterogeneous K fields were obtained through kriging and steady state hydraulic tomography. To assess the performance of the each characterization approach, *Illman et al*. [2010] conducted forward simulations of 16 independent pumping tests and 6 steady-state, flow-through tests

using these homogeneous and heterogeneous K fields. The results of these simulations were then compared to the observed data. Results showed that the mean K and heterogeneous K fields estimated through kriging of small scale K data (core and single-hole tests) produced biased predictions of drawdowns and flow rates under steady-state conditions. In contrast, the heterogeneous K distribution or "K tomogram" estimated via steady state hydraulic tomography yielded excellent predictions of drawdowns of pumping tests not used in the construction of the tomogram and very good estimates of total flow rates from the flow-through tests. Based on these results, *Illman et al*. [2010] suggested that steady-state groundwater model validation is possible, if the heterogeneous K distribution and forcing functions (boundary conditions and source/sink terms) are characterized sufficiently.

A1.3 Goal of this study

This study extends the work of *Illman et al*. [2010] who examined only various K characterization approaches and their performance in predicting independent test data under steady state conditions. In particular, the main goal of this study is to extend the work of *Illman et al*. [2010] to the transient case. Using the same sandbox aquifer as *Illman et al*. [2010], we jointly assess the performance of various characterization and modeling techniques that treat the aquifer to be either homogeneous or heterogeneous through the prediction of independent, transient cross-hole pumping tests not used in the characterization effort. Specifically, we characterize the 2D heterogeneous aquifer using both single-hole and cross-hole pumping tests. These data are then used to construct various forward groundwater models with homogeneous and heterogeneous K and S_s estimates. Two homogeneous or effective parameter models are constructed: 1) by averaging local scale K and S_s estimates from single-hole pumping tests and treating the medium to be homogeneous; and 2) using MMOC3 [*Yeh et al*., 1993] coupled with PEST [*Doherty*, 1994] to estimate K and S_s by simultaneously matching the transient drawdown data from all ports during a cross-hole pumping test.

Three heterogeneous models are constructed and consist of spatially variable K and S_s fields obtained via: 1) kriging single-hole K and S_s data; 2) accurately capturing the layering and calibrating the K and S_s values for these layers using a parameter estimation program (i.e., a calibrated geological model); and 3) conducting transient hydraulic tomography. The performance of these homogeneous and heterogeneous K and S_s fields are then quantitatively assessed by simulating 16 independent cross-hole pumping tests and comparing the simulated drawdowns to the observed drawdowns.

A2. EXPERIMENTAL METHODS

A2.1 Sandbox and synthetic heterogeneous aquifer description

A two-dimensional synthetic heterogeneous aquifer was constructed in a sandbox measuring 193.0 cm in length, 82.6 cm in height, and 10.2 cm deep. Forty eight ports, 1.3 cm in diameter, were cut out of the stainless steel wall to allow coring of the aquifer as well as installation of fully penetrating horizontal wells (Figure A1). Each port is instrumented with a 0 to 1 psig *Setra model 209* pressure transducer.

The synthetic heterogeneous aquifer was created through the cyclic deposition of sediments under varying water flow and sediment feed rates. Our goal in relying on sediment transport was to create a more realistic heterogeneity pattern with various scales of heterogeneity in an efficient manner. Table A1 summarizes the grain size characteristics, K estimates, and S_s estimates of the sands used to create the heterogeneous aquifer, and the layers in which these sands occur. Figure A2 is a photograph of the synthetic heterogeneous aquifer showing the interfingering nature of the deposits and layer numbers. Further details to this synthetic heterogeneous aquifer and its construction approach are provided in *Illman et al*. [2010].

The aquifer system was bounded by three connected constant head boundaries (one situated at the top of the tank, and one and each end). The remaining boundaries (front, back, and bottom) were all no flow boundaries.

A2.2 Characterization of synthetic heterogeneous aquifer (single-hole pumping tests)

The synthetic heterogeneous aquifer was characterized using single-hole pumping tests to obtain K and S_s estimates at each of the 48 ports. Since the support scale of parameters estimated via single-hole tests is unknown, we assume the length of the well screen open to the aquifer is representative [e.g., *Guzman et al*., 1996, *Illman and Neuman*, 2000; *Illman*, 2005] of the support scale. The tests were conducted by pumping water at each port at a constant rate and monitoring the transient head change within the pumped well using a pressure transducer. A constant pumping rate (Q = 1.25 cm^3/sec) was set for each single-hole pumping test. For each test, data collection started without the pump running in order to obtain the initial hydraulic head in the sandbox at all measurement ports. A peristaltic pump was then activated at the pumping port and allowed to run at a constant rate until the development of steady state flow conditions. The entire transient head data response was matched using VSAFT2 [*Yeh et al*., 1993] through manual calibration by treating the aquifer to be homogeneous. VSAFT2 was chosen for the analysis as opposed to traditional type curve models because the numerical model is able to more accurately describe the sandbox geometry and boundary conditions. Details to the numerical modeling and calibration effort are provided in *Craig* [2005]. The single-hole K estimates ranged from 0.01 cm/s to 0.32 cm/s with a geometric mean of 0.06 cm/s and a variance ($\sigma^2_{\ln K}$) of 0.38, and single-hole S_s estimates ranged from 1.0 x 10^{-4} cm^{-1} to 5.5 x 10^{-3} cm^{-1} with a geometric mean of 6.1 x 10^{-4} cm^{-1} and a variance ($\sigma^2_{\ln S_s}$) of 0.97. Here, we consider the geometric mean of the 48 local K and S_s estimates from the single-hole pumping tests to

represent an effective K and S_s for the entire aquifer. We also considered the use of arithmetic and harmonic means as alternatives to the geometric mean. However, the arithmetic mean is representative of flow along stratification and the harmonic mean is representative of flow across layers. Since these single-hole tests induce flow along and across the layers we feel the geometric mean provides the most reasonable estimate of effective parameters for K and S_s.

A2.3 Characterization of the synthetic heterogeneous aquifer (cross-hole pumping tests)

Twenty five cross-hole tests were also performed in the sandbox for the purposes of effective parameter characterization, calibration of a geological model (presented in section 4), transient hydraulic tomography (presented in section 5) and model validation (presented in section 6). The tests were conducted at each port along columns 2 (ports 2, 8, 14, 20, 26, 32, 38, and 44) and 5 (ports 5, 11, 17, 23, 29, 35, 41, and 47) and 9 additional pumping tests at various ports outside of these two columns (ports 13, 15, 16, 18, 21, 37, 39, 40, and 42) (see Figure A1). The cross-hole tests were conducted by pumping at rates ranging from 2.50 - 3.17 cm^3/sec at 25 separate ports indicated by open and dashed squares on Figure A1. During each test, head measurements in all 48 ports (and the constant head reservoirs) were recorded. Pumping continued until the development of steady state conditions, which was determined by observing the stabilization of all head measurements within the aquifer.

The cross-hole test at Port 21 was used to estimate effective homogeneous values of K and S_s. Analogous to the analysis of single-hole data, the observation head data from this test were analyzed by manually calibrating each observation port data by VSAFT 2 and treating the aquifer to be homogeneous. Estimates of K and S_s obtained between the pumping and observation intervals, when the medium is treated to be homogeneous, are considered to be an equivalent hydraulic conductivity (K_{eq}) or specific storage (S_{seq}) (*Renard and de Marsily*, 1997; *Neuman*, 2005). Analysis of the cross-

hole test yielded 48 estimates of K and S_s for the equivalent homogeneous medium. The K_{eq} estimates ranged from 0.054 cm/s to 0.42 cm/s with a geometric mean of 0.11 cm/s and a variance ($\sigma^2_{\ln K}$) of 0.22. The corresponding S_{seq} estimates ranged from 8.5 x 10^{-5} cm^{-1} to 7.5 x 10^{-3} cm^{-1} with a geometric mean of 3.3 x 10^{-4} cm^{-1} and a variance ($\sigma^2_{\ln S_s}$) of 1.03.

The cross-hole pumping test at Port 20 is used to estimate effective homogeneous values of K and S_s. The parameter estimation program PEST [*Doherty*, 1994] was coupled with the groundwater flow model MMOC3 [*Yeh et al.,* 1993] to simultaneously match the transient data recorded at all ports. The synthetic aquifer used for the parameter estimation was discretized into 741 elements and 1600 nodes with element dimensions of 4.1 cm × 10.2 cm × 4.1 cm. Both sides and top boundaries were set to the same constant head, while the bottom, front, and back boundaries of the sandbox were considered no-flow boundaries. Estimated K and S_s values are 6.76 x 10^{-3} cm/s and 6.80 10^{-4} cm^{-1} respectively.

A3. GEOSTATISTICAL ANALYSIS OF SINGLE-HOLE PUMPING TEST DATA

Kriging and other simplified interpolation methods are commonly used to characterize subsurface heterogeneity [e.g., *Sudicky*, 1986; *Adams and Gelhar*, 1992; *Chen et al.*, 2000; and *Sudicky et al.,* 2010]. As such, we used kriging of single-hole K and S_s data to generate heterogeneous distributions that could be used for forward modeling of cross-hole pumping tests. The exponential variogram model was fit to the experimental variograms in both horizontal and vertical directions, resulting in an anisotropic variogram model. Examination of the experimental variogram of the ln-S_s data revealed an increasing variogram with lag distance. We attributed this to a trend of ln-S_s values declining from high to low values from the top to the bottom of the sandbox. Such a trend was also observed in the S_s values from a separate sandbox packed in an entirely different way [*Liu et al*.,

2007]. We detrended the data [e.g., *Chen et al*., 2000] by fitting an anisotropic exponential model to the residuals of the original experimental variograms for ln-S_s data.

Table A2 lists the variogram parameters fit to the experimental variograms. For both model fits, the anisotropy ratio was determined to be 3 with the horizontal correlation length larger than the vertical correlation length due the layered nature of the deposits. The results (not shown here), in general, reveal smoother K and S_s fields in comparison to the interfingering layers shown in Figure A2, which is expected considering that there are only 48 data points used for kriging (see Figure SA1 in the Supplementary information section). For the purposes of inclusion this thesis, Supplementary Figures are included at the end of this appendix as Figures SAx.

A4. CONSTRUCTION OF A GEOLOGICAL MODEL

Groundwater flow and transport models are commonly built using various hydrogeologic data and often deterministically incorporate the knowledge of site geology. To compare the performance of groundwater flow models based on the knowledge of geology to other models, we constructed a numerical model by using Figure A2 as a reference to construct a parameter field that closely resembles the stratigraphy of the synthetic heterogeneous aquifer.

To construct the geological model, we assumed that the stratification is known for the entire simulation domain. In practice, a perfect knowledge of stratigraphy is not available, thus we consider this to be a best case scenario in terms of having information on stratigraphy. We further assumed that the stratification shown on the glass (Figure A2) was uniform throughout the thickness of the sandbox.

The parameter estimation program PEST [*Doherty*, 1994] coupled with the groundwater flow model MMOC3 [*Yeh et al.*, 1993] was used to estimate K and S_s values for each layer using the data collected during the cross-hole test at Port 20. In total 38 parameters were estimated (K and S_s for 19

layers). Only 18 layers are identified in Figure A2, however, layer 5 which is discontinuous because of erosion that occurred during the deposition of layer 8 is treated as two separate layers in the PEST estimation. The model domain used for this estimation is identical to that described in section 2.3. Figure A3 shows the ln K and ln S_s (Figures A3a and A3b respectively) distribution for the calibrated geological model.

A5. CROSS-HOLE PUMPING TESTS AND TRANSIENT HYDRAULIC TOMOGRAPHY ANALYSIS

The transient hydraulic tomography analysis of cross-hole pumping tests in the sandbox was conducted using the Sequential Successive Linear Estimator (SSLE) code developed by *Zhu and Yeh* [2005]. The inverse model assumes a transient flow field and the natural logarithm of K (ln-K) and S_s (ln-S_s) are both treated as multi-Gaussian, second-order stationary, stochastic processes. The model additionally assumes that the mean and correlation structure of the K and S_s fields are known *a priori*. Further details to the SSLE code can be found in *Zhu and Yeh* [2005].

A5.1 Input parameters and cross-hole tests used

The model domain used to obtain K and S_s tomograms with transient hydraulic tomography is identical to that described in section A2.3 for the calibration of the geological mode.

Inputs to the inverse model include initial guesses for the K and S_s, estimates of variances and the correlation scales for both parameters, volumetric discharge (Q_n) from each pumping test where n is the test number, available point (small-scale) measurements of K and S_s, as well as head data at various times selected from the head-time curve. Although available point (small-scale) measurements of K and S_s can be input to the inverse model, we do not use these measurements to condition the estimated parameter fields.

For this THT analysis, the initial parameter fields are homogeneous and represented by the mean value of the K and S_s obtained from the analysis described in section A2.3 of a cross-hole test at port 21 which treats the medium to be homogeneous. Estimates of variance for K and S_s which have been shown to have negligible effects on the resulting tomogram [*Yeh and Liu*, 2000] were based on estimates from the available small scale data and used as our input variance in the inverse model. Correlation scales represent the average size of heterogeneity, which is difficult to determine accurately without a large number of data sets in the field. The effects of uncertainty in correlation scales on the estimate based on the tomography are negligible because the tomography produces a large number of head measurements, reflecting the detailed site-specific heterogeneity [*Yeh and Liu*, 2000]. Therefore, the correlation scales were approximated based only on the average thickness and length of the discontinuous sand bodies.

Prior to the incorporation into the inverse model the transient head records were treated with various error reduction schemes discussed in *Illman et al*. [2007], while data from pumped ports were not included into the inverse model because of excessive noise resulting from the use of a peristaltic pump. Each drawdown curve was then fit with a 5^{th} or 6^{th}–order polynomial curve following *Liu et al*. [2007]. A 5^{th} or 6^{th}-order polynomial was found to best capture the overall drawdown behaviour for the majority of the data. We then manually extracted 5 points representative of the entire transient record.

Data curves that could not be properly fit due to excessive noise were manually excluded from the analysis. In total, we used 8 independent cross-hole tests with pumping taking place at ports 47, 44, 35, 32, 17, 14, 5 and 2 for the analysis. More specifically, we used 5 data points from 47 ports totalling 235 data in all tests, except for the pumping test in port 2. At port 2, 5 data points were obtained from 43 ports totalling 215 data. Some of the data points were excluded from this particular

test as the data were excessively noisy. In total, we used 1860 data points from 8 different tests in our transient inversions.

A5.2 Computation of ln-K and ln-S_s tomograms

All computations for transient hydraulic tomography analyses were executed using 44 of 48 processors on a PC-cluster consisting (of 1 master and 12 slaves each with Intel Q6600 Quad Core CPU running at 2.4 GHz with 16 GB of RAM per slave) at the University of Waterloo. The operating system managing the cluster was CentOS 5.3 based on a 64-bit system. The total computational time for inverting data from 8 pumping tests was about 14 minutes. Figure A4a-d are the ln-K tomograms obtained by inverting the transient head data from 2, 4, 6, and 8 pumping test, respectively. Figure A4a shows that with only 2 pumping tests, a coarse picture of the heterogeneity pattern emerges, although the distribution is still pretty smooth and many details of the heterogeneity and in particular, details to the stratification are missing. As more tests are included into the SSLE algorithm, we see that more detail of the heterogeneity structure emerges. In particular, the final ln-K tomogram obtained (Figure A4d) using 8 pumping tests reveals considerable detail to the heterogeneity structure including the connectivity of various high and low K layers. Figure A4e is a ln-K tomogram computed using the steady state hydraulic tomography algorithm of *Yeh and Liu* [2000] by *Illman et al*. [2010] and is included for comparison purposes.

Figures A5a through A5d show the corresponding ln-S_s tomograms that were estimated simultaneously. In contrast to Figures A4a–A4d, the layering structure visible in the ln-K tomogram is not visible for the ln-S_s tomogram. However, a decreasing trend in ln-S_s with depth in the synthetic aquifer is apparent. Physically speaking, this makes sense because the sands in the upper portion are less compressed, while the deeper sands are more compressed due to the stress exerted by the overlying material. This finding suggests that ln-K values are not significantly correlated with the ln-

S_s values in this sandbox and is in agreement with those found by *Liu et al*. [2007] for a different sandbox packed with a considerably different heterogeneity pattern. The only data available for comparison are the single-hole estimates presented in Table A1 and shown as a kriged distribution in Supplementary Figure SA1b. While not identical, the pattern is similar, and the values are in a similar range, suggesting the S_s-tomogram is physically based.

A5.3 Statistical summary of results

Table A3 summarizes the mean, variance, and correlation lengths of the resulting ln-*K* tomogram. The estimated geometric mean (K_G) of the ln-*K* tomogram after including data from 8 cross-hole tests was 1.0×10^{-1} cm/s, while the estimated variance ($\sigma^2_{\ln K}$) was 1.32. We note that the value of K_G is identical to that estimated using steady state hydraulic tomography by *Illman et al*. [2010], but the $\sigma^2_{\ln K}$ is slightly higher compared to a value of 1.12. The estimated K_G from transient hydraulic tomography is somewhat higher than the estimate of K_G obtained by taking the geometric mean (0.06 cm/s) of the 48 local *K* values from single-hole tests. In contrast, the estimate of $\sigma^2_{\ln K}$ from transient hydraulic tomography ($\sigma^2_{\ln K} = 1.32$) is considerably higher than that estimated from the 48 single-hole *K* data ($\sigma^2_{\ln K} = 0.38$).

It is of interest to note that there is little change in the K_G and the correlation lengths of the ln-*K* tomograms as more tests are included in the inverse analysis. On the other hand, $\sigma^2_{\ln K}$ increases as more cross-hole tests are included into the inverse model. These results imply that with as few as 2 pumping tests, one could reliably estimate the K_G and the correlation lengths of the *K* distribution in the synthetic aquifer, however the accurate estimation of $\sigma^2_{\ln K}$ requires more cross-hole. A similar

trend in the improvement of estimates of geostatistical parameters was observed in their hydraulic tomography analysis of steady state head data by *Illman et al*. [2010].

Table A4 summarizes the mean, variance, and correlation lengths of the ln-S_s tomogram. The estimated geometric mean (S_{sG}) of the ln-S_s tomogram after including data from 8 tests was 9.08×10^{-5} cm^{-1}, while the estimated variance ($\sigma^2_{\ln S_s}$) was 0.76. The estimate of S_{sG} obtained by taking the geometric mean of 48 single-hole S_s estimates yields 6.1 x 10^{-4} cm^{-1} which is somewhat higher than the estimate obtained through transient hydraulic tomography. The $\sigma^2_{\ln S_s}$ estimate obtained from the single-hole data ($\sigma^2_{\ln S_s} = 0.97$) is close to that estimated through transient hydraulic tomography. The estimated correlation lengths in the horizontal (λ_x) and vertical directions (λ_z) appear to stabilize as additional tests are included in the analysis.

A5.4 Visual comparison of ln-*K* tomogram to the deposits

A visual comparison of the ln-*K* tomogram (Figure A4e) to the deposits (Figure A2) shows that many of the features are captured, although due to the intralayer variability in *K*, we do not expect a 1:1 correlation of the ln-*K* tomogram and the stratification seen in Figure A2. The comparison of the ln-*K* tomogram from transient hydraulic tomography to the kriged *K* distribution (available online as supplementary Figure SA1a) shows a marked difference in the *K* distribution. We notice that many of the features captured by the ln-*K* tomogram are captured by the kriged map, but the latter is distinctively smoother. In addition, the connectivity of the layers captured in the ln-*K* tomogram are not visible in the kriged ln-*K* field. Finally, we point out that the kriged ln-*K* distribution covers the midrange values of the ln-*K* tomogram revealing the kriged ln-*K* field produces a reasonable ln-*K* field in an average sense, but lacks the details in the heterogeneity pattern that the ln-*K* tomogram reveals.

A5.5 Comparison of ln-*K* tomograms: transient vs. steady state hydraulic tomography

We next compare the ln-*K* tomogram obtained using transient hydraulic tomography (Figure A4d) to the one obtained using steady state hydraulic tomography (Figure A4e) computed previously by *Illman et al*. [2010]. In both cases, 8 cross-hole tests are used for the inverse analysis. Comparison of the two figures shows that overall ln-*K* distributions from the two approaches are similar, however, more details are visible in the ln-*K* tomogram from transient hydraulic tomography. To facilitate a pixel-by-pixel comparison, we include a scatterplot (Figure A6) of ln-*K* values from Figure A4d and A4e. The dashed line indicates a perfect 1:1 correlation; the solid line is a linear model fit to the data; and, R^2 is the coefficient of determination. Results show that the data cluster around the 1:1 line indicating agreement between the two cases, however, we acknowledge that there is some scatter and bias in the estimates.

A6. PERFORMANCE ASSESSMENT OF RESULTS

A6.1 Performance assessment by simulating individual tests

The various effective and heterogeneous K and S_s fields are assessed by simulating 16 independent cross-hole pumping tests. These cross-holes tests are considered independent as they were not used by the various methods described earlier to characterize the aquifer (the exception is the cross-hole test performed at Port 20 which was used to calibrate the geological model). If a given characterization technique can capture the salient features of the true heterogeneity of the aquifer, then the resulting prediction of the independent pumping tests should be accurate. That is, the discrepancy between observed and simulated drawdown values should be small. In contrast, if the predicted drawdown values are inaccurate, then we consider the approach used to idealize the heterogeneity to be poor. In particular, we construct forward numerical models using: a) the effective

K and S_s estimates from single-hole tests; b) the PEST estimated effective parameters of K and S_s calibrated to the cross-hole test at port 20; c) the kriged K and S_s fields from single-hole tests; d) the PEST calibrated geological model; and e) the K and S_s tomograms from transient hydraulic tomography. For comparison purposes only the results from the best performing cases (d and e) are illustrated with figures in the manuscript. Tables A5 to A8 summarizing various performance metrics are provided to allow for a direct comparison of all five cases. Additionally, figures for all cases are provided in the online supplementary material section.

Figures A7 shows scatterplots of observed versus simulated drawdowns from independent cross-hole tests 18, 23, 40, and 42. The simulated drawdown values were obtained through numerical simulations using the calibrated geological model. Figure A7 includes a dashed line indicating a perfect 1:1 correlation, a solid line which is the linear model fit to the data, and the coefficient of determination (R^2). Data plotted on the Figure A7 are drawdown values from 0.5, 2, 5, and 10 seconds since the pumping test began. We see that for most cases the points cluster around the 1:1 line with some positive or negative bias as indicated by the slope of the linear model fit. This same pattern is seen for most of the other tests except for those performed near the top of the aquifer, where the simulated drawdown tends to be smaller than the observed suggesting that the estimated K of the upper layer (layer 18) is too high. The R^2 values for the 16 cases range from 0.002 to 0.86 with an arithmetic mean value of 0.65. The slope and the intercept of the linear model fit also provide an indication of bias. The improvement of groundwater flow model prediction accuracy seen through the incorporation of lithofacies in conjunction with model calibration is in agreement with the findings of *Sakaki et al*. [2009].

In contrast, Figure A8 shows a significant improvement in the predictions of drawdowns for the 4 selected independent cross-hole pumping tests. The K and S_s distributions for all the forward models presented in this figure were obtained by transient hydraulic tomography. The R^2 values for

the 16 cases range from 0.82 to 0.99 with an arithmetic mean of 0.94 indicating a marked improvement over the other modeling approaches. In addition, the comparison of the scatterplots in Figure A8 to those from Figure A7 clearly shows that transient hydraulic tomography is able to better predict independent cross-hole tests.

The results of this comparison clearly show that transient hydraulic tomography yields considerably better parameter estimates and thus is better able to predict the drawdown for 16 independent cross-hole pumping tests. Scatterplots of all 16 independent cross-hole tests are available online in the supplementary materials section for cases a) to e) as Figures SA2-SA6 respectively.

Table A5 summarizes the minimum, maximum, and the mean values of the slope and intercept of the linear model fit as well as the R^2 for all 16 tests for cases a) to e). This table shows that the slope of the linear model fit is quite variable ranging from 0.01 to 3.27 for all characterization approaches and the mean values range from 0.75 to 1.00. Likewise, the intercept of the linear model fit ranges from -0.76 to 0.27 with a mean ranging between -0.44 to 0.02. Finally, the R^2 values range from 0.002 to 0.99 with a mean ranging from 0.25 to 0.94. Examination of the slope, intercept, as well as the R^2 values across all the characterization methods shows that in an average sense over 16 independent cross-hole pumping tests, the K and S_s tomograms computed by transient hydraulic tomography yields the best predictions of drawdown data with least bias and scatter.

To further quantitatively assess the correspondence between the simulated and observed drawdown values, we compute the mean absolute error (L_1) and the mean square error (L_2) norms of all cases examined. The L_1 and L_2 norms are computed as:

$$L_1 = \frac{1}{n}\sum_{i=1}^{n}\left|\chi_i - \hat{\chi}_i\right| \qquad [1]$$

$$L_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\chi_i - \hat{\chi}_i\right)^2 \qquad [2]$$

where n is the total number of drawdown data, i indicates the data number, and χ_i and $\hat{\chi}_i$ represent the estimates from the simulated and measured drawdowns, respectively. The L_1 and L_2 norms were calculated for each case by evaluating the observed and simulated drawdowns at 4 times (0.5, 2, 5, and 10 seconds) at each port, except for the port that was pumped due to excessive noise. Thus, each L_1 and L_2 norm represents 188 observations.

Table A6 summarizes the L_1 norm, while Table A7 summarizes the L_2 norm calculated for all the cases. The cells of each entry in the table are color-coded. The minimum value in the table is assigned a color of dark green, the maximum value a color of dark red, and the median value a color of yellow. Values intermediate to these anchor points are assigned appropriate intermediate colors. Both Tables A6 and A7 show that forward simulations using K and S_s tomograms from transient hydraulic tomography, consistently yields the lowest L_1 and L_2 norms for all 16 independent cross-hole pumping tests, suggesting that the approach yields the best predictions.

The scatterplots for cases a) to e) were also analyzed in an ensemble sense to see if the homogeneous cases were able to estimate the average behaviour of all of the 16 cross-hole tests. This comparison is presented for Table A8 and includes; L_1 norm, L_2 norm, slope, intercept, and R^2 for all cases when the data from all 16 cross-hole tests are analyzed collectively. The trends seen in the individual scatterplots are also seen in these ensemble scatterplots. The results for transient hydraulic tomography after 2, 4, 6, and 8 tests are also presented in Table A8. While not changing significantly with the inclusions of additional cross-hole tests, the matches do improve slightly.

A6.2 Predictability of transient drawdown curves

Finally, to further illustrate the robustness of transient hydraulic tomography, we also plot simulated and observed drawdown curves for a cross-hole test performed at port 40. Again only the

cases d) and e) are included in the manuscript. Full figures showing the comparison at all 48 ports are available online in the supplementary materials section as Figures SA7 to SA11 for cases a) to e) respectively. In particular, Figures A9 and A10 show double logarithmic plots of observed (small dots) and simulated (curves) drawdown records at 16 selected ports. In Figure A9, the simulated drawdown values are obtained through numerical simulations using the calibrated geological model, while in Figure A10, simulated drawdown values are obtained through numerical simulations with the K- and S_s-tomograms.

Figure A9 shows that the calibrated geological model does reasonable job of predicting drawdown at most of the ports, however, drawdown is significantly under estimated in the upper ports. This again suggests that the upper layer (layer 18) is not accurately characterized. In contrast, Figure A10 shows a drastically improved result when the simulated values are obtained from the forward simulation of the pumping test using the K and S_s tomograms from transient hydraulic tomography. Examination of Figure A10 shows that the match is not perfect, but overall, the drawdown curves are captured very well throughout the duration of the pumping test, which the other approaches failed to accomplish.

A7. SUMMARY AND CONCLUSIONS

We characterized and modeled a synthetic heterogeneous aquifer using 2 effective parameter approaches and 3 separate methods that consider the heterogeneity of the aquifer. Two effective parameter models are constructed: 1) by taking the geometric mean of single-hole test K and S_s estimates obtained by treating the aquifer to be homogeneous; and 2) using MMOC3 [*Yeh et al*., 1993] coupled with PEST [*Doherty*, 1994] to obtain effective K and S_s estimates by matching the transient drawdown data simultaneously from all ports during a cross-hole test. Heterogeneous models consist of spatially variable K and S_s fields obtained via: 1) kriging single-hole K and S_s data;

2) accurately capturing the layering (i.e., geological model) and calibrating K and S_s estimates of these layers with MMOC coupled to PEST; and 3) conducting transient hydraulic tomography [*Zhu and Yeh*, 2005].

The performance of these homogeneous and heterogeneous K and S_s fields were then tested through the forward numerical simulation of cross-hole pumping tests that were not used in the characterization effort. Drawdown values from 16 cross-hole pumping tests conducted in the sandbox aquifer were then directly compared through scatter plots. The comparison was done for individual tests and also for all 16 tests together. A linear model was fit to each of the scatter plots and the coefficient of determination (R^2) computed to quantitatively assess the goodness-of-fit between the observed and simulated drawdown values. The slope and intercept of the linear model fit provide information on prediction bias.

We found that the forward numerical simulations using the two homogeneous cases could not predict the 16 cross-hole pumping tests accurately for almost all tests and significant bias was shown when the observed and simulated drawdowns were compared through a series of scatter plots. The scatter and bias were evident in not just individual tests, but also when the data from all 16 tests were plotted together and examined in an ensemble sense.

The heterogeneous models were also tested in a similar fashion. We found that the forward simulations of 16 independent cross-hole pumping tests using the kriged K and S_s fields were unable to accurately match the observed drawdowns. The calibrated geological model showed a significant improvement, however, it did not seem to accurately represent the upper most layer, resulting in very poor matches in this region. It should be emphasized that for the calibrated geological model, the stratigraphy was assumed to be perfectly known. In reality, such knowledge is impossible to obtain. Therefore, the constructed and calibrated geological model should be considered to be a best case scenario.

The forward numerical modeling of the 16 tests using the ln-K and ln-S_s tomograms computed via transient hydraulic tomography convincingly showed that these distributions led to the smallest discrepancy between observed and simulated drawdowns. Our sandbox experiments demonstrate the robustness of transient hydraulic tomography conducted in a controlled environment. The repeated calibration from the inclusion of data from multiple cross-hole pumping tests leads to improved heterogeneity characterization of the connectivity of hydraulic parameters. The accurate estimation of hydraulic parameters and their distribution was critical for the accurate prediction of independent cross-hole pumping tests. We are optimistic about the capabilities of hydraulic tomography, but are also cautious in generalizing our sandbox results to the field scale. It remains to be seen whether hydraulic tomography will yield robust results under field conditions. We are currently conducting a comprehensive field assessment of hydraulic tomography and other heterogeneous characterization and modeling methods at the University of Waterloo. These results will be reported in the near future.

Table A1: Characteristics of each layer used to create a synthetic heterogeneous aquifer (after *Illman et al*., 2010).

Layer	Sand	d_{50} (mm)	K (cm/s) Shepherd	Core Permeameter K (cm/s)*	Single-hole K (cm/s)*	Single-hole S_s (cm/s)*
1	20/30	0.75	1.03×10^{-1}	3.20×10^{-2}	5.32×10^{-2}	2.12×10^{-4}
2	4030	0.35	2.99×10^{-2}	5.29×10^{-2}	5.67×10^{-2}	2.60×10^{-4}
3	F-85	0.15	7.29×10^{-3}	7.14×10^{-2}	5.70×10^{-2}	5.00×10^{-4}
4	20/40	0.58	6.68×10^{-2}	5.68×10^{-2}	5.10×10^{-2}	2.22×10^{-4}
5	mix	0.46	N/A	N/A	N/A	N/A
6	mix	0.46	N/A	8.16×10^{-2}	5.00×10^{-2}	4.00×10^{-4}
7	#12	0.52	5.70×10^{-2}	1.27×10^{-1}	7.35×10^{-2}	4.20×10^{-4}
8	F32	0.5	5.33×10^{-2}	1.34×10^{-1}	4.50×10^{-2}	1.75×10^{-4}
9	20/40	0.58	6.68×10^{-2}	8.69×10^{-2}	4.60×10^{-2}	2.15×10^{-4}
10	F-65	0.2	1.20×10^{-2}	1.13×10^{-1}	8.25×10^{-2}	1.14×10^{-3}
11	#12	0.52	5.70×10^{-2}	1.37×10^{-1}	2.05×10^{-1}	2.15×10^{-4}
12	16/30	0.87	1.32×10^{-1}	3.40×10^{-2}	4.95×10^{-2}	6.32×10^{-4}
13	20/30	0.75	1.03×10^{-1}	2.60×10^{-1}	1.05×10^{-1}	9.80×10^{-4}
14	F-75	0.17	9.22×10^{-3}	9.79×10^{-2}	5.70×10^{-2}	9.80×10^{-4}
15	20/40	0.58	6.68×10^{-2}	8.58×10^{-2}	7.50×10^{-2}	2.00×10^{-3}
16	mix	0.46	N/A	4.16×10^{-2}	2.69×10^{-2}	7.11×10^{-4}
17	F-85	0.15	7.29×10^{-3}	4.51×10^{-2}	4.47×10^{-2}	1.14×10^{-3}
18	20/30	0.75	1.03×10^{-1}	1.45×10^{-1}	1.16×10^{-1}	3.38×10^{-3}

*** If multiple ports are in the same layer then the geometric mean is presented**

Table A2: Geostatistical model parameters for kriging single-hole ln-K and ln-S_s data.

Data	Model	Nugget	Range (cm)	Sill	Anisotropy ratio
Single-hole ln-K	Exponential	0	30	0.40	3
Single-hole ln-S_s	Exponential	0	30	0.45	3

Table A3: Statistical properties of the estimated ln-*K* tomograms.

Case	$\overline{\ln K}$ (K ~ cms^{-1})	$\sigma^2_{\ln K}$	λ_x (cm)	λ_z (cm)
Hydraulic tomography (2 tests)	-2.42 (0.09)	0.62	25	10
Hydraulic tomography (4 tests)	-2.28 (0.10)	0.95	20	10
Hydraulic tomography (6 tests)	-2.28 (0.10)	1.17	24	12
Hydraulic tomography (8 tests)	-2.30 (0.10)	1.32	24	12

Table A4: Statistical properties of the estimated ln-S_s tomograms.

Case	$\overline{\ln S_s}$ (Ss ~cm^{-1})	$\sigma^2_{\ln S_s}$	λ_x (cm)	λ_z (cm)
Hydraulic tomography (2 tests)	-9.49 (7.55E-05)	0.15	15	15
Hydraulic tomography (4 tests)	-9.57 (6.98E-05)	0.19	40	20
Hydraulic tomography (6 tests)	-9.53 (7.30E-05)	0.27	34	17
Hydraulic tomography (8 tests)	-9.31 (9.08E-05)	0.76	38	19

Table A5: Statistics of the linear model fit and correlation of determination (R^2).

Case		Slope		Intercept			R^2		
	min	max	mean	min	max	mean	min	max	mean
Geometric mean K/S_s (single hole)	0.51	1.50	0.79	-0.25	0.09	-0.01	0.14	0.66	0.53
PEST Effective K/S_s	0.72	3.27	1.28	-0.76	-0.24	-0.44	0.08	0.38	0.25
Kriged K & S_s	0.62	1.50	0.87	-0.27	0.09	-0.13	0.17	0.73	0.58
Calibrated Geological Model	0.01	1.44	0.96	-0.25	0.27	0.02	0.002	0.86	0.65
Transient Hydraulic Tomography	0.87	1.21	1.00	-0.04	0.09	0.02	0.82	0.99	0.94

Table A6: L1 norms of observed versus simulated drawdowns from cross-hole tests 8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42.

	Port 8	**Port 11**	**Port 13**	**Port 15**	**Port 16**	**Port 18**	**Port 20**	**Port 23**	**Port 26**	**Port 29**	**Port 37**	**Port 38**	**Port 39**	**Port 40**	**Port 41**	**Port 42**
Geometric mean K/S_s (single hole)	0.20	0.31	0.23	0.35	0.40	0.30	0.61	0.39	0.62	0.57	0.51	0.66	0.58	0.76	0.65	0.42
PEST Effective K/S_s	0.39	0.62	0.48	0.65	0.78	0.57	0.91	0.72	0.91	0.89	0.81	0.99	0.93	1.18	0.97	0.67
Kriged K & S_s	0.19	0.33	0.23	0.34	0.38	0.30	0.56	0.37	0.55	0.50	0.48	0.57	0.48	0.65	0.53	0.36
Calibrated Geological Model	0.17	0.22	0.38	0.31	0.35	0.22	0.27	0.32	0.48	0.36	0.44	0.39	0.39	0.44	0.39	0.30
Transient Hydraulic Tomography	0.06	0.11	0.12	0.11	0.11	0.07	0.13	0.17	0.17	0.16	0.07	0.13	0.15	0.10	0.10	0.19
Max	1.18															
Min	0.06															

TableA7: L2 norms of observed versus simulated drawdowns from cross-hole tests 8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42.

	Port 8	Port 11	Port 13	Port 15	Port 16	Port 18	Port 20	Port 23	Port 26	Port 29	Port 37	Port 38	Port 39	Port 40	Port 41	Port 42
Geometric mean K/S_s (single hole)	0.14	0.18	0.13	0.26	0.25	0.16	0.58	0.25	0.61	0.57	0.50	0.81	0.61	1.00	0.78	0.37
PEST Effective K/S_s	1.97	2.06	1.92	2.26	1.99	1.89	2.25	1.97	2.09	2.24	2.08	2.50	2.28	2.82	2.25	1.71
Kriged K & S_s	0.12	0.18	0.14	0.24	0.24	0.16	0.49	0.24	0.49	0.45	0.43	0.61	0.43	0.75	0.53	0.27
Calibrated Geological Model	0.09	0.19	0.83	0.15	0.23	0.31	0.37	0.17	0.39	0.25	0.47	0.32	0.34	0.35	0.38	0.31
Transient Hydraulic Tomography	0.01	0.02	0.04	0.03	0.04	0.02	0.03	0.06	0.05	0.05	0.01	0.03	0.06	0.02	0.02	0.10
Max	2.82															
Min	0.01															

Table A8: Statistics of the linear model fit, correlation of determination (R^2), and L_1 and L_2 norms for the ensemble analysis of all cases (including transient hydraulic tomography after the inclusion of 2, 4, 6, and 8 tests)

Case	Slope	Intercept	R^2	L_1	L_2
Homogeneous K Field					
Geometric mean K/S_s (single hole)	0.61	-0.02	0.44	0.48	0.53
PEST Effective K/S_s	0.86	-0.25	0.20	0.78	2.14
Heterogeneous K Field					
Kriged K & S_s	0.73	-0.06	0.52	0.43	0.45
Calibrated Geological Model	0.99	0.18	0.67	0.34	0.32
Transient hydraulic tomography (Sequential - 2 tests)	1.06	0.01	0.92	0.17	0.07
Transient hydraulic tomography (Sequential - 4 tests)	1.03	0.04	0.95	0.13	0.04
Transient hydraulic tomography (Sequential - 6 tests)	1.05	0.06	0.96	0.14	0.04
Transient hydraulic tomography (Sequential - 8 tests)	1.01	0.02	0.95	0.12	0.04

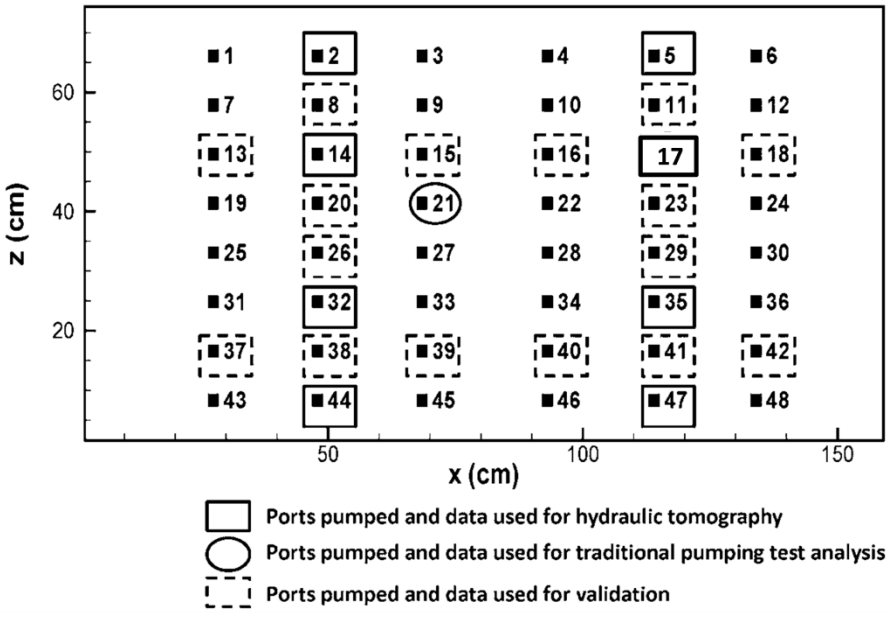


Figure A1: Schematic diagram of synthetic heterogeneous aquifer used for validation of transient groundwater flow models. Numbers next to solid squares indicate port numbers, open squares around numbers indicate the 8 ports (2, 5, 14, 17, 32, 35, 44, 47) used for hydraulic tomography, the open oval indicates the port (21) at which pumping test data was analyzed using VSAFT2 by treating the medium to be homogeneous (i.e., traditional pumping test analysis), and the dashed open squares around the 16 other

ports (8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42) indicate the pumping locations for the independent cross-hole pumping tests used for validation purposes [after *Illman et al*., 2010].

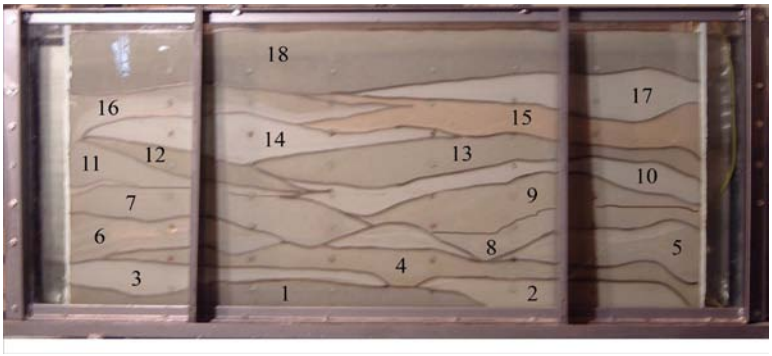


Figure A2: Photograph of synthetic heterogeneous aquifer created via cyclic flux of sediment-laden water [after *Illman et al*., 2010].

a)

b)

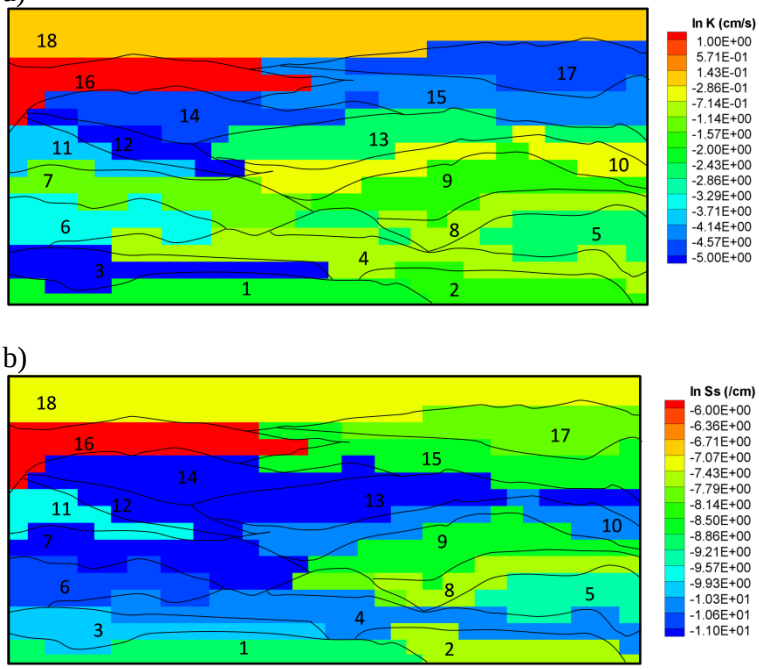
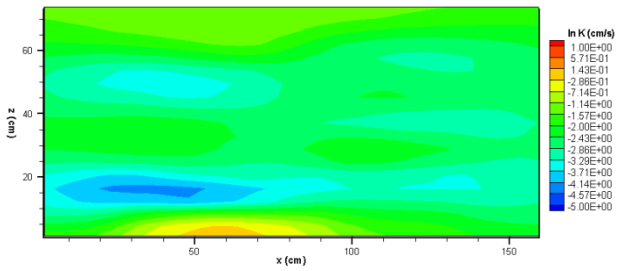
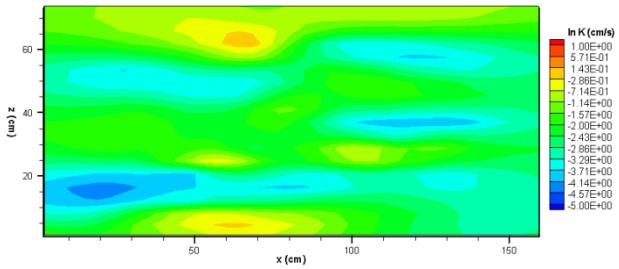


Figure A3: Distribution of a) ln-K and b) ln-S_s values for the calibrated geological model.

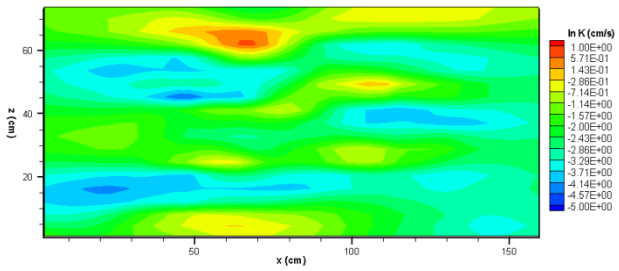
a)


ln K (cm/s)
1.00E+00
5.71E-01
1.43E-01
-2.86E-01
-7.14E-01
-1.14E+00
-1.57E+00
-2.00E+00
-2.43E+00
-2.86E+00
-3.29E+00
-3.71E+00
-4.14E+00
-4.57E+00
-5.00E+00
z (cm)
x (cm)

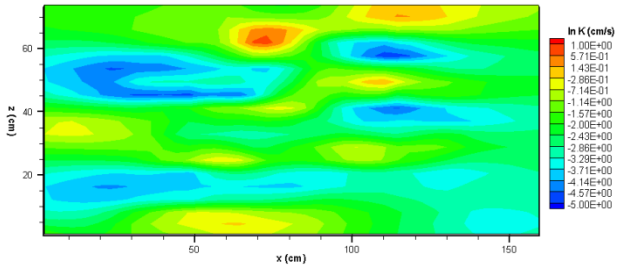
b)


ln K (cm/s)
1.00E+00
5.71E-01
1.43E-01
-2.86E-01
-7.14E-01
-1.14E+00
-1.57E+00
-2.00E+00
-2.43E+00
-2.86E+00
-3.29E+00
-3.71E+00
-4.14E+00
-4.57E+00
-5.00E+00
z (cm)
x (cm)

c)


ln K (cm/s)
1.00E+00
5.71E-01
1.43E-01
-2.86E-01
-7.14E-01
-1.14E+00
-1.57E+00
-2.00E+00
-2.43E+00
-2.86E+00
-3.29E+00
-3.71E+00
-4.14E+00
-4.57E+00
-5.00E+00
z (cm)
x (cm)

d)


ln K (cm/s)
1.00E+00
5.71E-01
1.43E-01
-2.86E-01
-7.14E-01
-1.14E+00
-1.57E+00
-2.00E+00
-2.43E+00
-2.86E+00
-3.29E+00
-3.71E+00
-4.14E+00
-4.57E+00
-5.00E+00
z (cm)
x (cm)

e)

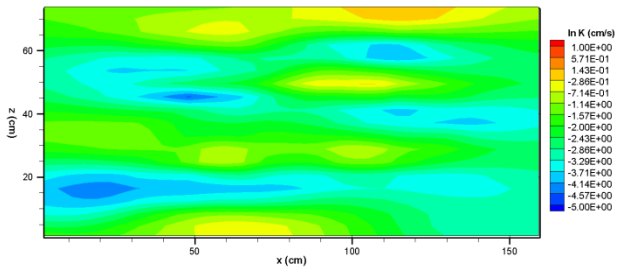
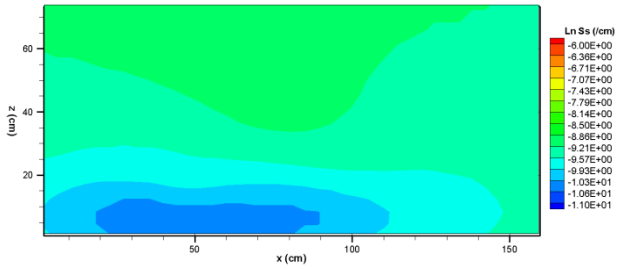
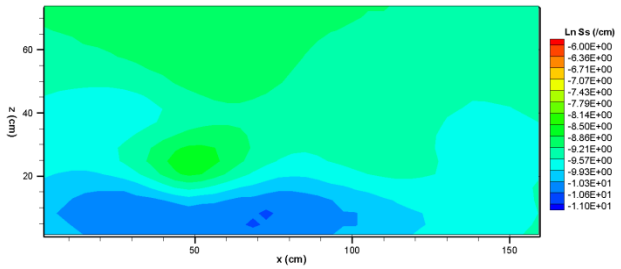


Figure A4: ln-*K* tomograms computed using: a) 2 pumping tests (ports 47 and 44); b) 4 pumping tests (ports 47, 44, 35, and 32); c) 6 pumping tests (ports 47, 44, 35, 32, 17, and 14); d) 8 pumping tests (ports 47, 44, 35, 32, 17, 14, 5 and 2) and e) 8 pumping tests (ports 47, 44, 35, 32, 17, 14, 5 and 2) but interpreted with the steady state hydraulic tomography algorithm (after *Illman et al*., 2010).

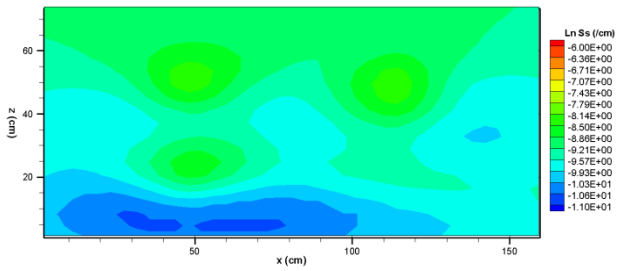
a)



b)



c)



d)

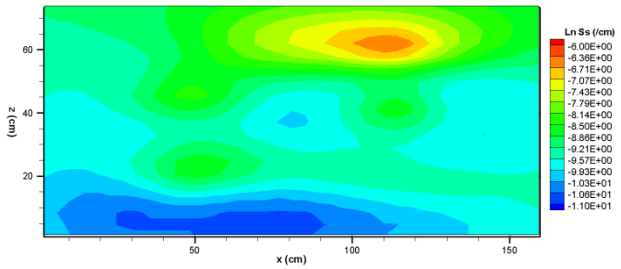


Figure A5: ln-S_s tomograms computed using: a) 2 pumping tests (ports 47 and 44); b) 4 pumping tests (ports 47, 44, 35, and 32); c) 6 pumping tests (ports 47, 44, 35, 32, 17, and 14); and d) 8 pumping tests (ports 47, 44, 35, 32, 17, 14, 5 and 2).

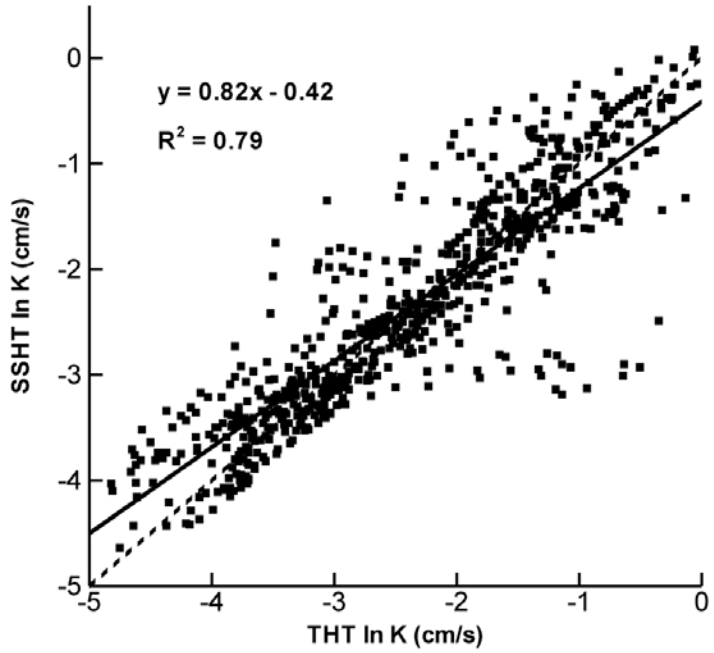


Figure A6: Scatterplot of ln-*K* from transient hydraulic tomography (this paper) to ln-*K* obtained using steady state hydraulic tomography (*Illman et al*., 2010). The dashed line is a 1:1 line indicating a perfect match. The solid line is a best fit line, and the parameters describing this line are only each plot.

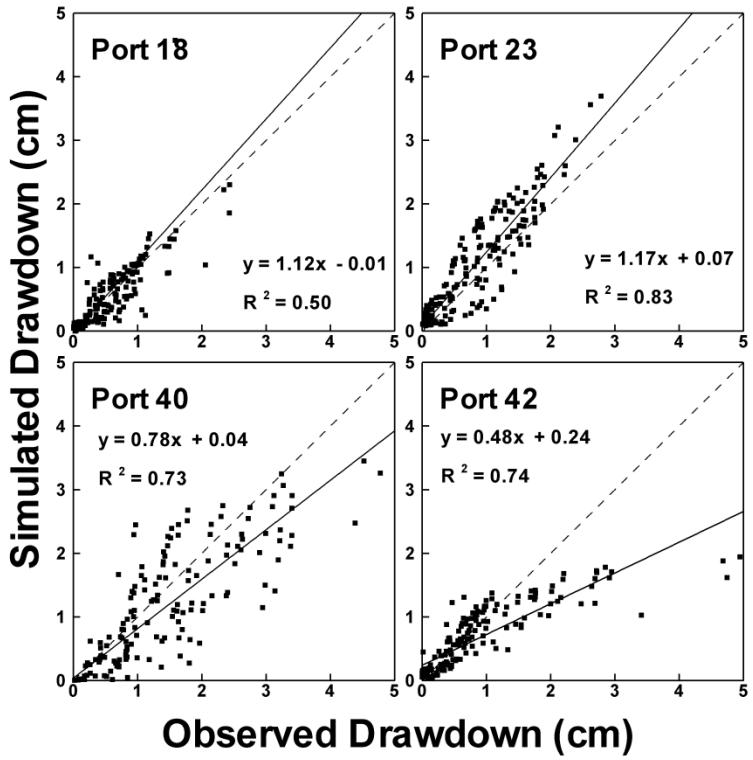


Figure A7: Scatterplots of observed vs simulated drawdowns from independent cross-hole tests 18, 23, 40, and 42 at t = 0.5, 2, 5, and 10 secs. Simulated drawdown values were obtained through numerical simulations with a calibrated geological model of K and S_s. The dashed line is a 1:1 line indicating a perfect match. The solid line is a best fit line, and the parameters describing this line are on each plot.

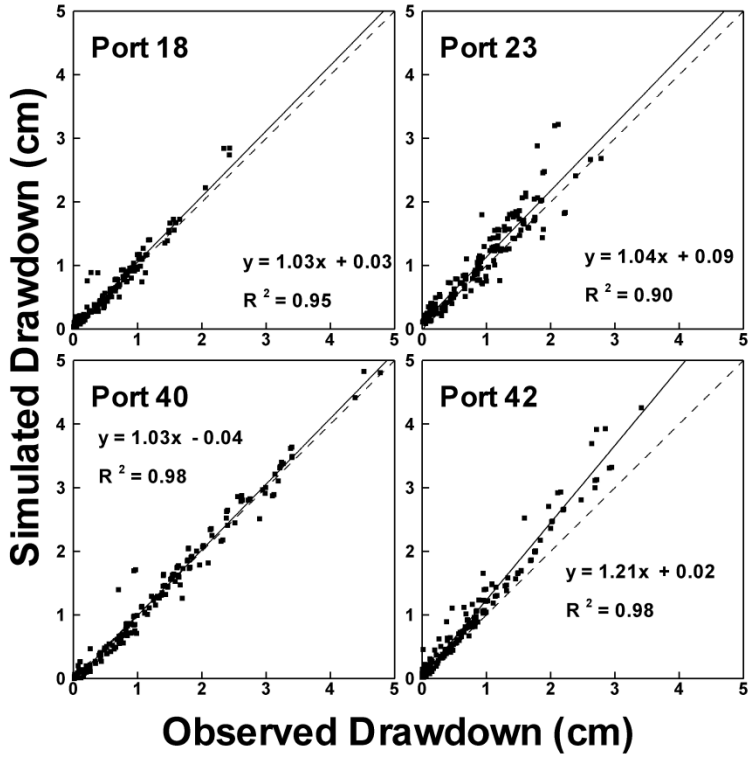


Figure A8: Scatterplots of observed vs simulated drawdowns from independent cross-hole tests 8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42 at t = 0.5, 2, 5, and 10 secs. Simulated drawdown values were obtained through numerical simulations with K and S_s tomograms from transient hydraulic tomography. The dashed line is a 1:1 line indicating a perfect match. The solid line is a best fit line, and the parameters describing this line are on each plot.

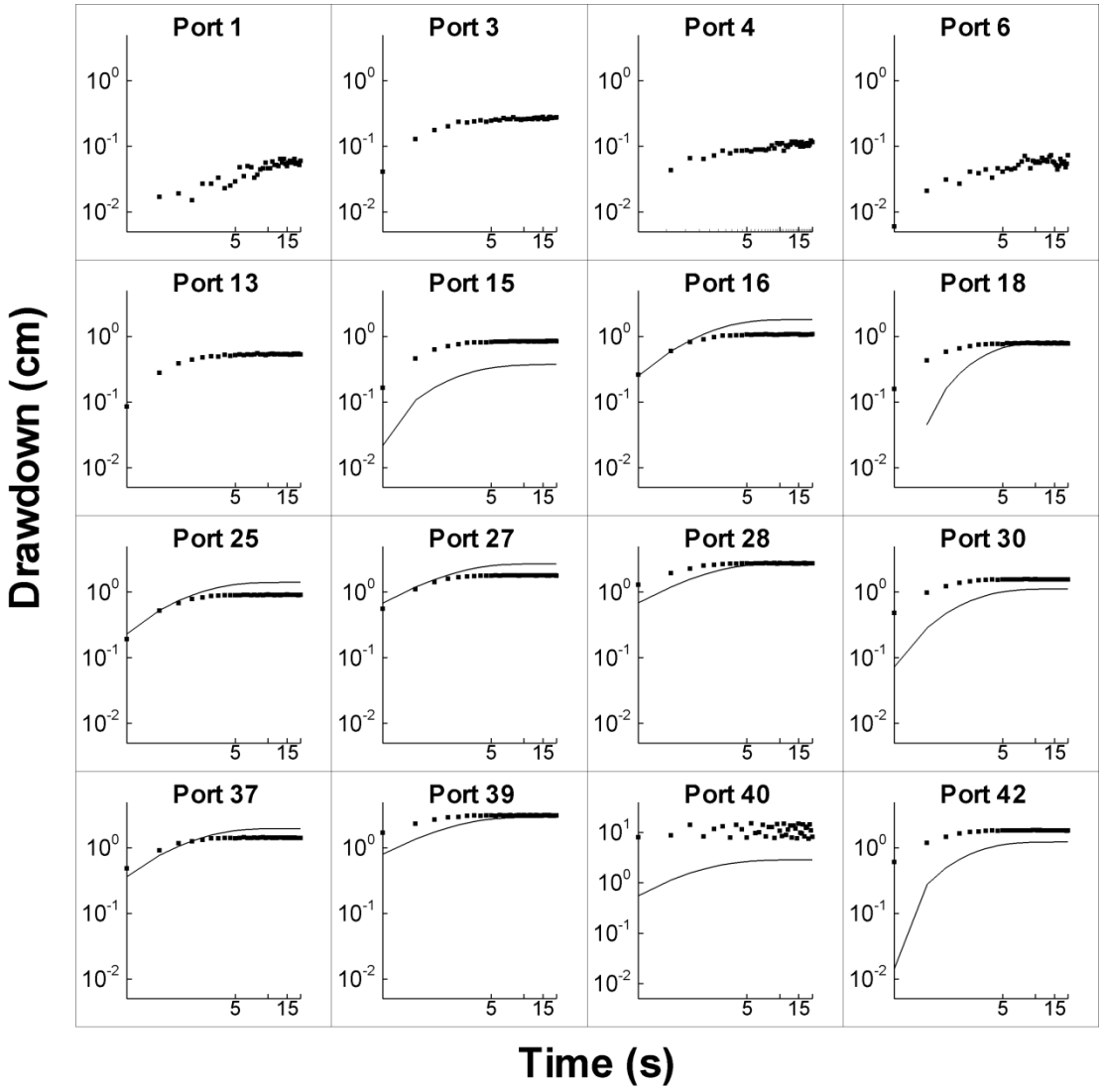


Figure A9: Observed (small dots) and simulated (curves) records of drawdown versus time (seconds) during cross-hole pumping test with pumping at port 40. Simulated values are obtained from the forward simulation of the pumping test using the calibrated geological model.

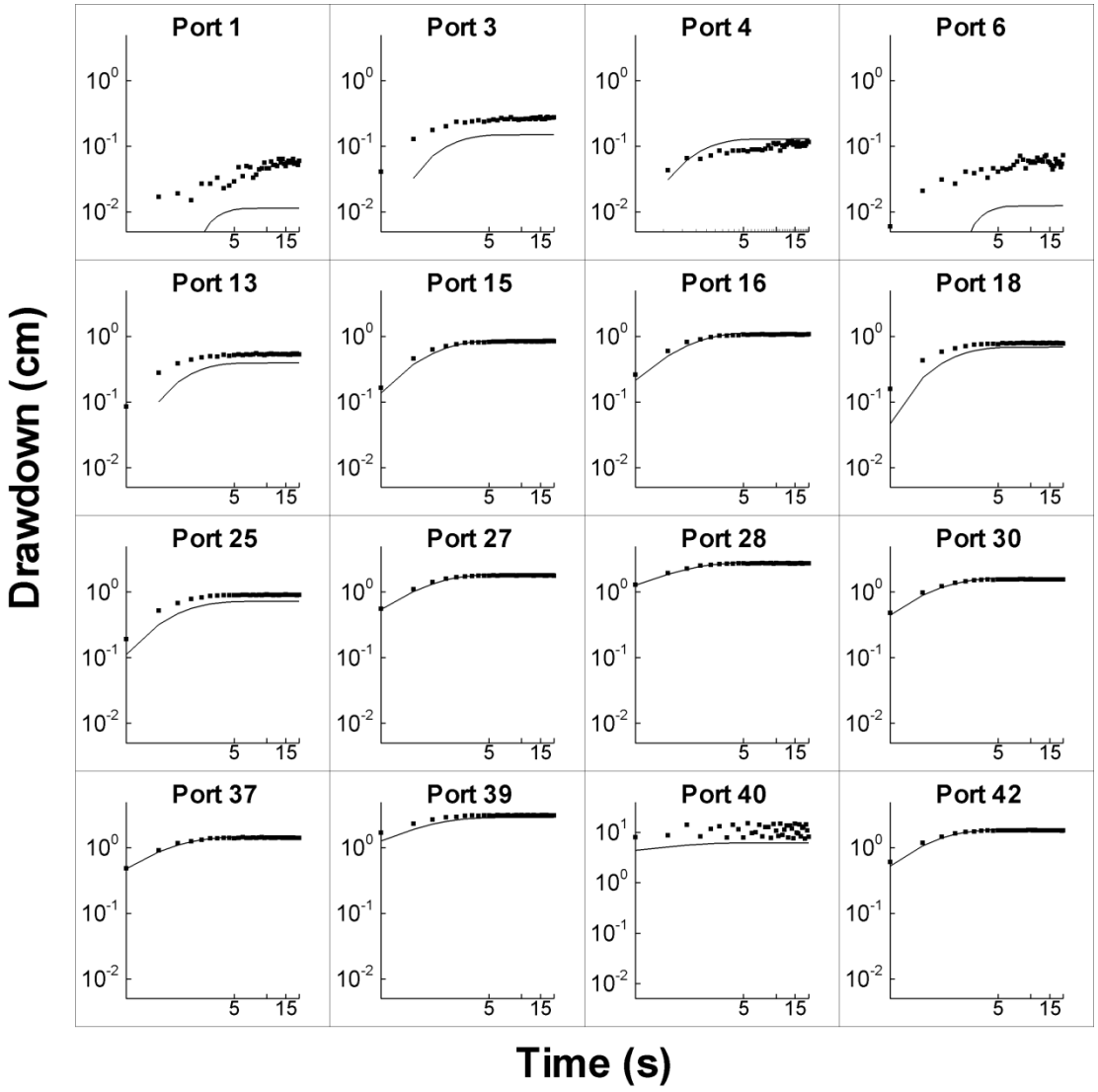
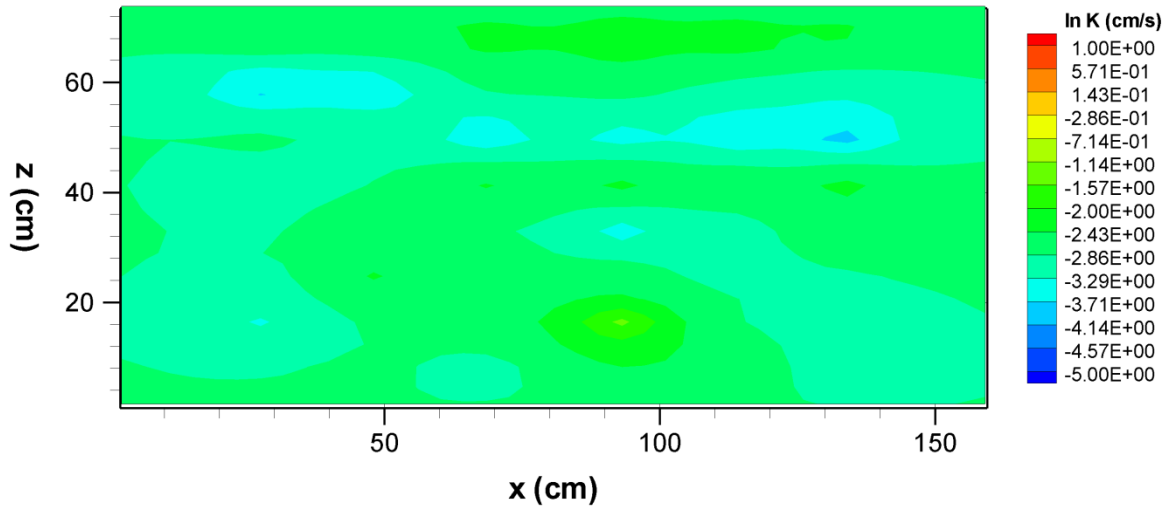


Figure A10: Observed (small dots) and simulated (curves) records of drawdown versus time (seconds) during cross-hole pumping test with pumping at port 40. Simulated values are obtained from the forward simulation of the pumping test using the K and S_s tomograms from transient hydraulic tomography.

a)



b)

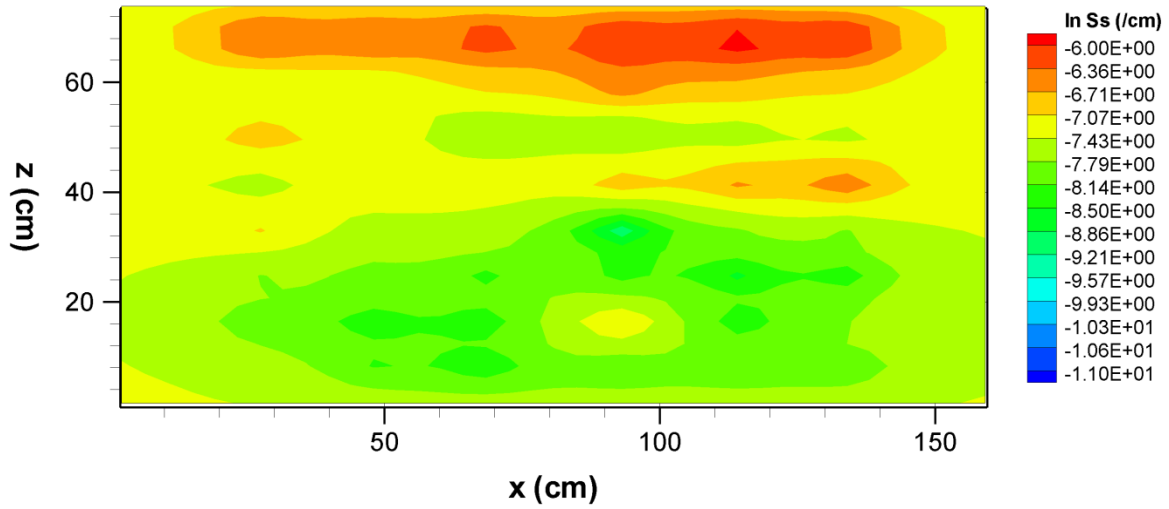


Figure SA1: Kriged K a) and S_s b) fields generated from 48 a single-hole pumping tests.

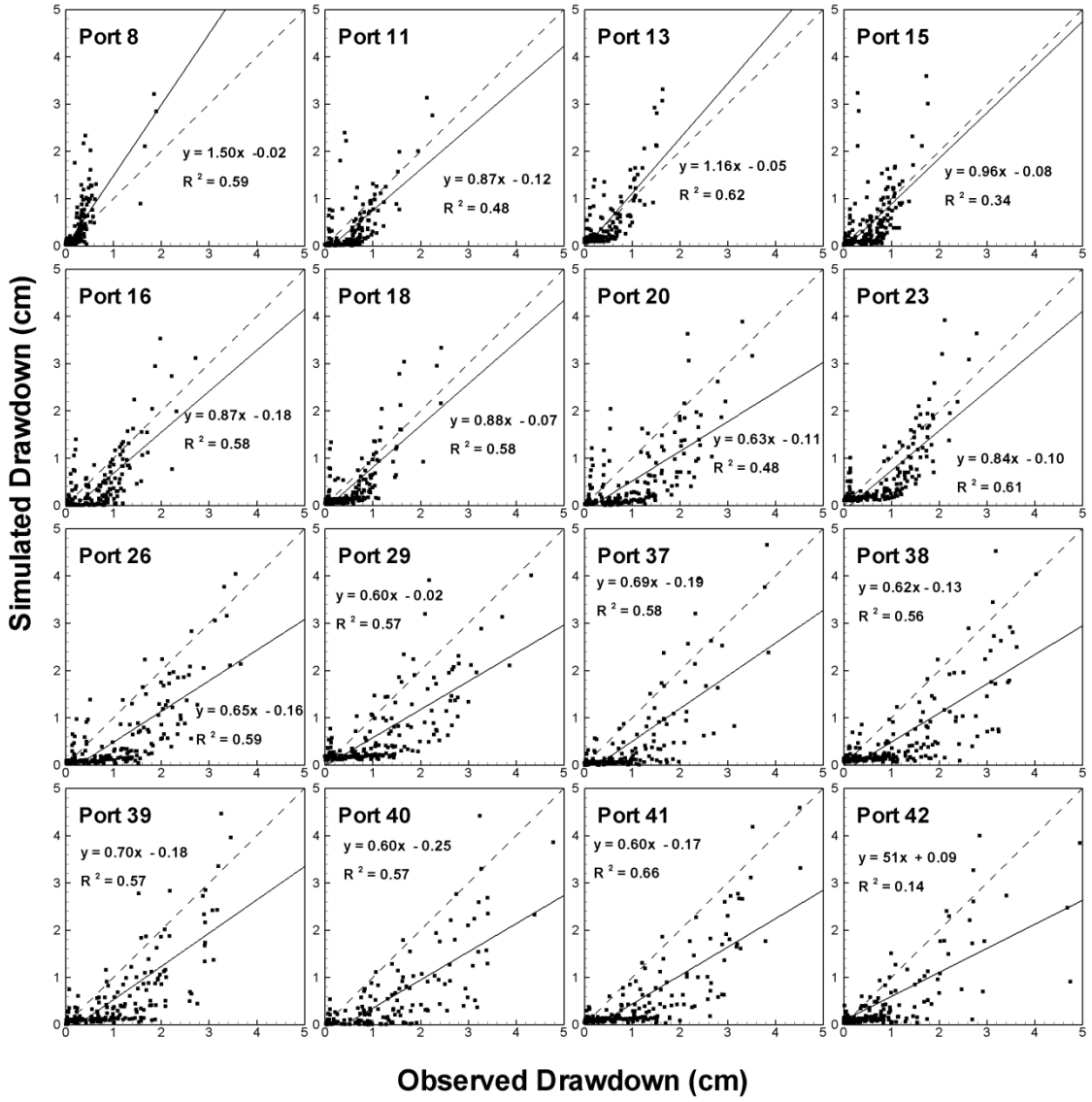


Figure SA2: Scatterplots of observed vs simulated drawdowns from independent cross-hole tests 8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42 at t = 0.5, 2, 5, and 10 seconds. Simulated drawdown values were obtained through numerical simulations with geometric mean of K and S_s estimates from single-hole tests. The dashed line is a 1:1 line indicating a perfect match. The solid line is a best fit line, and the parameters describing this line are on each plot.

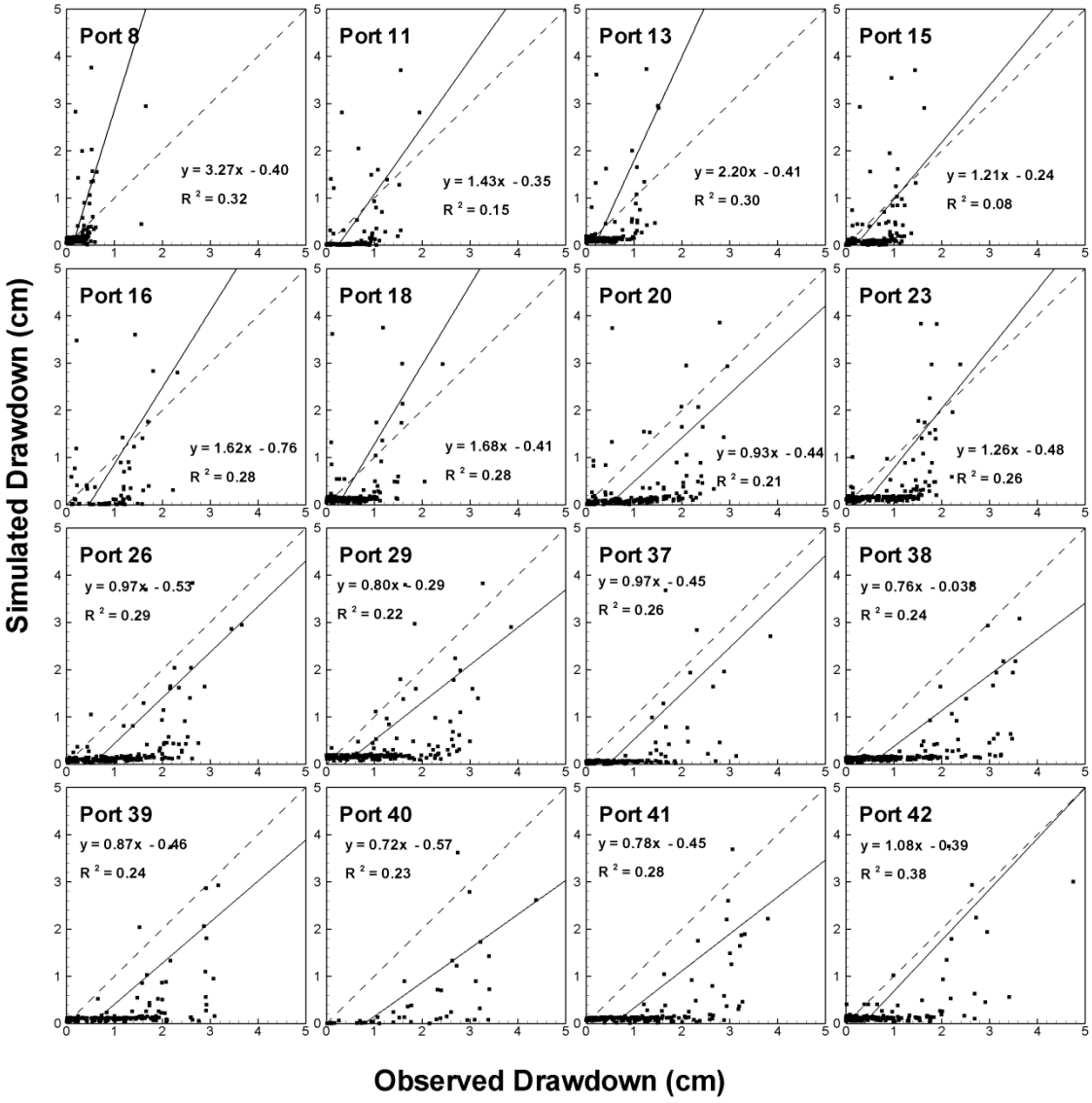


Figure SA3: Scatterplots of observed vs simulated drawdowns from independent cross-hole tests 8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42 at t = 0.5, 2, 5, and 10 seconds. Simulated drawdown values were obtained through numerical simulations the PEST estimated effective K and S_s values from the cross-hole test at port 21. The dashed line is a 1:1 line indicating a perfect match. The solid line is a best fit line, and the parameters describing this line are on each plot.

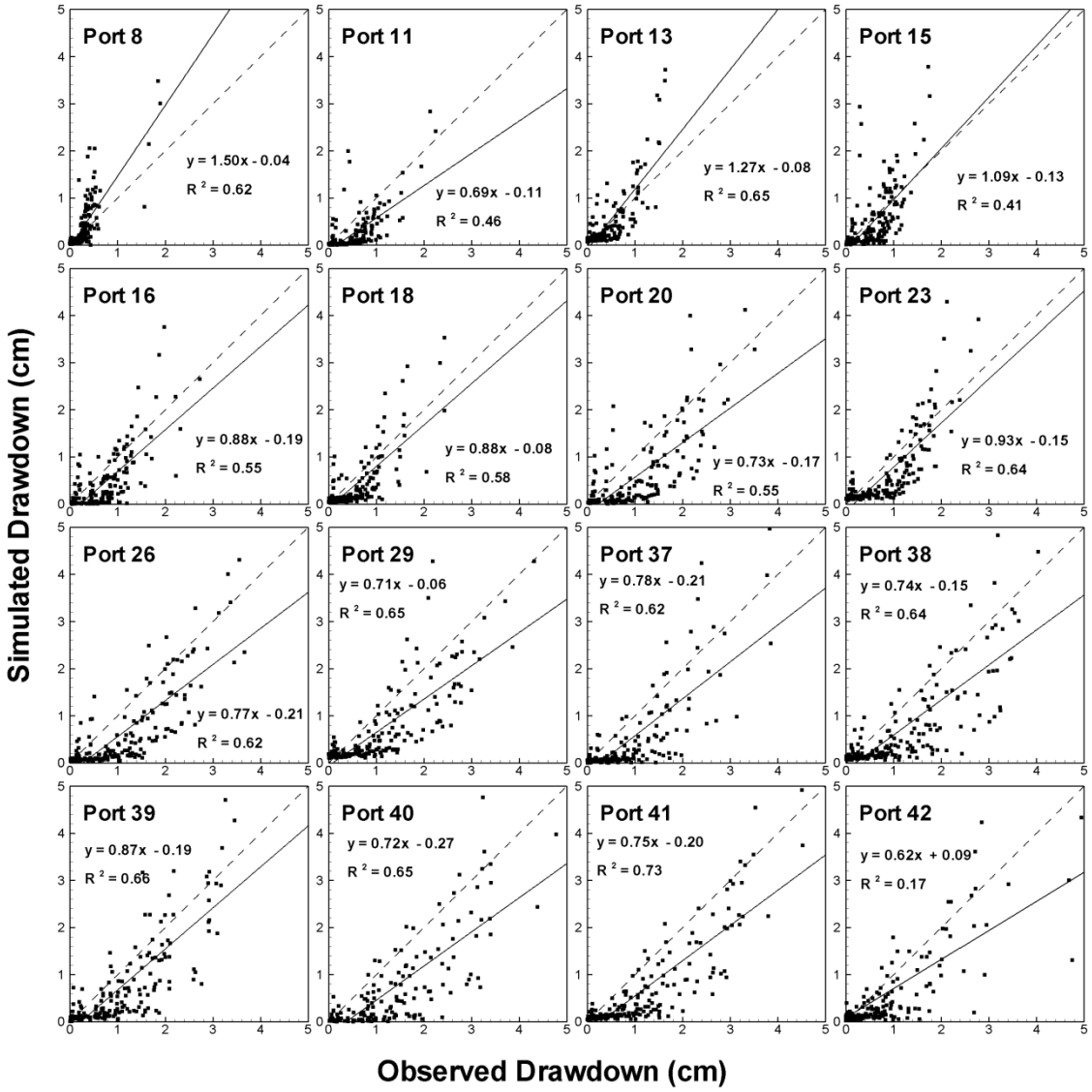


Figure SA4: Scatterplots of observed vs simulated drawdowns from independent cross-hole tests 8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42 at t = 0.5, 2, 5, and 10 seconds. Simulated drawdown values were obtained through numerical simulations with kriged K and S_s fields from single-hole tests. The dashed line is a 1:1 line indicating a perfect match. The solid line is a best fit line, and the parameters describing this line are on each plot.

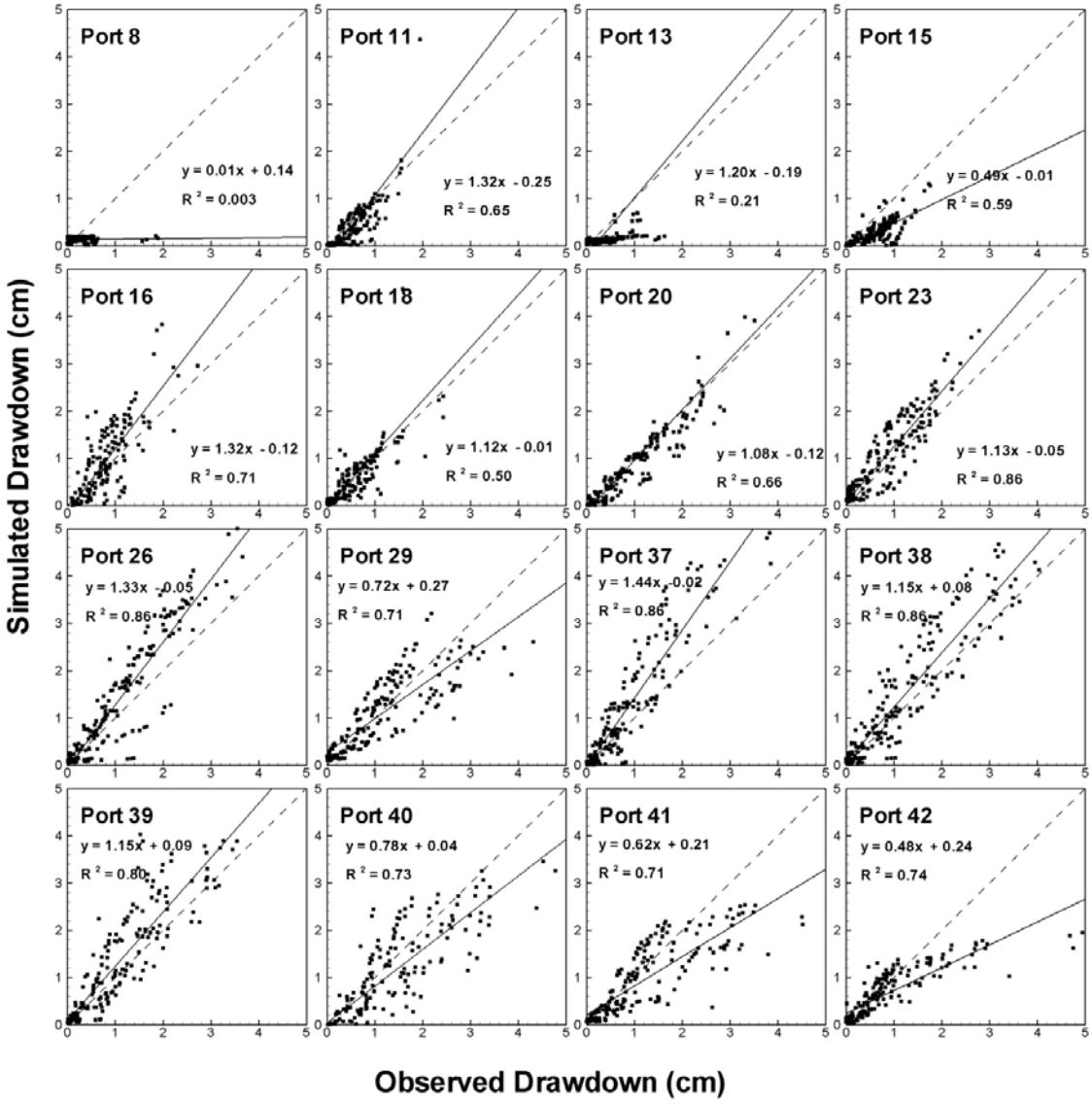


Figure SA5: Scatterplots of observed vs simulated drawdowns from independent cross-hole tests 8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42 at t = 0.5, 2, 5, and 10 seconds. Simulated drawdown values were obtained through numerical simulations with an calibrated geological model of K and S_s. . The dashed line is a 1:1 line indicating a perfect match. The solid line is a best fit line, and the parameters describing this line are on each plot.

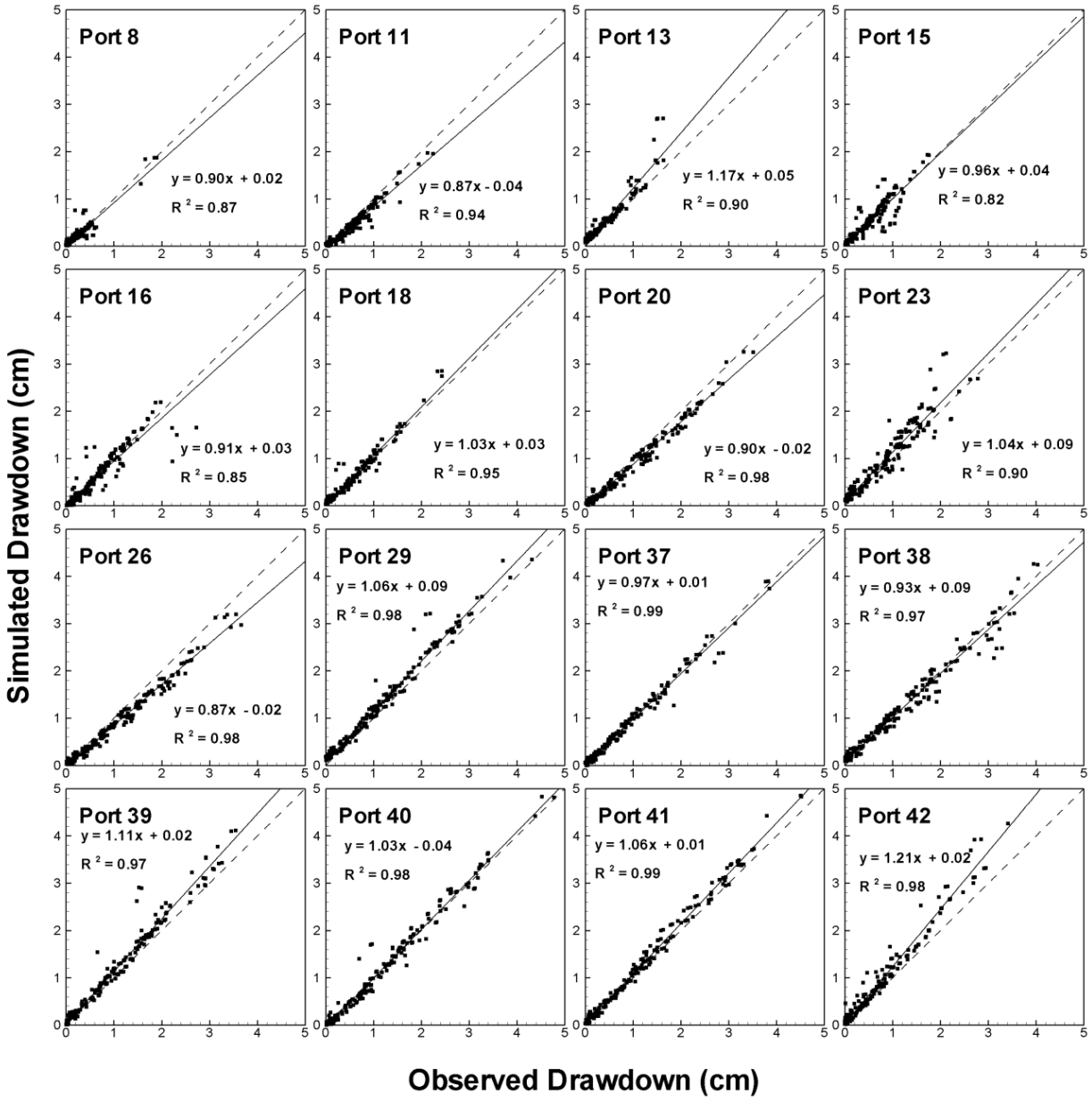


Figure SA6: Scatterplots of observed vs simulated drawdowns from independent cross-hole tests 8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42 at t = 0.5, 2, 5, and 10 seconds. Simulated drawdown values were obtained through numerical simulations with K and S_s tomograms from transient hydraulic tomography . The dashed line is a 1:1 line indicating a perfect match. The solid line is a best fit line, and the parameters describing this line are on each plot.

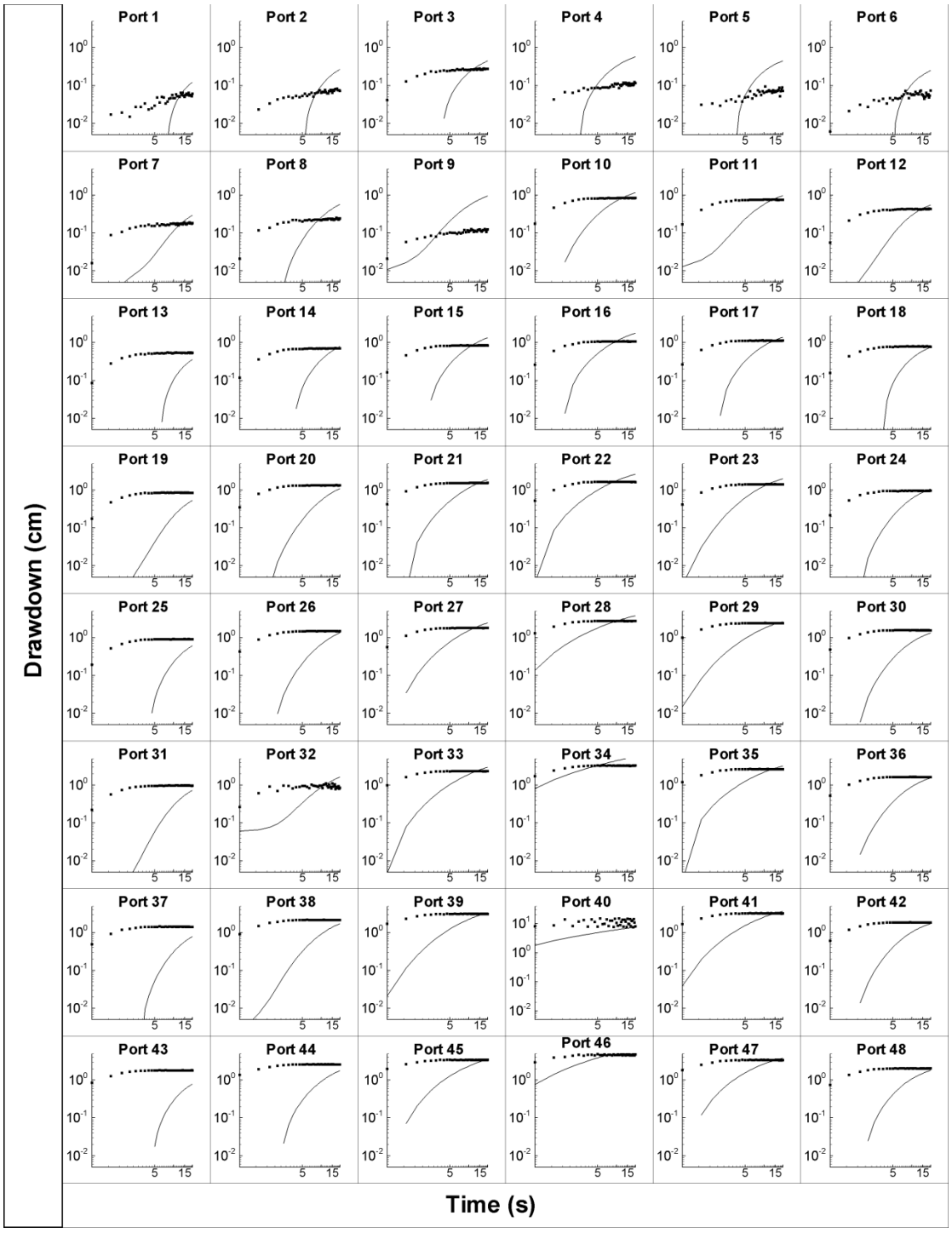


Figure SA7: Observed (small dots) and simulated (curves) records of drawdown versus time (seconds) during cross-hole pumping test with pumping at port 40. Simulated values are obtained from the forward simulation of the pumping test using the geometric mean of K and S_s estimates from single-hole pumping tests.

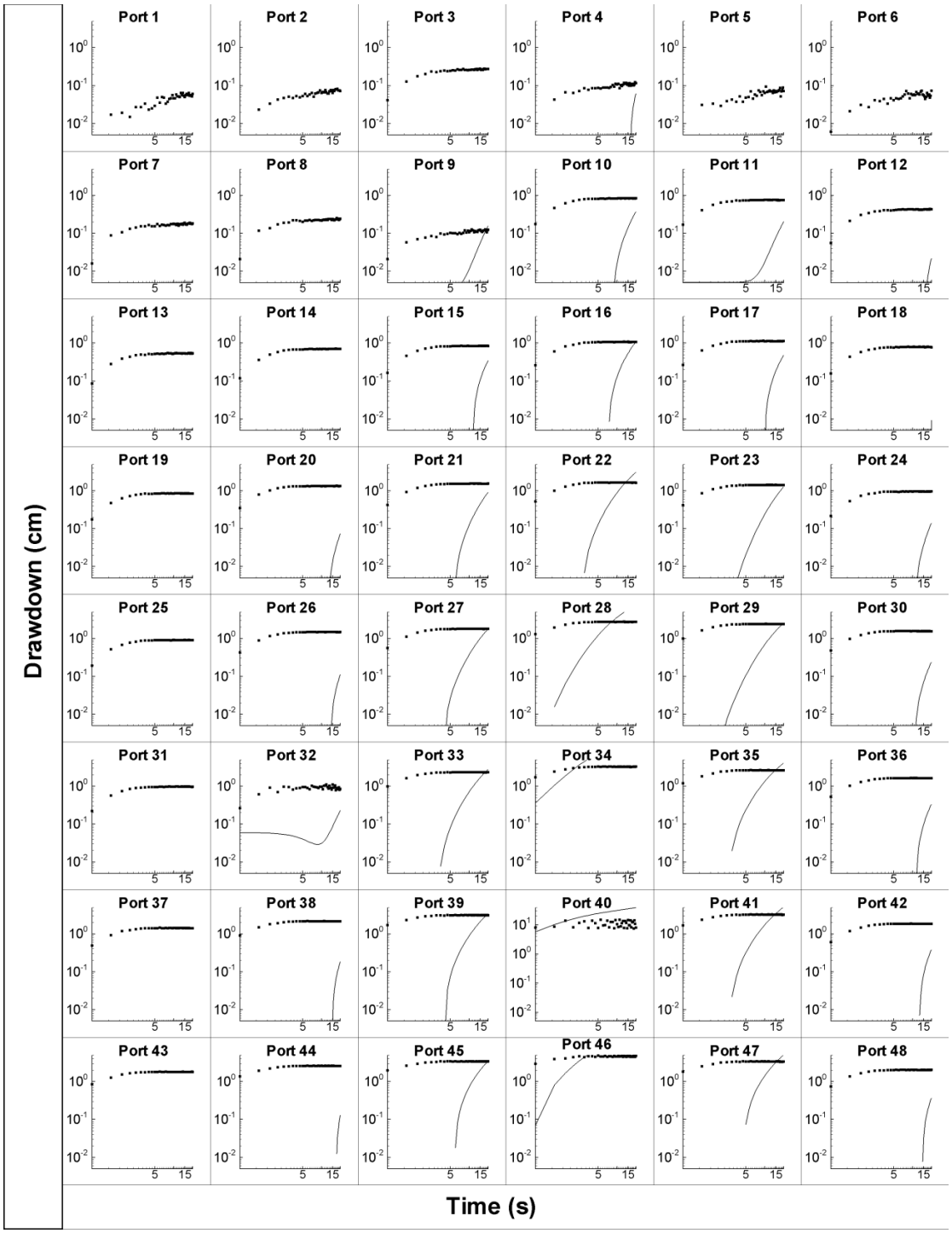


Figure SA8: Observed (small dots) and simulated (curves) records of drawdown versus time (seconds) during cross-hole pumping test with pumping at port 40. Simulated values are obtained from the forward simulation of the pumping test using the PEST estimated effective K and S_s values from the cross-hole test at port 20.

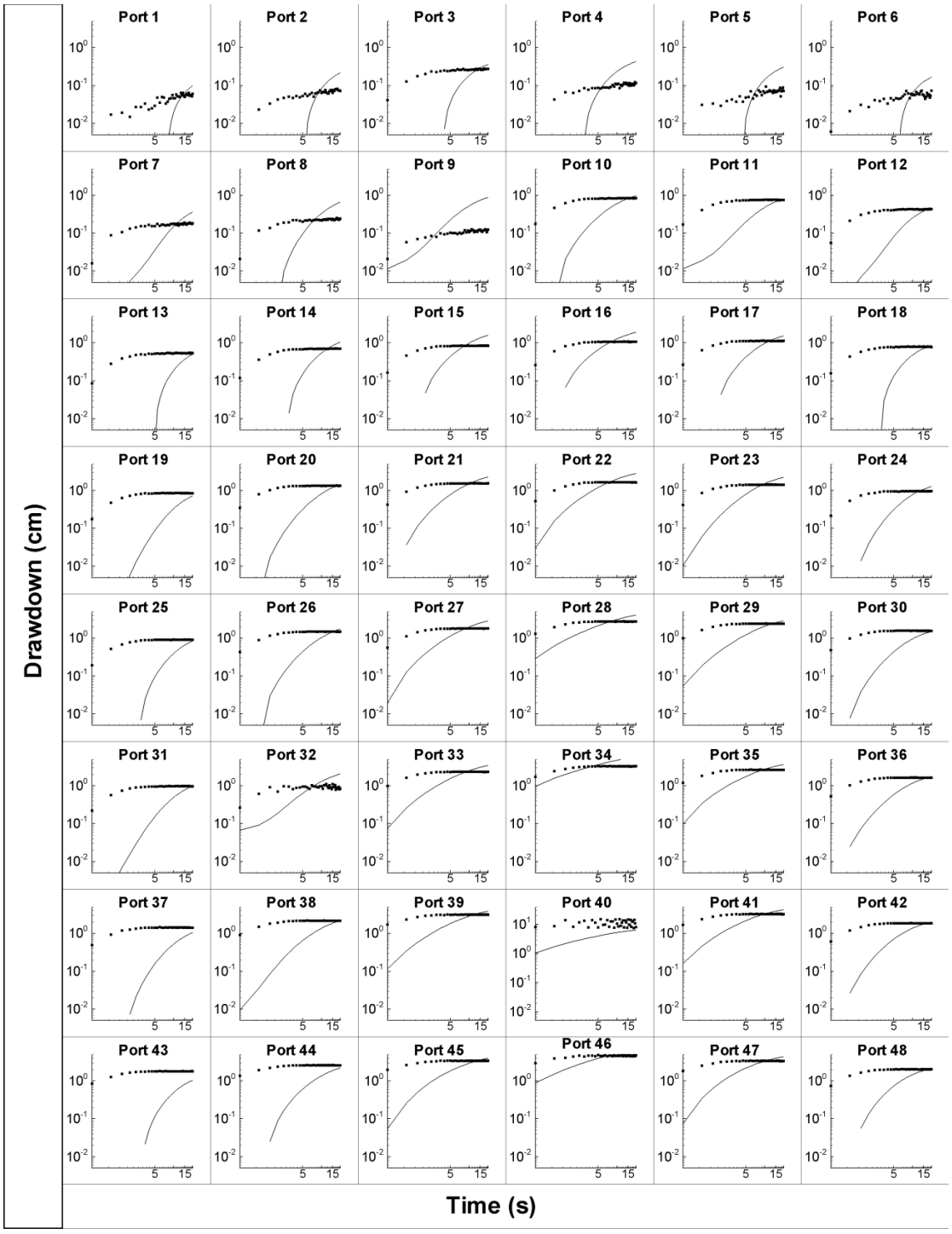


Figure SA9: Observed (small dots) and simulated (curves) records of drawdown versus time (seconds) during cross-hole pumping test with pumping at port 40. Simulated values are obtained from the forward simulation of the pumping test using the kriged K and S_s distribution.

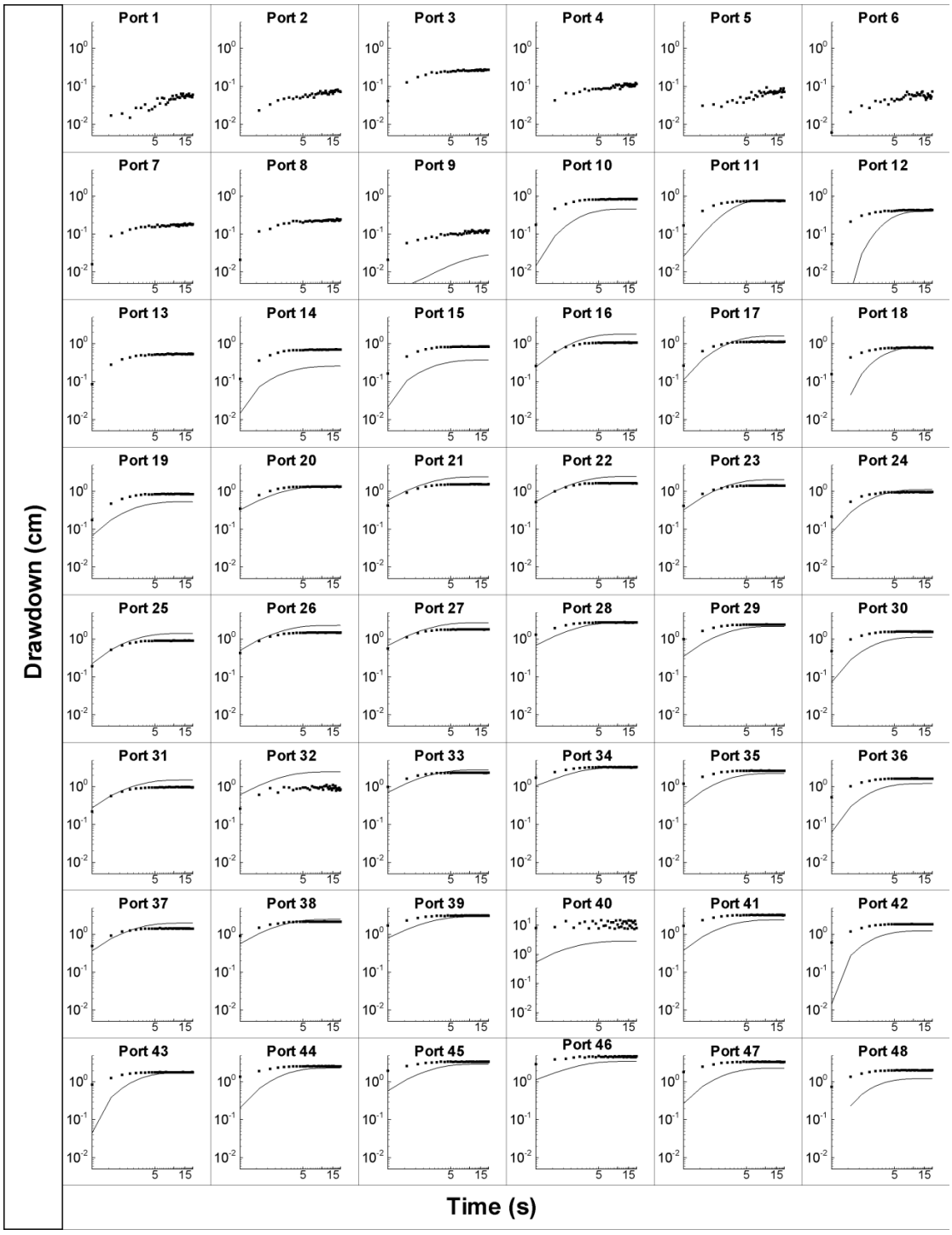


Figure SA10: Observed (small dots) and simulated (curves) records of drawdown versus time (seconds) during cross-hole pumping test with pumping at port 40. Simulated values are obtained from the forward simulation of the pumping test using the calibrated geological model.

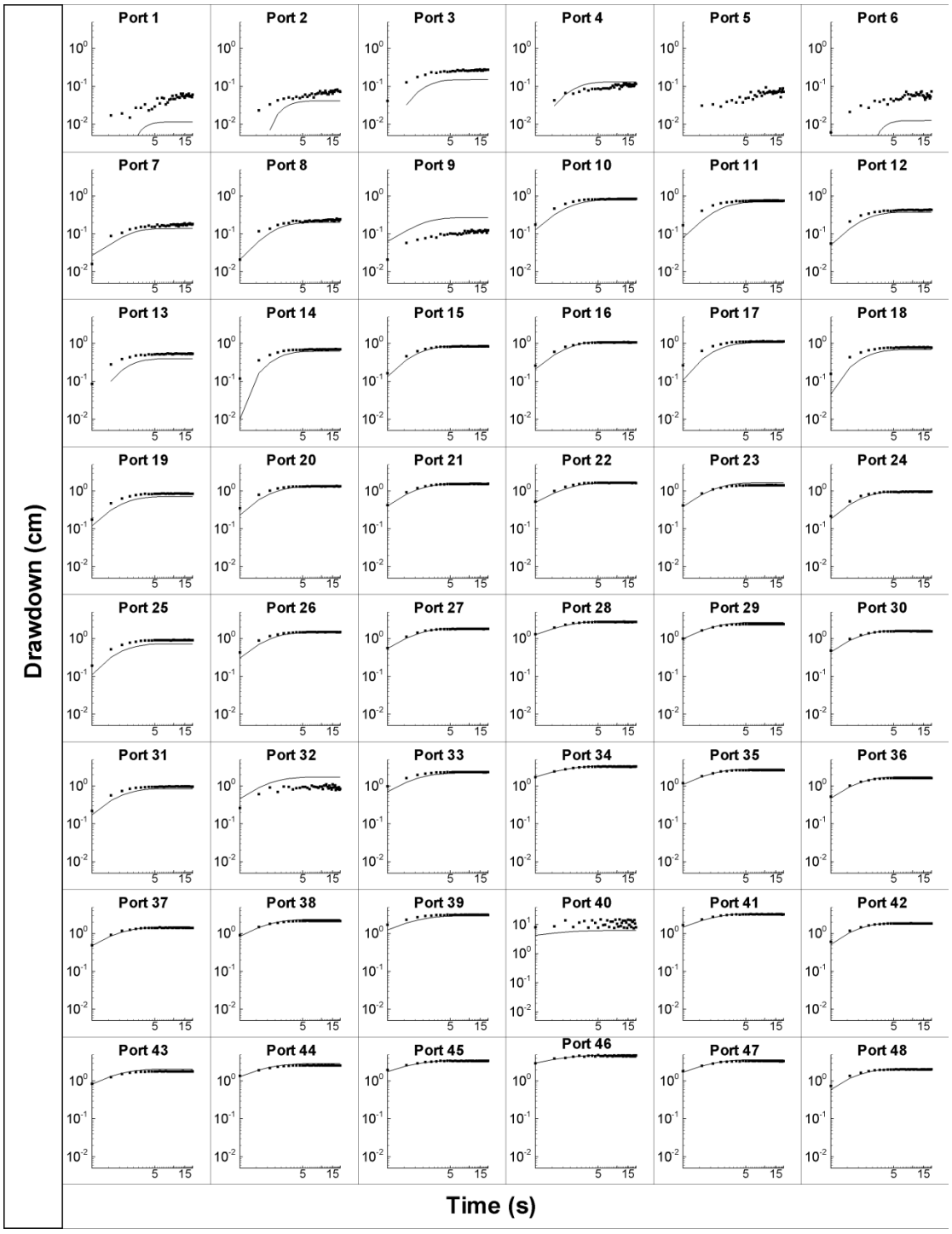


Figure SA11: Observed (small dots) and simulated (curves) records of drawdown versus time (seconds) during cross-hole pumping test with pumping at port 40. Simulated values are obtained from the forward simulation of the pumping test using the K and S_s tomograms from THT.

Appendix B

Comparison of Approaches For Predicting Solute Transport: Sandbox Experiments

This manuscript by *Illman W.A., S.J Berg,, and T-C. J. Yeh* is currently under review.

EXECUTIVE SUMMARY

The main purpose of this paper was to compare various heterogeneity characterization methods in predicting solute transport. The three methods compared were: 1) the effective parameter approach (Gelhar and Axness, 1983); 2) ordinary kriging; and 3) hydraulic tomography. We conducted our comparison in a heterogeneous sandbox aquifer. The aquifer was first characterized by taking 48 core samples to obtain local-scale hydraulic conductivity (K). The spatial statistics of these K values were then used to calculate the effective parameters. These K values and their spatial statistics were also used for kriging to obtain a heterogeneous K field. In parallel, we performed a hydraulic tomography survey using cross-hole hydraulic tests conducted in a dipole fashion with the drawdown data analyzed using the SSLE code of Yeh and Liu (2000) to yield the K distribution (or K tomogram). The effective parameters, heterogeneous K field from kriging, and the K tomogram were then used in forward simulations of a conservative tracer test conducted in a dipole configuration. The breakthrough curves and the temporal moments from the simulations were compared to those obtained from the actual experiment conducted in the sandbox aquifer. Results show an improvement in predictions of drawdown behavior and tracer transport during the dipole tracer test when the K distribution obtained from hydraulic tomography was used. This suggests that the high-resolution prediction of solute transport is possible without collecting a large number of small scale samples to estimate flow and transport properties that are costly to obtain at the field scale.

Introduction

Improving our ability to predict the temporal and spatial evolution of solute transport in geologic media has been a topic of great interest over the last several decades. This is because it can improve our ability to safeguard groundwater supplies, and increase the efficiency of groundwater pollution cleanup efforts. The recognition of the importance of the spatial variability of hydraulic conductivity (K) for understanding and predicting solute transport in geologic media, as well as the quantification of associated uncertainty has led to the development of various stochastic effective parameter/macrodispersivity approaches (see Gelhar, 1993). These approaches upscale the spatial statistics of small-scale K to derive the effective K for an equivalent homogeneous media; then, macrodispersivity is applied to account for the effects of spatial variability of small-scale K on solute transport that were omitted in the estimation of effective K.

The stochastic effective parameter approaches have yielded much optimism about our abilities to predict solute transport behavior both spatially and temporally over the past few decades. For example, various field (e.g., MacKay et al., 1986; LeBlanc et al., 1991; Boggs et al., 1992) and laboratory experiments (e.g., Fernàndez-Garcia et al., 2005 and others) have been conducted to show the applicability of the effective K and macrodispersivity approach. In particular, Sudicky (1986) characterized the spatial variability of K in great detail in a small area of a tracer test experiment site at the Borden site in Canada, by conducting permeability measurements on a series of cores. Sudicky (1986) and Freyberg (1986) then used the theoretical expressions for effective K and macrodisperivity developed by Gelhar and Axness (1983) and Dagan (1982, 1984) to predict the evolution of the spatial moments (mean position, and spread) of the observed tracer plume. Results of the prediction are very satisfactory. More recently, at a landfill site in North Bay, Canada, Sudicky et al. (2010) also demonstrated that the use of the effective parameters based on stochastic theories can yield a reasonably good prediction of the large-scale behavior of a tracer plume in geological media.

Research at other sites, most notably the MADE site, however, suggests that these stochastic theories may not be able to predict even the average plume behavior and the failure has been attributed to

the non-Fickian behavior of the plume (Rehfeldt et al., 1992). This has led to a race for more complex mathematical approaches such as space-time nonlocal stochastic advective-dispersive flux theory (e.g., Neuman, 1993; Cushman and Ginn, 1993; Zhang and Neuman, 1996), continuous time random walk (CTRW), (e.g., Scher and Lax, 1973; Berkowitz et al., 2006), and advection dispersion equations based on fractional derivatives to describe solute transport in strongly heterogeneous porous and fractured geological media (e.g., Meerschaert et al., 1999; Benson et al., 2000). Neuman and Tartakovsky (2009) recently offered their perspective on differences, commonalities, and relative merits of these approaches. These new theories nevertheless still assume homogeneity of geologic media and as such are limited to the description of integrative behaviors of a plume (the mean spread, or averaged breakthrough over a compliance surface or the breakthrough at a pumping well) rather than the detailed temporal and spatial distribution of the plume.

Limitations of these effective hydraulic parameter/macrodispersivity approaches for predicting fine-scale behavior of the tracer plume were demonstrated in the experiments at the Georgetown site in South Carolina, USA, where Yeh et al. (1995) and McCarthy et al. (1996) conducted two-well, forced-gradient tracer experiments over an area of 5 m x 5m in a coastal sandy aquifer of a thickness of 3 m. The evolution of three-dimensional chloride plumes during two tracer experiments was observed in a dense sampling network. While the sandy aquifer appears to be homogeneous, a large portion of the bromide tracers injected uniformly over the entire depth of the aquifer was found to migrate preferentially to the pumping well through the bottom portion of the aquifer which is composed of coarser-textured sand. Using a three-dimensional flow and transport model, Yeh et al. (1995) showed that although assuming homogeneity, using the effective K and macrodispersivity can satisfactorily produce the chloride breakthrough observed at the pumping well (an integrative behavior), but the preferential migration of tracer through the bottom of the aquifer can only be captured by using the detailed three-dimensional K field measured from a large number slug tests.

Results of these field studies corroborate the view by Yeh (1992) that the upscaled effective K and macrodispersivity approaches for large-scale field problems are practical and useful. However, the

approaches can yield only the integrative or general behavior of solute spread and migration in the subsurface, which may not satisfy the resolution of interest; they serve, at best, as a tool for a first-cut analysis. In order to meet the resolution of interest, more detailed information on the spatial distribution of small-scale *K* heterogeneity is needed to describe fine scale behavior.

Following the school of thought of Yeh (1992), a high-resolution aquifer characterization approach (i.e., hydraulic tomography) has been developed (e.g., Gottlieb and Dietrich, 1995; Yeh and Liu, 2000; Vesselinov et al., 2001; Bohling et al., 2002; Brauchler et al., 2003; McDermott et al., 2003; Zhu and Yeh, 2005, 2006; Li et al., 2005; Fienen et al., 2008; Castagna and Bellin, 2009; Xiang et al., 2009; Liu and Kitanidis, 2011). Laboratory experiments have been conducted in a number of studies (Liu et al. 2002; Illman et al., 2007, 2008; 2010a; Liu et al., 2007; Yin and Illman, 2009) to evaluate hydraulic tomography. These studies show not only that hydraulic tomography can map the heterogeneity in detail, but it also yields hydraulic parameter fields that can be used to predict flow under different stresses. The robustness of hydraulic tomography in terms of imaging the heterogeneity patterns in hydraulic parameters is also evident through various field experiments (Bohling et al., 2007; Li et al., 2007; Straface et al., 2007; Illman et al., 2009; Cardiff et al., 2009; Berg and Illman, 2010).

Recently, Ni et al. (2009) conducted numerical simulations to show that hydraulic tomography is able to capture a sufficient spatial distribution of *K* so that detailed, high-resolution, solute transport behaviors can be predicted. In particular, these authors used a synthetic two-dimensional aquifer to assess the accuracy of predicted concentration breakthrough curves (BTCs) on the basis of the *K* fields estimated by geometric mean, kriging, and hydraulic tomography. Such *K* fields represent different degrees of flow resolutions as compared with the synthetically generated one. It was found that without accurate estimates of local-scale dispersivities and porosities, the flow field based on the *K* distribution or (*K* tomogram from now on) can yield accurate predictions of breakthrough arrival and peaks. However, such a promising result has not been demonstrated to date either in the laboratory or in the field.

The main objective of this paper is to compare three different approaches for predicting solute transport. The three methods compared are: 1) the effective parameter approach of Gelhar and Axness

(1983); 2) point samples of K and ordinary kriging; and 3) the K field obtained from hydraulic tomography. We conducted our comparison by creating a heterogeneous aquifer through the cyclic flux of sediment-laden water in a laboratory sandbox. The aquifer was first characterized by taking core samples to obtain local scale estimates of K which were then used to calculate effective parameters, and to perform kriging. We then conduct a hydraulic tomography survey using the inverse code developed by Yeh and Liu (2000) to map the heterogeneous K tomogram. Upon completion of the hydraulic tomography survey, we conduct a conservative tracer test in a dipole configuration. Water samples collected at various locations within the synthetic aquifer allowed us to obtain breakthrough curves. The tracer test is then simulated through a forward transport simulation using the effective parameter field, the kriged K field, and the K tomogram with a constant porosity value. We compare the evolution of the tracer plume qualitatively by examining the photographs of the dyed plume and comparing those to the various numerical simulation cases based on the effective parameter concept, kriging, and hydraulic tomography. We then make a direct comparison of the simulated and observed breakthrough curves. Finally, we quantitatively assess the results through the comparison of temporal moments computed from the simulated and observed breakthrough curves.

Experimental Methods

Sandbox and Synthetic Aquifer Construction

A synthetic heterogeneous aquifer was constructed in a vertical, laboratory sandbox to validate various fluid flow and solute transport algorithms. The sandbox is 193.0 cm in length, 82.6 cm in height, and has a depth of 10.2 cm. The aquifer was created through the cyclic deposition of sediments under varying water flow and feed rates of sediments (Illman et al., 2010a). Previous efforts by our group (Illman et al., 2007, 2008; Liu et al., 2007; Yin and Illman, 2009; Illman et al., 2010b) involved packing the sandbox in a relatively simplistic pattern. Our goal in relying on sediment transport was to create a more realistic heterogeneity pattern with various scales of heterogeneity in an efficient manner. Table B1 summarizes the sands used to create the synthetic aquifer including d_{50}, which is the particle diameter for

which 50% of the grains are finer, hydraulic conductivity (*K)* estimates obtained using Shepherd's empirical model (Shepherd, 1989), and hydraulic conductivity (*K)* estimates obtained using a constant head permeameter device (Klute and Dirksen, 1986). Figure B1 is a photograph of the frontal view of the synthetic aquifer, showing the interfingering nature of the deposits with numbers indicating the layers. Port locations are also shown on this figure. Further details to this synthetic heterogeneous aquifer and its construction approach are provided in Illman et al. (2010a).

Forty eight ports, 1.3 cm in diameter, have been cut out of the stainless steel wall to allow coring of the aquifer as well as installation of horizontal wells. Each well was constructed by making 6 cuts spaced 1.46 cm apart in sections of brass tubing. The cuts were then covered with a stainless steel mesh that was bonded to the tubing with corrosion resistant epoxy. Extreme care was taken to avoid the epoxy filling the mesh which could impede water flow. The wells, which penetrate the horizontal thickness of the synthetic aquifer were installed after the deposition of the layers. Each well location is monitored by a pressure transducer, and can be used for pumping, injection of tracers, or for sampling (Illman et al. 2010a).

For this particular study, all boundaries around the synthetic aquifer were set as no-flow boundaries to achieve better mass control for our dipole cross-hole and tracer experiments. This was found to be critical during our initial studies as our previous sandboxes had constant head reservoirs which could potentially diminish the tracer signals via dilution.

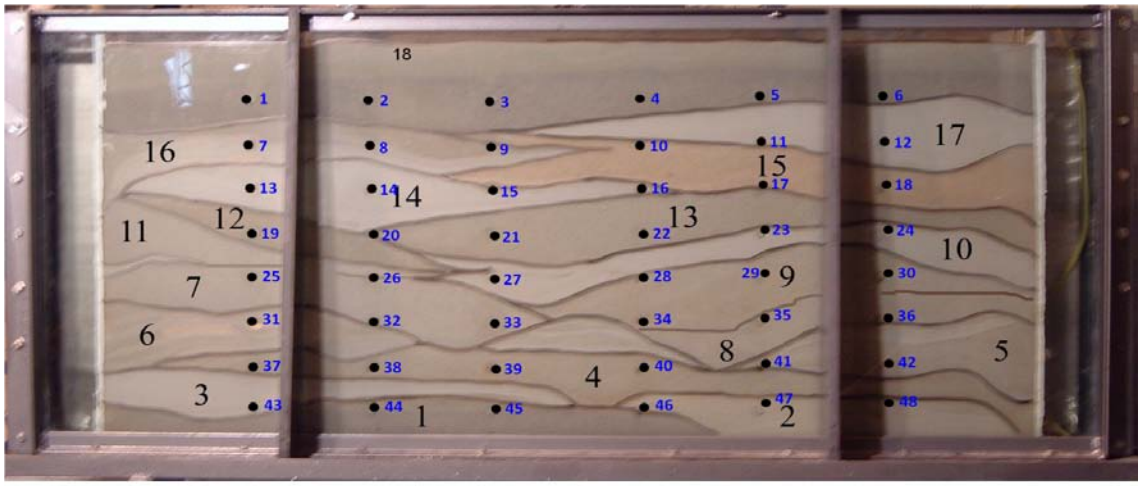


Figure B1: Photograph of the sandbox showing the synthetic heterogeneous aquifer in which the dipole cross-hole and conservative tracer tests were conducted. Large black numbers indicate layer numbers, solid circles indicate port locations, and small blue numbers indicate port numbers. Layer 1 = 20/30; layer 2 = 4030; layer 3 = F85; layer 4 = 20/40; layer 5 = mix; layer 6 = mix; layer 7 = #12; layer 8 = F32; layer 9 = 20/40; layer 10 = F65; layer 11 = #12; layer 12 = 16/30; layer 13 = 20/30; layer 14 = F75; layer 15 = 20/40; layer 16 = mix; layer 17 = F85; layer 18 = 20/30. Note: The layers labelled "mix" consisted of equal volumes of #14, F75, and 16/30 sands (after Illman et al. 2010a).

The data acquisition system used for the laboratory experiments consisted of three major components. Pressure measurements were made with 50 Setra model 209 gauge pressure transducers with a range of 0 to 1 psi, 48 of which measured hydraulic head in the aquifer. These pressure transducers were installed at each of the 48 ports in the stainless steel wall of the sandbox.

Table B1: Characteristics of porous media used to create a synthetic heterogeneous aquifer (after Illman et al. 2010a).

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

Dipole tracer test

Description of dipole tracer test

We conducted a dipole tracer test using bromide as a conservative tracer. Prior to the injection of the tracer, a dipole flow field was established by injecting tap water at port 42 at a rate of 372.4 mL/min and extracting at port 7 at 304 mL/min. A water mass balance was then achieved by setting 12 sampling lines at a cumulative rate of 68.4 mL/min. Figure B2 is a schematic diagram showing the injection, extraction and sampling ports for the dipole conservative tracer test.

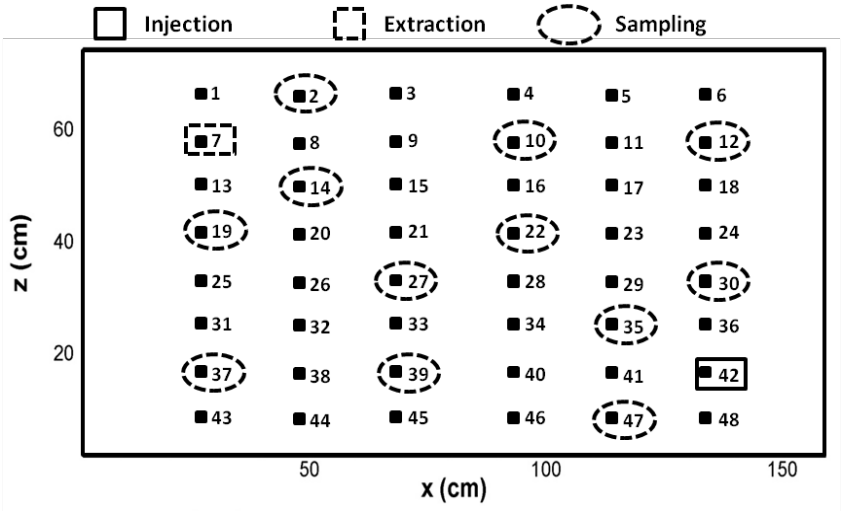


Figure B2: Schematic diagram showing the injection, extraction, and sampling ports during the dipole conservative tracer test.

Achieving a mass balance was crucial because all 6 boundaries of the sandbox were no-flow boundary conditions. Water flow rates in each of the sampling lines were monitored frequently and the pressure transducers were also used to measure the pressure in the tank during the tracer test. Deviations in the mass balance could result in an over-pressurization of the tank which could cause the glass holding the synthetic aquifer to rupture.

Once steady state flow conditions were reached (injection, extraction, and sampling rates were stable and in equilibrium) a valve in the injection line was switched from water to a solution containing

bromide (Fisher Scientific) as the conservative tracer (151.3 mg/L). Food coloring (70 ml of blue 'Tone's' Food Coloring) was added to the tracer solution so that we could observe the movement of the tracer plume in the sandbox through the glass. The tracer solution was injected for 10 minutes and then the injection line was switched back to tap water. Each of the sampling ports (and the main extraction port) were sampled to record the movement of the bromide tracer through the sand box. Crude estimates of tracer travel time were obtained using Darcy's law to estimate sampling times at each port. Water sampling from each one of the sampling lines was made possible with a manifold that fed the water samples into twelve 1.5 ml vials simultaneously. Sixty water samples were collected from each of the 12 sampling lines (plus the effluent line) during the dipole tracer test for a total of 780 samples. The dipole tracer test lasted approximately 5 hours.

Description of water sample analysis with ion chromatograph

Water samples taken during the tracer test were chilled in a refrigerator at 4^{o}C until sample analysis. The conservative tracer (Br^-) concentrations were determined using an ion chromatograph (IC) (Dionex ICS 2000) equipped with an AS40 auto sampler. The IC analysis method consisted of an AS-18 4 mm column, EGC II KOH eluent generator with an eluent concentration of 30mM, SRS-Ultra II 4mm suppressor with a current of 75 mA, DS6 heated conductivity cell with a cell temperature of 35°C, and a flow rate of 1 mL/min. Five external calibration standards were prepared for bromide with 1000mg/L Fluka IC standard diluted with de-ionized water at levels of 1, 5, 10, 50, 100 mg/L. The calibration plots were created by injecting triplicates of each calibration level and were linear ($R^2 > 0.99$). Calibrations were made in each analysis sequence consisting of 15 calibration injections (triplicates of the 5 standards) followed by 60 tracer samples.

Aquifer Characterization Methods

Permeameter analysis of core samples

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

Dipole hydraulic tomography tests

Hydraulic tomography tests were conducted by pumping water from one well and injecting the pumped water at another location (dipole cross-hole tests from now on) and then monitoring other sampling ports in the sandbox. In the dipole tests, a mass balance of water injected and extracted was maintained using a peristaltic pump by connecting the injection and extraction ports in a single loop. Eight pairs of ports consisting of tests 1 through 8 (e.g., test 1: extraction at port 2 and injection at port 47, from now on E2/I47 (E=Extraction port and I for injection port); test 2: E42/I7; test 3: E4/I45; test 4: E15/I34; test 5: E24/I25; test 6: E17/I32; test 7: E23/I26; and test 8: E20/I29) were chosen for these dipole cross-hole tests (see Figure B1 for port locations). Injection and pumping rates for all dipole tests averaged 480 mL/min. Prior to each dipole test, all pressure transducers were calibrated to ensure accurate data collection. We then collected hydraulic head data for several minutes in all pressure transducers to establish a static, initial condition. After establishment of static conditions, we pumped from each port using a peristaltic pump, injecting the pumped water into another port, while taking head measurements at all 48 ports. For each test, pumping continued until the development of steady state conditions, which was determined by observing the stabilization of all head measurements. After reaching steady state, the pump was shut off to collect recovery head data until its full recovery.

Heterogeneity characterization methods

Geostatistical analysis and kriging of core K estimates

Geostatistical analysis of the 48 core *K* data was conducted using the Surfer 8 software developed by Golden Software, Inc (www.goldensoftware.com). The exponential variogram model was used to fit the experimental variograms in both horizontal and vertical directions, resulting in an anisotropic variogram model. However, results of the analysis of range (i.e., correlation scale) are not satisfactory since the sample interval is larger than the average thickness of the sand bodies and the number of samples in vertical and horizontal directions are not sufficient. Therefore, the correlation scales were approximated based on the average thickness and length of the discontinuous sand bodies. The variogram parameters for the experimental variograms to be used in kriging thus include the geometric mean (K_G = 0.08 cms^{-1}), the variance ($\sigma^2_{\ln K}$ = 0.87) and the correlation lengths (λ_x = 60.00 cm and λ_z = 20.00 cm).

We then kriged the core *K* estimates using this anisotropic variogram model. The kriged domain was discretized into 73 vertical elements and 162 horizontal elements. Figure B3 shows the kriged *K* field which delineates some of the major layers but definitely has a smoothed appearance. Effects of correlation lengths on the estimate were investigated and were found to not be significant.

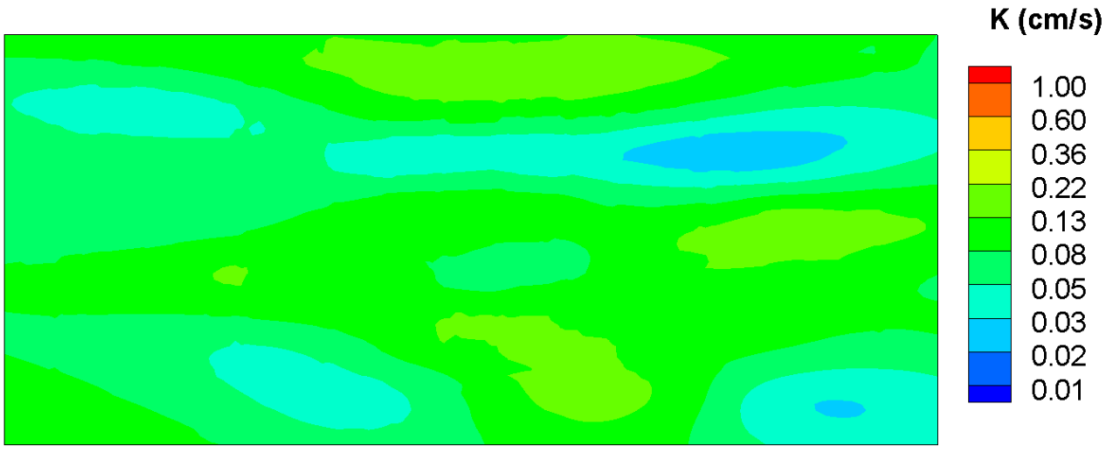


Figure B3: *K* distribution obtained through kriging of core scale *K* data from permeameter analysis.

Estimates of effective hydraulic conductivity and macrodispersivity

The stochastic theories of effective *K* and the macrodispersion parameter by Gelhar and Axness (1983), Dagan, (1982, 1984), and Neuman et al. (1987) provide a methodology for determining large-scale effective flow and transport parameters from knowledge of the spatial statistics (mean, variance, and correlation scales) describing the spatial variations of the underlying local-scale *Ln(K)* process. These theories assume that if the local-scale mixing process is sufficient, the flow and transport processes will reach the ergodic condition, where the ensemble mean flow and transport behaviors derived from the stochastic approaches will be equivalent to those observed in the field. As a result, these approaches allow replacement of the real heterogeneous aquifer with a macroscopic homogeneous equivalent for the purposes of mathematical modeling conveniences and practicality.

The relationship between the asymptotic macrodispersivity tensor (A_{ii}, i = 1 and 2) where 1 and 2 denotes x and z direction, respectively, and the spatial statistics describing the spatial variability of the *K* field have been derived by Gelhar and Axness (1983) under the assumption of uniform flow. For the case of statistical anisotropy in the vertical and horizontal directions where $\lambda_1 > \lambda_2 > \lambda_3$ (Case 2, Gelhar and Axness, 1983), the longitudinal macrodispersivity A_{11} is given by

$$A_{11} = \sigma_Y^2 \lambda_1 \lambda_2 \Big/ \left[\xi^2 \left(\lambda_1^2 \sin^2 \phi + \lambda_2^2 \cos^2 \phi \right)^{1/2} \right] \qquad (1)$$

where

$$\xi = \exp\left[\sigma_Y^2 \left(0.5 - g_{22} \right) \right] \Big/ \left(\sin^2 \phi + \bar{K}_{22} / \bar{K}_{11} \cos^2 \phi \right) \qquad (2)$$

$$A_{11} = \sigma_Y^2 \lambda_1 \qquad (3)$$

and ϕ is the angle in the horizontal plane between the mean flow direction and the longitudinal axis of the effective hydraulic conductivity tensor ($\bar{K}_{ii}$). It should be noted that a longitudinal macrodispersion coefficient calculated using (3) must be augmented by the value of the local longitudinal dispersion coefficient. When the mean flow direction coincides with the λ_1 direction (ϕ = 0^o), the transverse

macrodispersivity values are zero, thus indicating that the transverse macrodispersion process is controlled by local transverse dispersion. The effective hydraulic conductivity tensor used above is given by

$$\bar{K}_{ii} = K_g \exp\left[\sigma_Y^2 \left(0.5 - g_{ii}\right)\right] \quad (4)$$

where the flow integrals, g_{ii} are functions of the correlation lengths defined by Gelhar and Axness (1983).

Using estimated values for λ_1 and λ_2 equal to 60.00 cm and 20.00 cm, respectively, and a variance of $\sigma_Y^2 = 0.87$, as estimated from the two-dimensional variogram analysis, the principal values of the effective hydraulic conductivity tensor are given as K_{11} = 0.096 cm/s and K_{22} = 0.062 cm/s for the permeameter estimates of *K*. The computed value of the longitudinal macrodispersivity using (3) equals 52.08 cm on the basis of the geostatistical parameters derived from the two-dimensional variogram analysis.

Analysis of dipole hydraulic tomography tests

Inverse modeling approach

The steady state hydraulic tomography analysis of dipole cross-hole tests in the sandbox was conducted using a sequential successive linear estimator (SSLE) approach developed by Yeh and Liu (2000). We only provide a brief description of the inversion approach here. The inverse model assumes a steady flow field and the natural logarithm of *K* (ln *K*) is treated as a stationary stochastic process. The model additionally assumes that the mean and correlation structure of the *K* field is known *a priori*. The algorithm essentially is composed of two parts. First, the Successive Linear Estimator is used for each cross-hole test. The estimator begins by cokriging to create a conditional, mean removed ln *K* (*f*, i.e., perturbation of ln *K*) field, using the initial estimate of K_{eff} and observed heads collected in one pumping test during the tomographic sequence. Cokriging does not take full advantage of the observed head values for estimating *K* because it assumes a linear relationship (Yeh and Liu, 2000) between head and *K*,

while the true relationship is nonlinear. To circumvent this problem, a linear estimator based on the differences between the simulated and observed head values is successively used to improve the estimate.

The second step of Yeh and Liu (2000)'s approach is to use the hydraulic head data sets sequentially instead of including them simultaneously in the inverse model. In essence, the sequential approach uses the estimated *K* field and covariances, conditioned on previous sets of head measurements as prior information for the next estimation based on a new set of pumping data. This process continues until all the data sets are fully used.

To obtain a *K* tomogram from the available dipole cross-hole tests, we solve an inverse problem for steady-state flow conditions. Boundary conditions were set to be no-flow for all sides. We created a fine grid with the synthetic aquifer discretized into 11,826 elements and 24,124 nodes with element dimensions of 1.0 cm × 10.2 cm × 1.0 cm.

Inputs to the inverse model include the mean, variance and the correlation scales for *K*, volumetric discharge (Q_n) from each pumping test where *n* is the test number, as well as steady-state head data. The mean, variance and the correlation scales used here are identical to those in the kriging analysis.

For the steady state hydraulic tomography analysis, we used the 8 dipole cross-hole tests (test 1: E2/I47; test 2: E42/I7; test 3: E4/I45; test 4: E15/I34; test 5: E24/I25; test 6: E17/I32; test 7: E23/I26; and test 8: E20/I29) and the corresponding steady-state head observations at the rest of 46 ports during each test as data sets. The steady state head value from each port was obtained by averaging the steady-state portion of the record. We elected to not use the head data from the injection and extraction ports from each test because these ports could be affected by skin effects (Illman et al., 2007).

Prior to the computation of the *K* tomogram with the SSLE algorithm, we pre-processed the hydraulic head data. This is because Illman et al. (2008) found from the analysis of cross-hole pumping test data obtained in a laboratory sandbox aquifer that the signal-to-noise ratio can be critical in inverse modeling of cross-hole pumping test data. Despite calibrating the pressure transducer prior to the start of each pumping test, there is a minute level of drift always present in each of the pressure transducers.

Therefore, we removed this drift by shifting the head value to a common one for all transducers in the sandbox prior to each test. The collected data were additionally processed by taking the average of the steady state head values collected over a period of time in a given monitoring port.

All computations for steady state hydraulic tomography analyses were executed using 44 of 48 processors on a PC-cluster consisting (of 1 master and 12 slaves each with Intel Q6600 Quad Core CPU running at 2.4 GHz with 16 GB of RAM per slave) at the University of Waterloo. The operating system managing the cluster was CentOS 5.3 based on a 64-bit system. The total computational time for inverting data from 8 pumping tests was about 330 minutes (5.5 hours).

Figure B4 is the *K* tomogram obtained by inverting the steady state head data from 8 dipole cross-hole tests. Spatial statistics of this estimated *K* tomogram are the geometric mean (K_G = 0.15 cms^{-1}), the variance ($\sigma^2_{\ln K}$ = 1.70) and the correlation lengths (λ_x = 60 cm and λ_z = 20 cm).

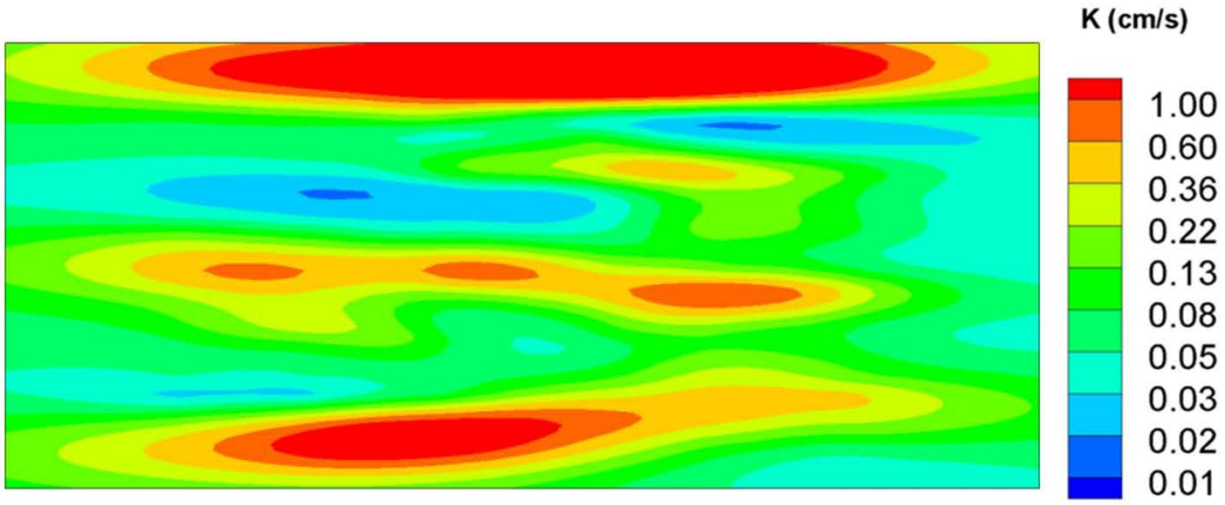


Figure B4: *K* tomograms computed using the steady state hydraulic tomography algorithm of Yeh and Liu (2000).

Results: Dipole Tracer Test

We next report on the results from the dipole tracer test by first examining the evolution of the tracer plume throughout the experiment. To achieve this, numerous photographs were taken during the dipole tracer test. Figure B5 shows six such photos taken at t = 12 min, 27 min, 62 min, 100 min, 175 min, and

286 min. At t = 0 min, the injection of the tracer solution began and at t = 10 min, the injection was switched from tracer solution to tap water.

The photographs were used for qualitative assessment of the tracer migration. Figure B5a was taken at t = 12 min, 2 minutes after the injection of the tracer solution at port 42 completed in layer 5 ended. Layer 5 consisted of a mixture of equal volumes of #14, F75, and 16/30 sands. We notice from this figure that the tracer solution uniformly distributes in layer 5. However, this figure also shows that the tracer solution is preferentially transported along layer 4 (20/40) which has a K of 2.05×10^{-1} cm/s.

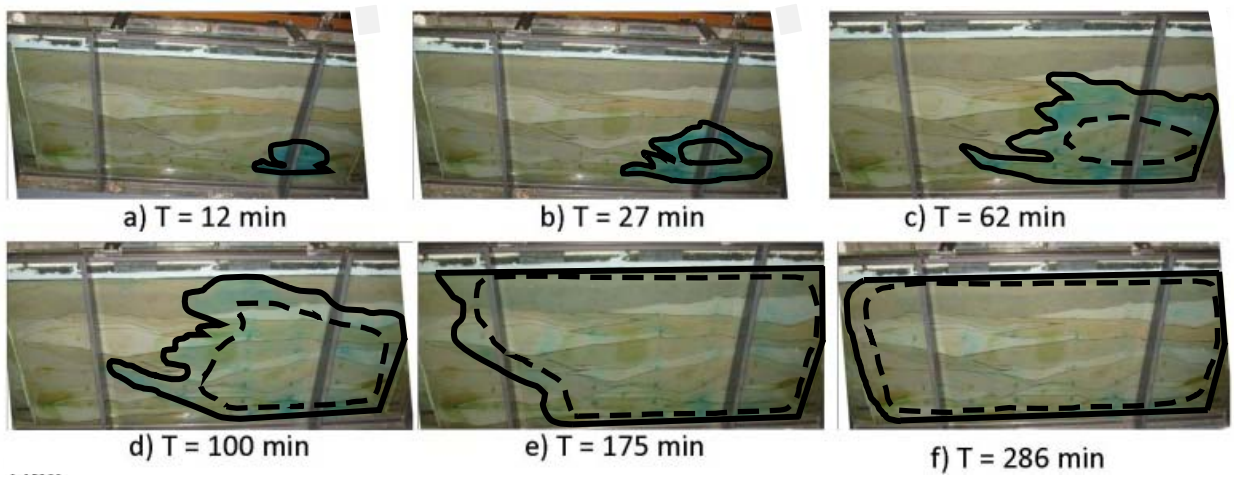


Figure B5: Photographs of sandbox during tracer test 1 at various times: a) t = 12 min; b) t = 27 min; c) t = 62 min; d) t = 100 min; e) t = 175 min; and f) t = 286 min. The solid line represents the edges of the dyed tracer plume. The dashed lines are approximations of the extent of the tracer plume. Note: at t = 175 and t = 286, most of the dye has left the system, however, some dye is trapped around the wells ever after the tracer has been flush through the system giving the tank a blue-green tinge at these times.

At 27 minutes (Figure B5b), 17 minutes after the injection was switched from tracer solution to tap water, we notice that the blue dye becomes cleared around the injection port. Figure B5b shows that the tracer continues to move rapidly through layer 4. The tracer solution also moves laterally and upwards through layer 8 (F32), layer 9 (20/40) and into layer 10 (F65).

At 62 min. (Figure B5c), we notice that the tracer solution becomes more diluted through the injection of tap water. We also notice that the tracer has migrated into layers 13 (20/30) and 15 (20/40). It also begins to migrate into layer 17 (F85) which has the lowest K in the aquifer at 1.35×10^{-2} cm/s.

At 100 min. (Figure B5d), the tracer appears to breakthrough layer 17 (F85) vertically and reaches layer 18 (20/30). Because of the high K of the 20/30 sand (3.12×10^{-1} cm/s), the tracer then rapidly migrates laterally through layer 18. The tracer also appears to migrate rapidly through layer 7 (#12) which has a K of 2.05×10^{-1} cm/s.

By 175 min. (Figure B5e), the tracer appears to reach port 7 which is the extraction port that is completed in layer 14 (F75). We notice that despite the fact the tracer solution appears to be diluted, the tracer plume has reached most parts of the aquifer and all of the layers between the injection and extraction ports.

By 286 min. (Figure B5f), we observe that the tracer has been removed from the high K layers. However, there is a noticeable amount of blue dye in low K layers including: layer 3, layer 8, layer 10, layer 14, and layer 17. This suggests that the low K layers can contribute to store tracers for a long period requiring a long period to flush the tracers out of the synthetic aquifer. Breakthroughs of the tracer during the experiment are shown in Figure B8.

Prediction of dipole tracer test

Description of tracer transport simulations

A two-dimensional, saturated flow and transport model of the synthetic aquifer was developed using the finite-element code MMOC3 (Yeh et al., 1993). For the effective parameter approach, we used the effective hydraulic conductivity (K_{eff}) to simulate groundwater flow during the dipole tracer test, and a longitudinal macrodispersivity (A_{11}) to simulate the migration of the conservative tracer. To address the issue that dipole tracer tests conducted in a bounded aquifer could potentially yield smaller dispersivity estimates, we also conducted an additional simulation with macrodispersivity reduced by 30% to examine its sensitivity to tracer transport. We also simulated groundwater flow and plume migration using the kriged K field and the K tomogram. For the transport simulation using the kriged K field and the K tomogram, we set the dispersivity value equal to zero, while the effective porosity was set to 0.36 for all cases. In total, 3 different cases were considered.

For all cases, flow conditions were established by simulating steady state groundwater flow with injection, extraction, and sampling ports. A tracer solution with a bromide concentration of 151.3 mg/L is then injected for 10 minutes and the solute transport is simulated under transient conditions. As in the experiment, all boundaries are set as no-flow and no-flux boundaries.

Groundwater flow results

We first modeled the steady-state drawdown values during the dipole tracer test at the sampling ports for the effective and heterogeneous *K* distributions. Figure B6 shows scatter plots of simulated versus observed drawdowns for the 3 cases that we examined. In particular, Figure B6a shows the comparison when the K_{eff} computed using the Gelhar and Axness (1983) solution is used to simulate the drawdown behavior. Likewise, Figure B6b shows the same when the kriged *K* field is used for the forward simulation. These comparisons show that the K_{eff} and kriged K field both yield biased predictions of drawdowns at various locations. In addition, we observe a significant scatter suggesting that the K_{eff} and kriged *K* field may not be suitable for predicting drawdowns within this synthetic aquifer. In contrast, Figures B6c shows a much improved prediction of steady state drawdowns using the *K* tomogram. There is some scatter and bias, but it is evident that the *K* tomogram yields considerably improved results in terms of drawdown predictions in comparison to the effective parameter approach.

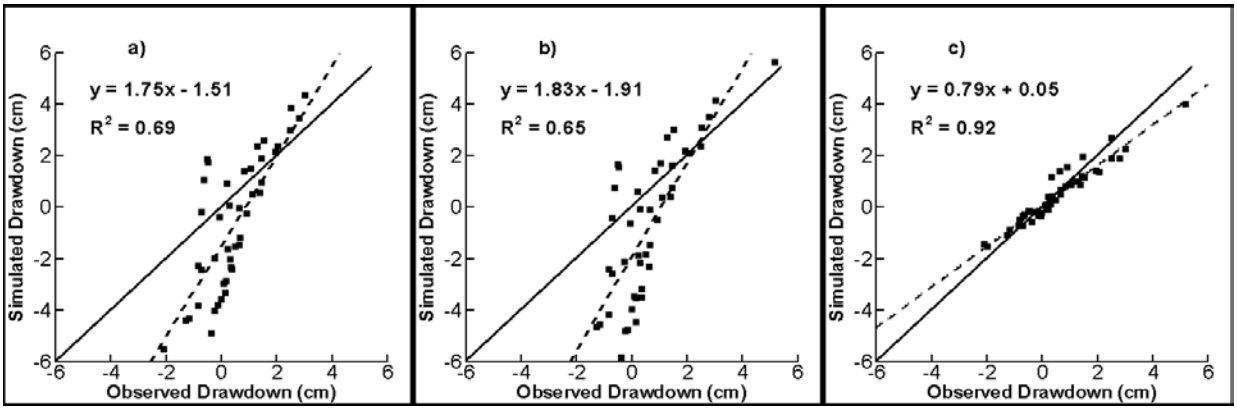


Figure B6: Simulated versus observed drawdowns at 48 ports during the dipole tracer test. Simulated values were computed using: a) K_{eff} from Gelhar and Axness (1983) solution with statistics of permeameter analysis of core samples; b) kriged *K* field; and c) *K* tomogram.

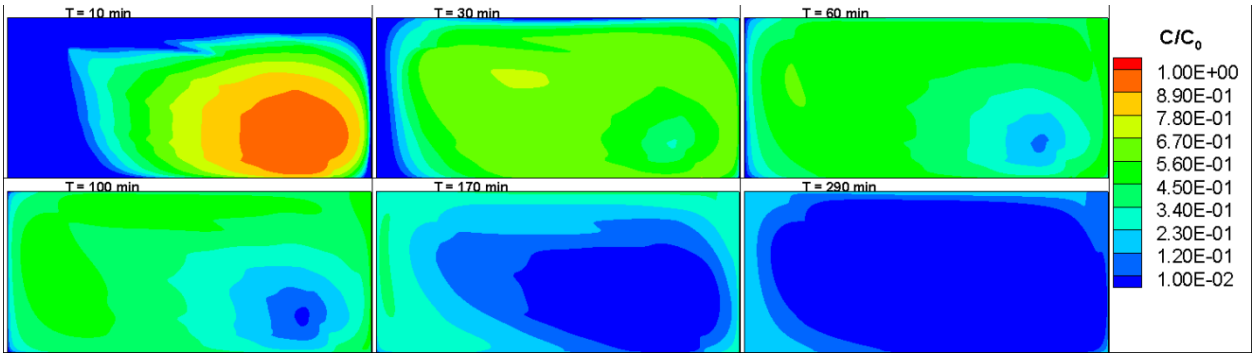
Dipole tracer transport results

We next modeled the tracer migration using the various effective and heterogeneous parameter fields. Figure B7 shows the concentration distributions from the various tracer transport simulations. In particular, Figure B7a shows the spatial distribution of tracer concentrations at various times using the K_{eff} and macrodispersivity estimates. The results reveal that the concentration distribution evolves quite uniformly with significant spreading of the tracer concentration. Not shown here are results of simulations in which we reduced the macrodispersivity by 30%. As expected, reducing macrodispersivity by 30% did not significantly influence our results.

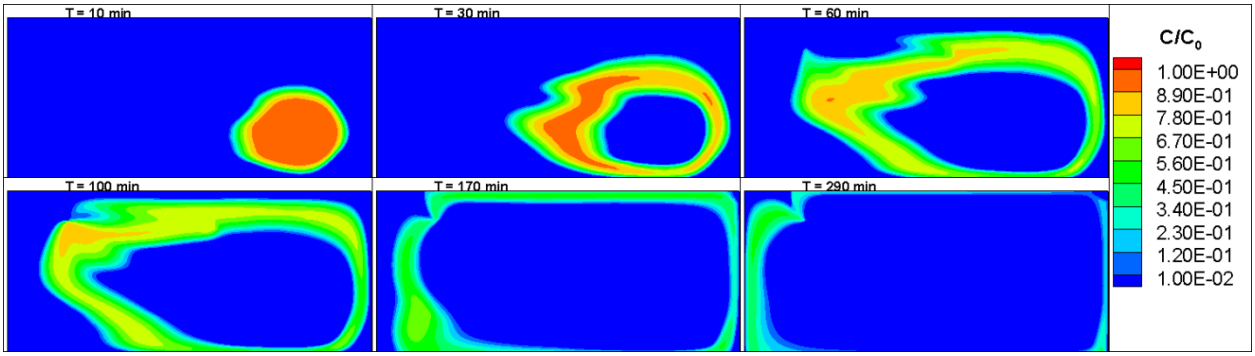
Figure B7b is the result using the kriged *K* field while Figure B7c is the result based on the *K* tomogram. We note that dispersivity was set to zero for both cases. Both Figures B7b and B7c show that due to the mapping of the heterogeneity, the concentration distributions are less smooth in comparison to the effective parameter case (Figure B7a). These figures reveal a much more heterogeneous distribution of tracer concentrations with tracers following preferential pathways with hydraulic tomography providing a more heterogeneous concentration distribution in comparison to kriging.

We next compare the results from Figure B7a through B7c with Figure B5a through B5c. The time when the photographs were taken and the transport simulation results are not exactly coincident, but the two figures nonetheless can be compared qualitatively. The comparison shows that the migration of the tracer represented by the migration of the dye is better represented by Figure B7b and best represented by Figure B7c. This is expected as the transport simulations involving effective parameters do not consider the *K* heterogeneity, while kriging and steady state hydraulic tomography yields increasingly realistic distributions of *K* heterogeneity that can better represent the migration of tracers.

a)



b)



c)

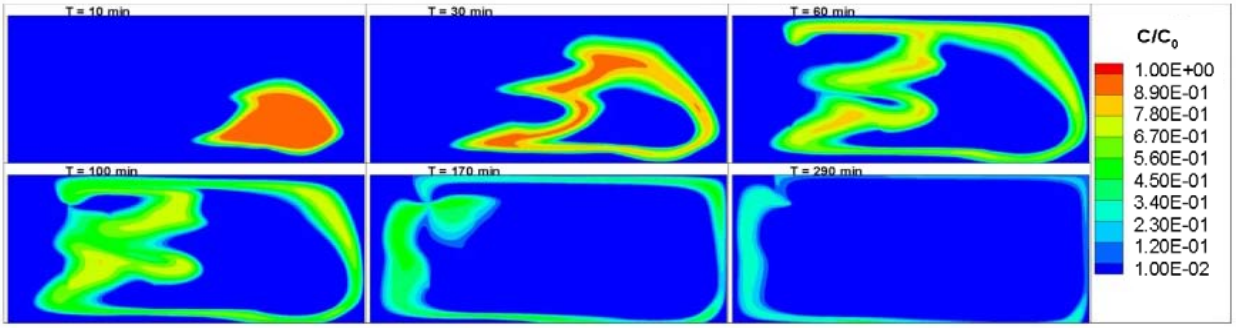


Figure B7: Concentration distributions from tracer transport simulation with: a) K_{eff} and A_{11} computed using Gelhar and Axness (1983) solution with statistics of permeameter analysis of core samples (case 1); b) kriged K field (case 2); and c) K tomogram (case 3).

Simulated versus observed breakthrough curves

We next make a direct comparison of the breakthrough curves obtained through numerical simulations and the actual tracer data (Figure B8). The matches of the breakthrough curves obtained from ports (30, 35, and 47) close to the injection port is quite good for all cases. In particular, the arrival time, peak

concentrations, and time for tracer concentrations to reach background levels is quite consistent for all cases.

Differences in the quality of matches begin to emerge when the breakthrough curves from the intermediate distance ports (10, 12, 22, 27, 37, and 39) are examined. We observe that case 1 (effective parameters) consistently underpredicts the peak concentration at all ports in this category and show an earlier arrival of tracers. In contrast, cases 2 (kriging) and 3 (K tomogram) show a marked improvement in the quality of the fits with case 3 performing the best visually, on the average.

Examination of the furthest ports (2, 14, 19) from the injection port shows mixed results. In general, we observe that case 3 performs better than cases 1 and 2 in terms of better predicting the peak concentrations and arrival times of tracers. However, we also note that the matches are far from perfect.

At the extraction port (7), none of the three approaches yields a perfect prediction, however, we find case 3 captures the first 2 peaks that arrive at different times. Case 1, on the other hand, because it is based on the effective parameter approach, cannot capture the multiple peak behavior. Likewise, case 2 with the smoother K field does not capture this behavior. To compare results quantitatively, we next compute and compare the temporal moments.

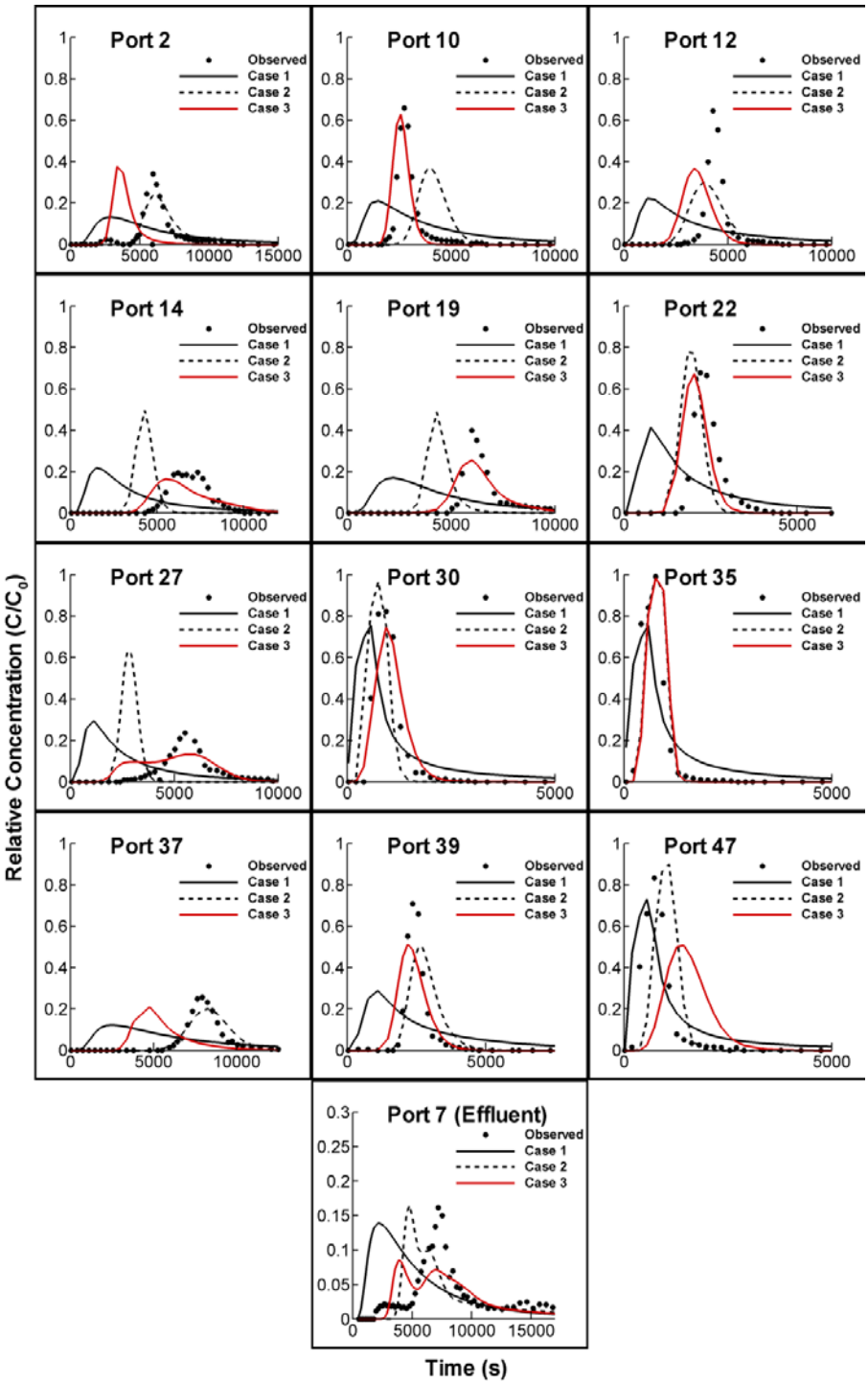


Figure B8: Breakthrough curves from the dipole tracer test and corresponding match of forward simulations. Tracer transport simulations were performed using various *K* fields.

Temporal moment analysis

The method of moments for temporal concentration distribution was used to characterize the bromide breakthrough data at all wells. The n^{th} temporal moments (M_n) of concentration (C) at location (x, y) at time (t) are given by:

$$M_n = \int_0^\infty t^n C(x, y, t) dt \qquad (5)$$

where t is time, $C(x_i, t)$ is tracer concentration. One can compute the zeroth (M_0), first (M_1), and second (M_2) temporal moments using the moment generating function (eqn. 5) by setting $n = 0, 1$ and 2 respectively. The temporal moments were obtained through numerical integration of the breakthrough data using the trapezoidal rule.

The total mass of solute passing through the sandbox at each sampling point is obtained by computing the M_0 from the breakthrough curves. The first normalized moment of breakthrough curves at each sampling ports were used to estimate the mean arrival time of the center of bromide mass (μ):

$$\mu = \frac{M_1}{M_0}. \qquad (6)$$

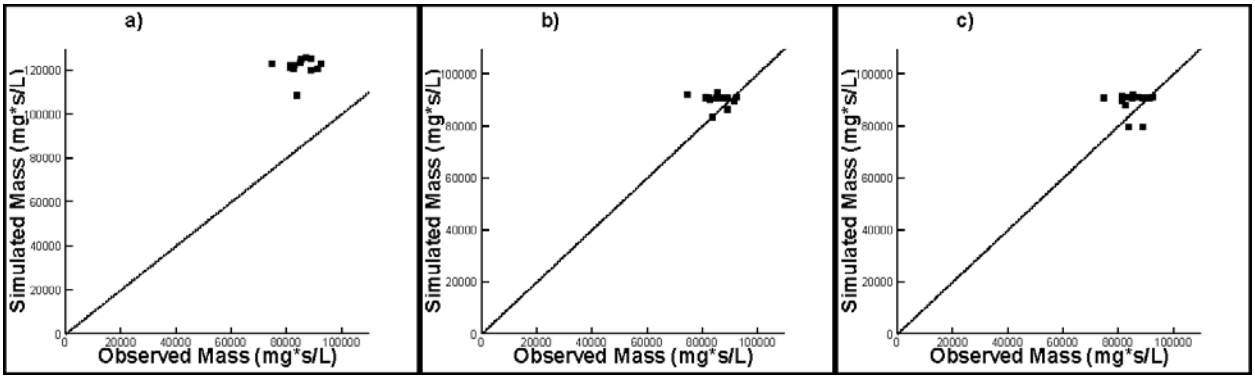
The variance of the breakthrough curve is then calculated by,

$$\sigma = \frac{M_2}{M_0} - \left(\frac{M_1}{M_0}\right)^2. \qquad (7)$$

In general, the variance represents the spread of the concentration distribution and is influenced by mechanical dispersion and molecular diffusion.

Figures B9 through B11 show the M_0, μ, and σ computed from the temporal moment analyses of simulated and observed breakthrough curves. In particular, Figure B9 reveals the estimates of M_0 at these observation ports are, on average, higher for the simulated breakthrough curves for case 1. Figures B9a and B9b indicate that when the heterogeneity in K is considered (i.e., Cases 2 and 3), the estimation of the

total mass at a given sampling point improves. This suggests that the heterogeneity estimated by kriging and HT appears to capture the flow paths of the tracer reaching these sampling ports.



FigureB 9: Total mass computed by the M_0 for each sampling port: a) case 1; b) case 2; and c) case 3.

The arrival time of the center of mass (μ) for the 3 cases is plotted on Figure B10. Results of the three cases show that the arrival times of simulated breakthroughs for the ports close to the injection port are similar to those observed. At ports that are far away from the injection point, the discrepancy between the simulated and observed values increases. While the scatter is centered around the 1:1 line for all cases, case 3 shows considerably less scatter. There are two sampling points 2 and 37 that causes the simulated versus observed relationship to be biased at larger times. These two ports are at a large travel distance and near boundaries. If we choose to disregard these two points, then the relationship between the simulated and observed values becomes quite similar. This suggests that the mapping of heterogeneity via hydraulic tomography yields more accurate estimates of the arrival times of the center of mass for this synthetic aquifer.

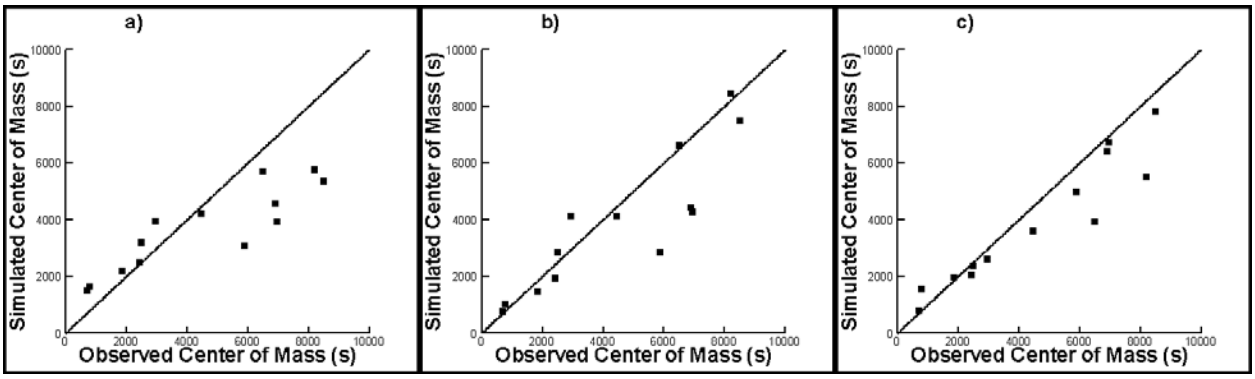


Figure B10: Mean arrival time of the center of mass (μ) computed for each sampling port: a) case 1, b) case 2, and c) case 3.

The variance of the breakthrough curves (σ) was also calculated at each sampling point and plotted on Figure B11. The result suggests that the approach based on effective parameters over predicts the temporal spreading of the plume. The variance estimates at the extraction point for case 1 are, however, satisfactory, suggesting that an integrative behavior (such as breakthrough at the extraction well) can be predicted by low-spatial resolution approach. When we examine case 2 in which the kriged *K* field is used, the bias in the variance estimates diminishes at sampling intervals close to the injection well. In contrast, a remarkable improvement is seen in case 3 in which both the bias and scatter diminish considerably. This suggests that hydraulic tomography can adequately map the heterogeneity, which lead to a higher-resolution prediction of the tracer migration. These results are consistent with those obtained by Yeh et al. (1995) at Georgetown site, where detailed heterogeneity was characterized using 308 slug tests.

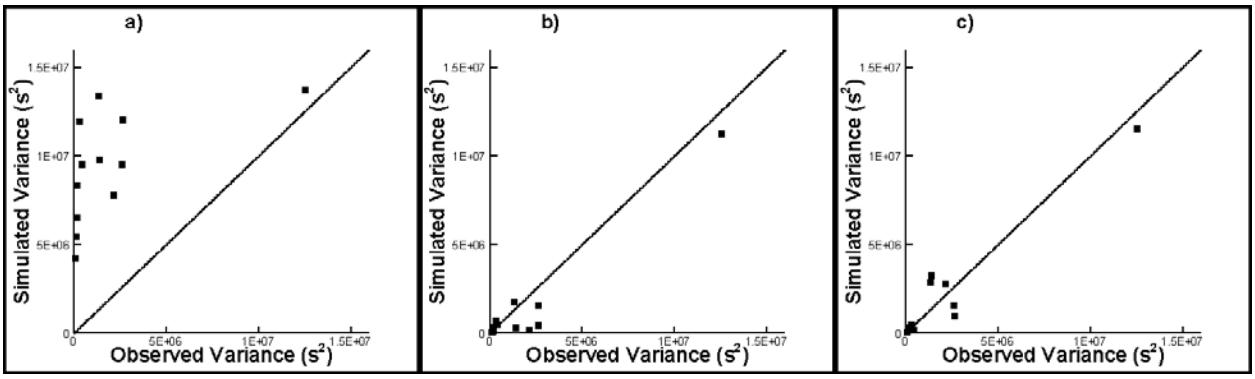


Figure B11: The variance of the breakthrough curve (σ) at each sampling port: a) case 1; b) case 2; and c) case 3.

Discussion and Summary

As stated by Yeh (1992) and demonstrated through field studies (e.g., Sudicky, 1986; Sudicky et al., 2010), the effective parameter and macrodispersion concept is a useful and practical approach for predicting the bulk behavior of the system. In this study, we show that the effective *K* approach resulted in biased predictions of drawdowns during the dipole tracer test. While the macrodispersivity approach yields the general spatial pattern of the tracer migration, the details are quite different from the observed. In addition, lowering macrodispersivity by 30% did not affect our results. This suggests that macrodispersivity is relatively insensitive to the arrival time, peak concentration, and time to which concentration drops to a background level. This is in line with the sensitivity analysis of macrodispersivity conducted by Sudicky et al. (2010) based on data at the North Bay site in Canada.

Results also confirmed that the heterogeneous *K* field estimated from HT yielded a much better prediction of drawdowns during the dipole tracer test than the effective *K* approach and the approach based on kriging and core samples. Both qualitative and quantitative evaluations of predictions of the tracer experiment revealed that while mapping detailed heterogeneity using either kriging/core samples or HT approach can improve the prediction, the approach based on HT yielded a better result. Nevertheless, we must emphasize that the *K* distribution obtained from hydraulic tomography still could not capture all the details of the tracer breakthrough. These results corroborate the findings by Yeh et al. (1995) that prediction of the tracer breakthrough at individual location in a field will be a very difficult task. In

comparison with results of the work by Ni et al. (2009), which evaluated these methods for predicting flow and transport under uniform flow using numerical experiments, we may attribute the discrepancy in this study to 1) effects of noise in head measurements used in the HT, 2) the diverging and converging flow fields forcing tracers to migrate through areas near the impermeable boundaries of the sandbox, where heterogeneity cannot be mapped satisfactorily using HT, 3) our omission of spatial variability of porosity and dispersivity, and 4) the less diffusive nature of the tracer which demands a much higher resolution mapping of the *K* field. Finally, a joint inversion using both HT and tracer (e.g., Yeh and Zhu., 2007, Li and Yeh, 1999) perhaps may be helpful. This is beyond the current manuscript but is a topic for further research in the future.

Finally, our study suggests that accurate predictions of solute transport can become possible without collecting a large number of small scale samples to estimate effective parameters or ascribing dispersivity estimates that are costly to obtain at the field scale through tracer tests. All one has to conduct are well designed hydraulic tomography surveys to capture the *K* heterogeneity pattern at a given site.

Appendix C

Flow to a Well in a Heterogeneous Unconfined Aquifer: Insights From an Intermediate Scale Sandbox

This manuscript by Berg, S.J. and W.A. Illman is currently under review.

EXECUTIVE SUMMARY

Interpretation of pumping tests in unconfined aquifers has largely been based on the use of analytical solutions that disregard aquifer heterogeneity. In this study, we investigate whether the prediction of flow to a well in a heterogeneous unconfined aquifer (and the unsaturated zone above) using variably saturated groundwater flow models is possible. To investigate this question, we designed and conducted laboratory sandbox experiments to characterize the saturated and unsaturated properties of a heterogeneous unconfined aquifer. Specifically, we conducted pumping tests under fully saturated conditions and interpreted the response from multiple monitoring points by treating the medium to be either homogeneous or heterogeneous. Unsaturated zone parameters describing the primary drainage curve were determined using the hanging column method. We then conducted another pumping test and allowed partially saturated conditions to develop (i.e. unconfined pumping test). Simulations conducted using a variably saturated flow model, MMOC3 [*Yeh et al*., 1993] revealed: 1) homogeneous parameters in the saturated and unsaturated zones did not adequately predict the response of the heterogeneous unconfined aquifer and the unsaturated zone above it; 2) heterogeneous hydraulic parameter distributions obtained via transient hydraulic tomography yield significantly improved predictions in the saturated zone; and 3) considering heterogeneity of unsaturated zone parameters produced a minor improvement in predictions in the unsaturated zone, but not the saturated zone. However, the match between simulated versus

observed drawdown responses in the unsaturated zone was far from perfect, indicating a need for inverse techniques to jointly interpret the responses in the saturated and unsaturated zones.

C1. Introduction

In recent decades the topic of flow to wells during pumping in unconfined aquifers has been both a topic of great interest, and, great debate. The need to better understand and interpret the observations from these tests has led to the development of various analytical solutions [e.g., *Boulton*, 1954, 1963; *Dagan*, 1967; *Kroszynski and Dagan*, 1975; *Streltsova*, 1973; *Neuman*, 1972, 1974; *Moench*, 1997; *Mathias and Butler*, 2006]. *Mishra and Neuman* [2010] provided a comprehensive review of these solutions offering their perspectives on the advancement of the theoretical development of unconfined aquifer analysis. *Mishra and Neuman* [2010] also presented an analytical solution for flow to a partially penetrating well in a compressible unconfined aquifer that infers saturated and unsaturated hydraulic properties from drawdowns recorded in the saturated and/or unsaturated zone. Their work is an extension of the solution developed by *Tartakovsky and Neuman* [2007] and adds: i) a more flexible representation of unsaturated constitutive properties; and ii) a finite thickness for the unsaturated zone. The type curve approach was used in conjunction with PEST [*Doherty*, 1994] to simultaneously analyze seven observed drawdown records from a pumping test conducted at the Cape Cod site by *Moench et al*. [2001]. The analysis yielded comparable estimates of hydraulic conductivity (K) and specific storage (S_s) and somewhat higher values of specific yield (S_y) when compared to those obtained by *Moench et al*. [2001] and *Tartakovsky and Neuman* [2007]. Their estimates of the van Genuchten-Mualem parameters were also found to be comparable to laboratory estimates obtained for similar materials in the area. Despite the completeness and usefulness of *Mishra and Neuman's* [2010] solution, analytical solutions that treat the medium to be homogeneous cannot address the issue of heterogeneity. Recently, the homogeneity assumption required by analytical solutions for heterogeneous saturated aquifers was brought into question by *Wu et al*. [2005]. *Wu et al*. [2005] used random Gaussian fields (of transmissivity and storativity) to demonstrate

that parameter estimates made using the *Theis* [1935] solution varied throughout the duration of the pumping test. Transmissivity estimates approached the geometric mean and storativity estimates were dominated by the material between the pumping well and the observation point. It is reasonable to expect similar limitations for analytical solutions used to estimate parameters from unconfined pumping tests.

To date, a limited number of investigations have explicitly examined the effect of heterogeneity on pumping tests performed in unconfined aquifers [e.g., *Bunn et al*., 2010; *Mao et al*., in review]. For example, using multiple realizations of heterogeneous fields, *Bunn et al*. [2010] examined the connection between *K* heterogeneity and the capillary fringe extension phenomenon observed during a pumping test conducted by *Bevan et al*. [2005] at the Borden site in Canada. The ensemble mean hydraulic heads were able to reproduce the field observations quite well, however, these simulations were unable to reproduce the capillary fringe extension observed in the field data.

More recently, *Mao et al*. [in review] used the stochastic moment approach to better understand the role of heterogeneity in the development of the S-shaped drawdown curves and a cross-correlation analysis to show that drawdowns at locations in a heterogeneous unconfined aquifer are mainly affected by local heterogeneity near the pumping and observation wells. *Mao et al*. [in review] also looked at the sensitivity of the S-shaped curve to the spatial variability of hydraulic parameters and concluded that it is most sensitive to hydraulic conductivity (K), specific storage (S_s) and saturated water content (θ_s) and insensitive to the variability of unsaturated hydraulic parameters.

Mapping of *K* heterogeneity in unconfined aquifers was recently reported by *Cardiff et al*. [2009] under steady-state conditions. These authors analyzed multiple drawdown data collected from fully-screened wells completed in an unconfined aquifer located at the Boise Hydrogeophysical Research Site to obtain depth-averaged *K* distributions through their hydraulic tomography algorithm. Various inverse methods have been developed for hydraulic tomography, which utilize pumping test data simultaneously or sequentially [e.g., *Gottlieb and Dietrich*, 1995; *Yeh and Liu*, 2000; *Bohling et al*., 2002; *Brauchler et al*., 2003; *Zhu and Yeh*, 2005, 2006; *Li et al*., 2005; 2007; *Fienen et al*., 2008; *Cardiff et al*., 2009; *Castagna and Bellin*, 2009; *Xiang et al*., 2009; *Liu and Kitanidis*, 2011], but to our knowledge, there are

no hydraulic tomography algorithms that can interpret the transient drawdown records from a heterogeneous, unconfined aquifer. One potential reason for this is that pumping in unconfined aquifers induces flow regimes that are considerably more complex than in confined aquifers, as it involves flow in both the saturated and unsaturated zones with the water table acting as a time-varying interface [*Mao et al*., in review]. In fact, *Mao et al*. [in review] advocates the use of a variably saturated flow equation for the interpretation of drawdown responses in an unconfined aquifer.

In this study, we ask the question: "if we have the necessary parameters and forcing functions to populate a variably saturated flow model, can we accurately model the responses to pumping in a heterogeneous unconfined aquifer in both the saturated and unsaturated zones?" To answer this question, we designed an intermediate-scale laboratory sandbox containing a heterogeneous aquifer that was instrumented with pressure transducers, tensiometers, and water content sensors. A number of cross-hole pumping tests were performed while the tank was fully saturated for characterization purposes. These cross-hole tests were used to estimated both homogeneous and heterogeneous K and S_s distributions. Unsaturated parameters of the different material types in the sandbox were measured using hanging water column experiments.

We next conduct a pumping test in which we allowed the water table to fall, while the drawdown in the sandbox was monitored at multiple locations in the saturated zone with pressure transducers. Concurrently, we monitored the unsaturated zone with tensiometers (equipped with pressure transducers) and water content sensors. After this, we modeled the pumping test using MMOC3 [*Yeh et al*., 1993], a variably saturated flow and transport model to evaluate the ability of different homogeneous and heterogeneous representations of the sandbox aquifer in both the saturated and unsaturated zones to reproduce the observed drawdown responses. By comparing various idealizations of saturated and unsaturated zone parameters, we would determine the parameters having the greatest impacts on making accurate predictions of responses in the unsaturated and saturated zones.

C2. Experimental setup

C2.1 Sandbox design and instrumentation

A synthetic heterogeneous unconfined aquifer was constructed in a vertical, laboratory sandbox. The sandbox was 244 cm in length, 122 cm in height, and had a depth of 9.4 cm. A glass plate covers the front of the sandbox and a stainless steel plate covers the back. A total of one hundred and four ports were drilled into the stainless plate to allow the sandbox to be instrumented with sensors (pressure transducers, tensiometers, and water content sensors).

Fifty eight of these ports were used for the installation of fully penetrating horizontal wells. Each well was constructed by drilling fourteen 0.5-cm diameter holes along a section of brass tubing. The holes were then covered with a stainless steel mesh that was bonded to the tubing with corrosion resistant epoxy. These wells, which penetrate the thickness of the synthetic aquifer were installed after the packing of the sandbox aquifer. Additionally, 47 of these wells contain 0-2 psig pressure transducers (Model S35) from BHL Instruments. Eleven wells did not contain pressure transducers and were still able to be used for pumping purposes.

Twenty two of the ports (located in the upper portion of the sandbox) were used for the installation of column tensiometers (CL-029B Flow Cell Tensiometer, Soil Measurement Systems) equipped with Microswitch pressure transducers. The remaining 24 ports (also located in the upper portion of the sandbox) were instrumented with EC5 water content sensors (Decagon Devices Inc.). All sensors were connected to a National Instruments Compaq Data Acquisition System (NIDAQ) (Model #9178) which allows for real-time monitoring of signals during pumping tests. Figure C1 is a photograph of the sandbox showing the locations of the sensors.

For this particular study, all boundaries except the upper boundary are no-flow boundaries. The top boundary was set as a constant head boundary for the saturated cross-hole tests performed for characterization purposes, and left open to the atmosphere for the unconfined pumping test.

C2.2 Packing of sandbox and types of sands used

Prior to the installation of the wells and tensiometers, and concurrent with the installation of the water content sensors, the sandbox was packed with commercially sieved silica sand (six types; F35, F45, F65, F75, F85, and F110) and silt (four types; Sil-co-sil 45, Sil-co-sil 53, Sil-co-sil 106, and Sil-co-sil 250) from U.S. Silica. This material was packed to match a heterogeneity pattern drawn on the glass side of the sandbox. The sand was slowly added to the sandbox in thin layers (approximately 0.5 to 1 cm thick) and packed between each addition to achieve uniform compaction.

After the sandbox was filled, CO_2 gas was injected to displace the air trapped during packing. Over a period of 3 days, CO_2 was injected at a low pressure (2 - 5 psi) from a number of different ports working from the bottom of the sandbox to the top. The top of the sandbox was sealed except for a small hole that allowed the gas to exit. After flushing with CO_2 de-aired water was added at port 3 under a head difference of approximately 10 - 20 cm. This low gradient meant that any air above the water level would be displaced out the top of the sandbox. Since the air in the sandbox was replaced with CO_2, any trapped gas would readily dissolve into the water phase. This procedure minimized the risk of air entrapment and ensured the sandbox was fully saturated.

The predominant materials in the sandbox are the medium sands F35 and F45. Within these sands are discontinuous layers of finer grained material that average 0.8 m to 1 m in length and average 10 to 15 cm thick. The most notable features of this heterogeneity pattern are the three discontinuous lenses near the top of the sandbox (Sil-co-sil 106 in the top left, Sil-co-sil 53 in the upper middle, and F110 in the top right). These lenses (in particular the silt lenses) are expected to have significantly different saturated and unsaturated properties than the surrounding F35 and F45 sand. Near the bottom of the sandbox, the most salient feature is a layer of silt (Sil-co-sil 250) that extends along the length of the sandbox expect for near the middle where a window is present. This window connects the F35 sand layer below with the F35 layer above.

Figure C1: Photograph of the sandbox showing all sensor locations (● = pressure transducers; ● without pressure transducers; ○ = tensiometers; and **x** = water content sensors) and the various layers that were packed. Sensor locations are approximate.

C3. Characterization of Aquifer Heterogeneity

The analysis of pumping tests performed in unconfined aquifers requires knowledge of both saturated and unsaturated parameters. Our analysis will require knowledge of K, S_s, θ_s (saturated moisture content), θ_r (residual moisture content), and the *van Genuchten* [1980] fitting parameters describing the primary drainage curve (α and n).

C3.1 Characterization of K and S_s

A number of cross-hole pumping tests were performed for characterization purposes while the sandbox was fully saturated. During these tests the top boundary of the sandbox was maintained as a constant head boundary by having 3 cm of standing water in the top portion of the sandbox (above the sand). The extraction line was connected to the desired well and the pump was operated long enough to

fill the effluent line (which discharged into the top reservoir, creating a loop and ensuring mass balance). Once the line was filled the pump was turned off and the system was allowed to return to equilibrium (static), before the cross-hole test was started.

Sixteen ports were chosen for these cross-hole tests (Figure C2). During these tests, one port was pumped, while the pressure responses within the sandbox were simultaneously monitored at 47 points sampling at a rate of 4 Hz. The pumping rates for all cross-hole tests was 240 mL/min. Prior to the cross-hole pumping tests, all pressure transducers were calibrated to ensure accurate data collection. We then collected hydraulic head data for several minutes in all pressure transducers to establish a static, initial condition. After establishment of static conditions, we pumped from each port using a peristaltic pump, while taking head measurements at all 47 ports. For each test, pumping continued until the development of steady state conditions (approximately 1-2 minutes), which was determined by observing the stabilization of all head measurements on the data logger connected to a computer. After reaching steady state, the pump was shut off and recovery data were collected.

Analysis methods for interpreting pumping test data, whether for confined, semi-confined, or unconfined aquifers often assume the aquifer to be homogeneous, however, this is rarely the case. As such, the cross-hole data collected in this sandbox were analyzed in two ways; 1) a single cross-hole test is used to estimate effective K and S_s values (section 3.1.1) and, 2) eight cross-hole tests were used to create heterogeneous K and S_s estimates using hydraulic tomography (section 3.1.2). The remaining eight cross-hole tests were then used to validate these transient hydraulic tomography generated heterogeneous fields.

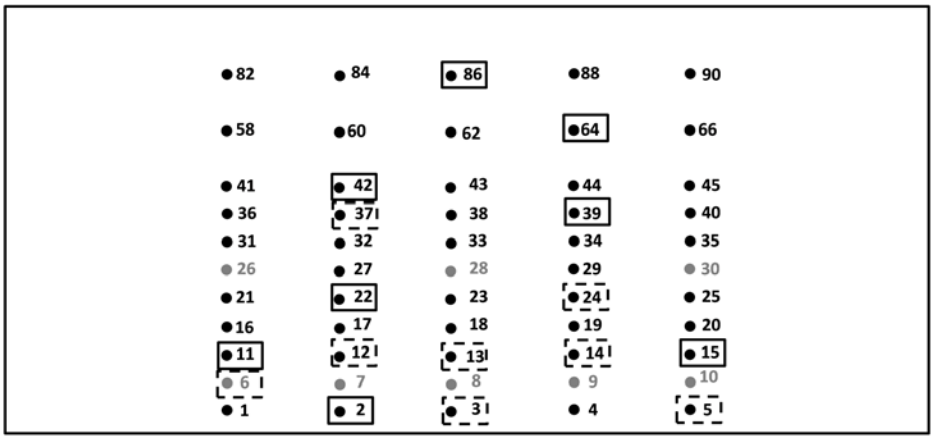


Figure C2: Schematic diagram showing the ports that were pumped for the cross-hole pumping tests. Solid black circles represent locations of pressure transducers and solid grey circles represent ports that were not instrumented. Solid squares are pumping locations used for the estimation of heterogeneous K and S_s fields. Dashed squares are pumping locations for the validation of the results from transient hydraulic tomography.

C3.1.1 Estimation of Effective K and S_s

Using a cross-hole pumping test performed at port 22, the parameter estimation program PEST [Doherty, 1994] was coupled with MMOC3 [*Yeh et al.*, 1993] to estimate effective K and S_s values representing the entire domain by fitting all the drawdown curves simultaneously. The synthetic aquifer was discretized into 3,645 elements and 7,544 nodes with average element dimensions of 3.0 cm × 9.4 cm × 2.5 cm. The top boundary was set to be a constant head boundary and the remaining boundaries of the sandbox were no-flow boundaries. Six data points (at t = 1, 5, 10, 20, 40, and 80 seconds) from 46 ports totalling 276 data points were used for parameter estimation. The initial estimates of K and S_s were 0.1 cm/s and 1 x 10^{-4} cm^{-1} respectively.

The estimated effective K was 1.85 x 10^{-2} cm/s (with lower and upper 95% confidence bounds of 1.76 x 10^{-2} cm/s and 1.94 x 10^{-2} cm/s respectively), and the estimated effective S_s was 3.94 x 10^{-5} cm^{-1} (with lower and upper 95% confidence bounds of 3.40 x 10^{-5} cm^{-1} and 4.57 x 10^{-5} cm^{-1} respectively). The estimate parameters were then used to simulate the cross-hole test at port 22, essentially testing the calibration of the effective parameters. Figure C3 is a calibration scatter plot (observed vs. simulated drawdown) for the cross-hole test at port 22. This plot shows the ability of these effective values to

reproduce the observed drawdown. The dashed line represents the 1:1 line which would be a perfect match between simulated and observed. The linear model fit (solid line) has a slope near 1 suggesting that the effective values capture the average response in the sandbox for this cross-hole test. The scatter as indicated by the coefficient of determination (R^2) suggests that the details of the heterogeneity are not accurately captured. This is expected since homogeneous values of K and S_s were used for this simulation.

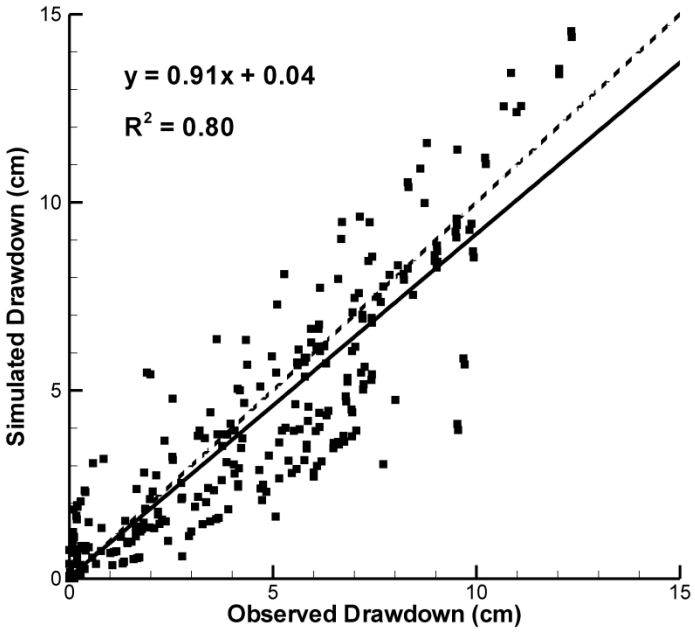


Figure C3: Simulated versus observed drawdown for the cross-hole test at port 22 using effective K and S_s values.

C3.1.2 Estimation of Heterogeneous K and S_s

The use of hydraulic tomography for the characterization of heterogeneity in sandbox aquifers has been shown to be a robust technique that outperforms other methods which often rely on the interpolation of point scale data [*Illman et al*., 2007; *Liu et al*., 2007; *Illman et al.* 2010; and *Berg and Illman*, in review]. Thus, we elect to use transient hydraulic tomography [*Zhu and Yeh*, 2005] for the characterization of K and S_s within this sandbox.

The Sequential Successive Linear Estimator (SSLE) code developed by *Zhu and Yeh* [2005] was used to conduct transient hydraulic tomography with multiple cross-hole tests performed under fully saturated conditions to obtain a heterogeneous distribution of K and S_s (from now on referred to as K and S_s tomograms). The model domain used for this inversion is identical to the one used for estimation of effective K and S_s described in section 3.1.1.

Inputs to the inverse model include initial guesses for the K (0.1 cm/s) and S_s (1 x 10^{-4} cm^{-1}), estimates of variances (σ^2_{lnK} = 3.0 and σ^2_{lnSs} = 3.0) and estimated correlation scales for both K (λ_x = 150 cm and λ_z = 20 cm) and S_s (λ_x = 150 cm and λ_z = 20 cm), volumetric discharge (Q_n) from each pumping test where n is the test number, available point (small-scale) measurements of K and S_s, as well as head data at various times selected from the head-time curve. Although available point (small-scale) measurements of K and S_s can be input to the inverse model, we do not use these measurements to condition the estimated parameter fields to test the inversion algorithm.

In total, we used 8 independent cross-hole tests with pumping taking place at ports 2, 11, 15, 22, 39, 42, 64 and 86 for the transient hydraulic tomography analysis. More specifically, we used 6 data points (at t = 1, 5, 10, 20, 40, and 80 seconds) from 46 ports totalling 276 data for each test. We excluded the data from the pumped port because of possible skin effects. In total, we used 2208 data points from 8 different tests in our transient inversions.

All computations for transient hydraulic tomography analyses were executed using 44 of 48 processors on a PC-cluster consisting (of 1 master and 12 slaves each with Intel Q6600 Quad Core CPU running at 2.4 GHz with 16 GB of RAM per slave) at the University of Waterloo. The operating system managing the cluster was CentOS 5.3 based on a 64-bit system. The total computational time for inverting data from 8 pumping tests was approximately 24 hours.

The resulting K and S_s tomograms are shown on Figure C4a and C4b. A visual comparison of the K tomogram to the deposits (Figure C1) shows that many of the important features are captured. Of particular note are the low K zones which compare very favourably with the real distribution. The two low K layers extending towards the middle from both sides near the bottom of the sandbox are present, as

well as the two low *K* zones near the top left (there is even some indication in the tomogram that these may not be connected, which is the case). Additionally, the Sil-co-Sil 106 layer near the middle right of the sandbox is captured, however, it does not extend all the way to the boundary. This is not surprising considering there are no observation points in this region, thus, estimates here have greater uncertainty. The *K* tomogram also preserves the connectivity of the high *K* units, particularly the window in the silt (Sil-co-sil 250) layer near the bottom middle, and the large window in the silt near the top right. Thus, we feel confident that the *K* tomogram reflects the heterogeneity of the sandbox.

Examination of the S_s tomogram (Figure C4b) reveals a trend of increasing S_s values with increasing elevation within the sand box. Such a trend was also observed in the S_s values from separate sandbox studies packed using different methods, and containing different heterogeneity patterns [*Liu et al*., 2007; *Berg and Illman*, in review]. A likely explanation for this pattern being observed in several sandbox studies is that the material near the bottom of the sandbox is much less compressible due to the weight of the overlying material, thus, resulting in low S_s values near the bottom of the sandbox and higher S_s values near the top. Additionally some heterogeneity in the distribution of S_s is present within this trend from top to bottom, however, it does not appear to be strongly correlated with any of the features seen in the *K* tomogram (Figure C4a) or the picture of the sandbox (Figure C1). This suggests that the vertical trend in S_s is the dominant feature controlling the drawdown response in the sandbox.

To quantitatively assess the results, we plotted the simulated and observed drawdown responses (Figure C5a) for each cross-hole test used for the inverse modeling effort (i.e., calibration) on separate scatter plots. A linear model (solid line) was fit to each case without forcing the intercept to zero. All 276 points included (6 data points at t = 1, 5, 10, 20, 40, and 80 seconds from 46 ports) are presented on the scatter plots. The dashed line represents the 1:1 line which would be a perfect match between simulated and observed. The linear model fit (solid line and the coefficient of determination (R^2) provided on each plot give an indication of scatter and bias. The R^2 is a statistic that provides a quantitative measure of similarity between the simulated and measured drawdown values. A high R^2 value means that the

simulated and observed drawdown values are linearly correlated, even though the mean values could be different.

Illman et al. [2007, 2008, 2010] and *Liu et al*. [2007] previously found that the best way to test the results from hydraulic tomography was to use the hydraulic parameter distributions in predicting one or more independent drawdown events. Therefore, to further assess the reliability of the K and S_s tomograms, we simulated 8 additional cross-hole tests performed at ports 3, 5, 6, 12, 13, 14, 24, and 37 using the forward groundwater model MMOC3 [*Yeh et al*., 1993]. The results were then assessed by plotting individual scatterplots of simulated and observed drawdown (Figure C5b) and fitting a linear model as described earlier. The scatter plots in Figure C5b show that the K and S_s tomograms were able to accurately predict independent cross-hole tests. The largest error is seen for the independent cross-hole test at port 5. There are 6 points that are significantly below the 1:1 line were measure at port 4 (adjacent to port 5). This suggests that the heterogeneity in this portion of the sandbox may not be accurately captured. Despite this, the remaining ports for the cross-hole test at port 5 are matched accurately suggesting that the rest of the sandbox is fairly accurately characterized. In addition, the other validation tests in Figure C5b show good agreement between observed and simulated suggesting that the heterogeneity pattern of the sandbox is accurately captured.

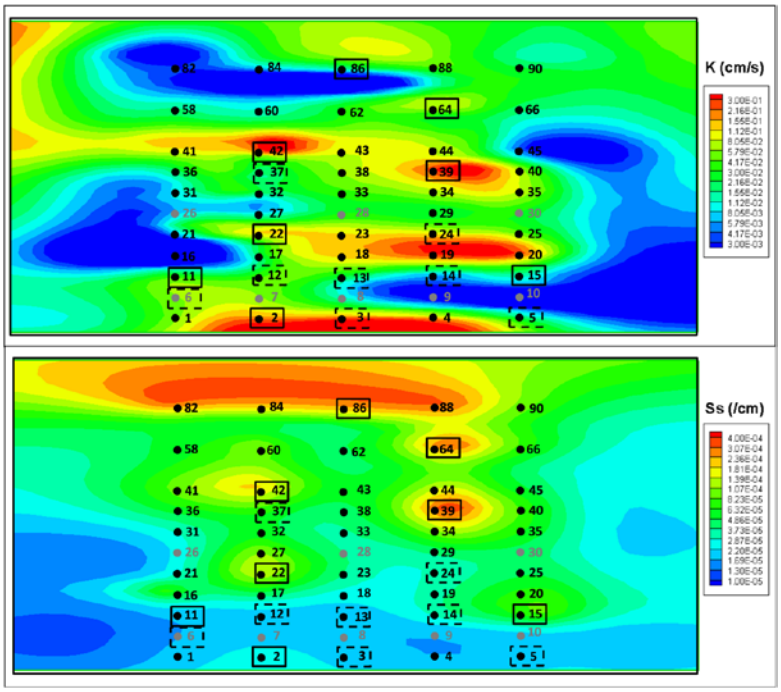
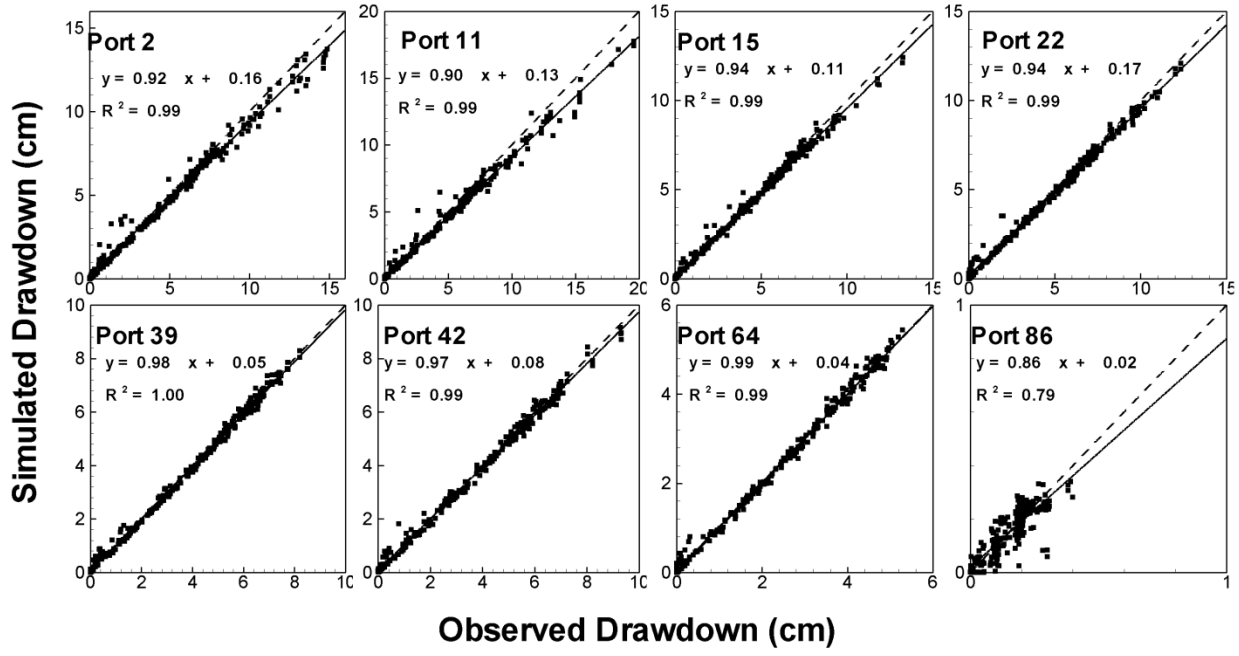


Figure C4: a) K and b) S_s tomograms computed using the transient hydraulic tomography algorithm of *Zhu and Yeh* [2005] SSLE with 8 cross-hole pumping tests. Solid squares are pumping locations used for transient hydraulic tomography dashed squares are pumping locations for the validation of the results from transient hydraulic tomography.

a)



b)

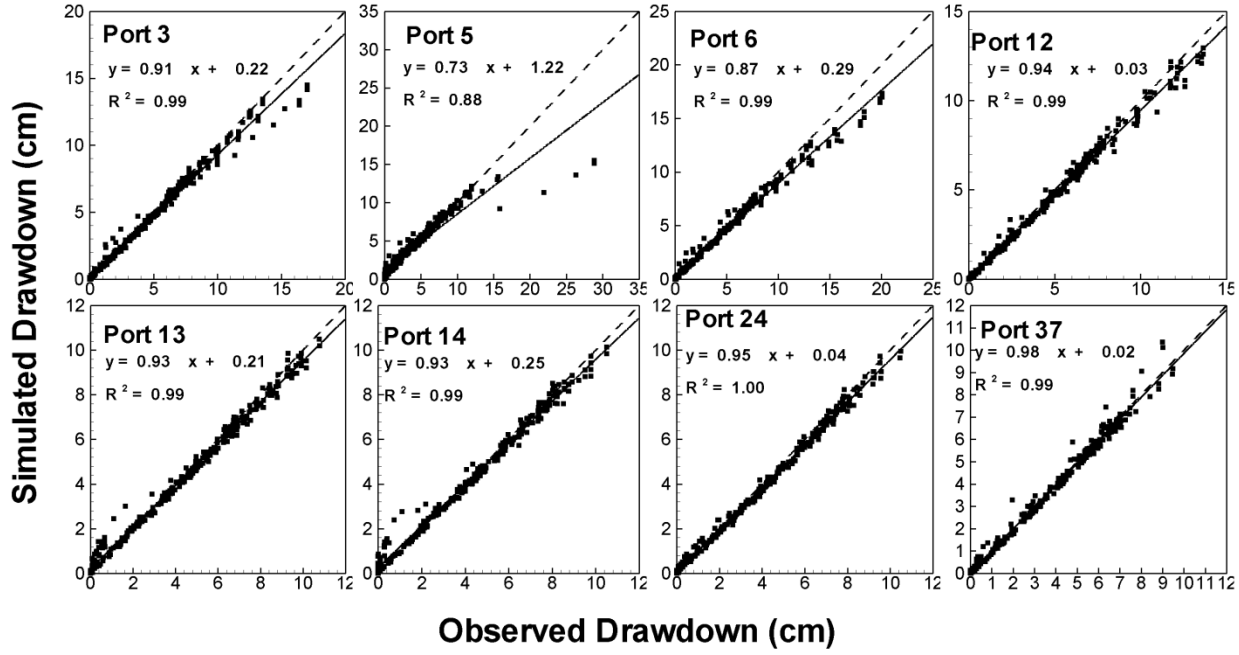


Figure C5: a) Simulated versus observed drawdowns from 8 cross-hole pumping tests used for calibration purposes and b) additional tests used for the validation of the K and S_s tomograms.

3.1.3 Estimation of θ_s, θ_r, α and n

Using the hanging water column method [*Stephens*, 1995] the primary drainage curve was measured for five of the materials packed in the sandbox (F35, F45, F110, Sil-co-Sil 53, and Sil-co-Sil 106). These materials were selected because they are located in the upper portion of the sandbox and are likely to experience negative pressure heads (and as a result could possibly release water) when the water table is lowered during drainage or an unconfined pumping test.

Negative pressure head was increased incrementally until either residual water content was reached, or a value of -100 cm was achieved. A value of -100 cm was selected because it is unlikely that values greater than this will be seen during unconfined pumping or drainage tests in the sandbox. The *van Genuchten* [1980] model was fit to the primary drainage curves for each material by varying the α and n parameters using a generalized reduced gradient nonlinear routine implemented in Excel (except for Sil-co-Sil 53 and 106 which were fit manually because the solver failed to converge for these two cases). In

this analysis, the residual water content was estimated from the moisture characteristic curve and fixed during the regression. The model fit to the moisture characteristic curve, optimized parameters and the associated coefficient of determination (R^2) are presented on Figure C6. The parameter values obtained for each of these materials (Figure C6a-e) can be used to construct a heterogeneous case for the model domain, however, this level of detail is rarely available and a single 'effective' drainage curve would be used to describe the entire domain.

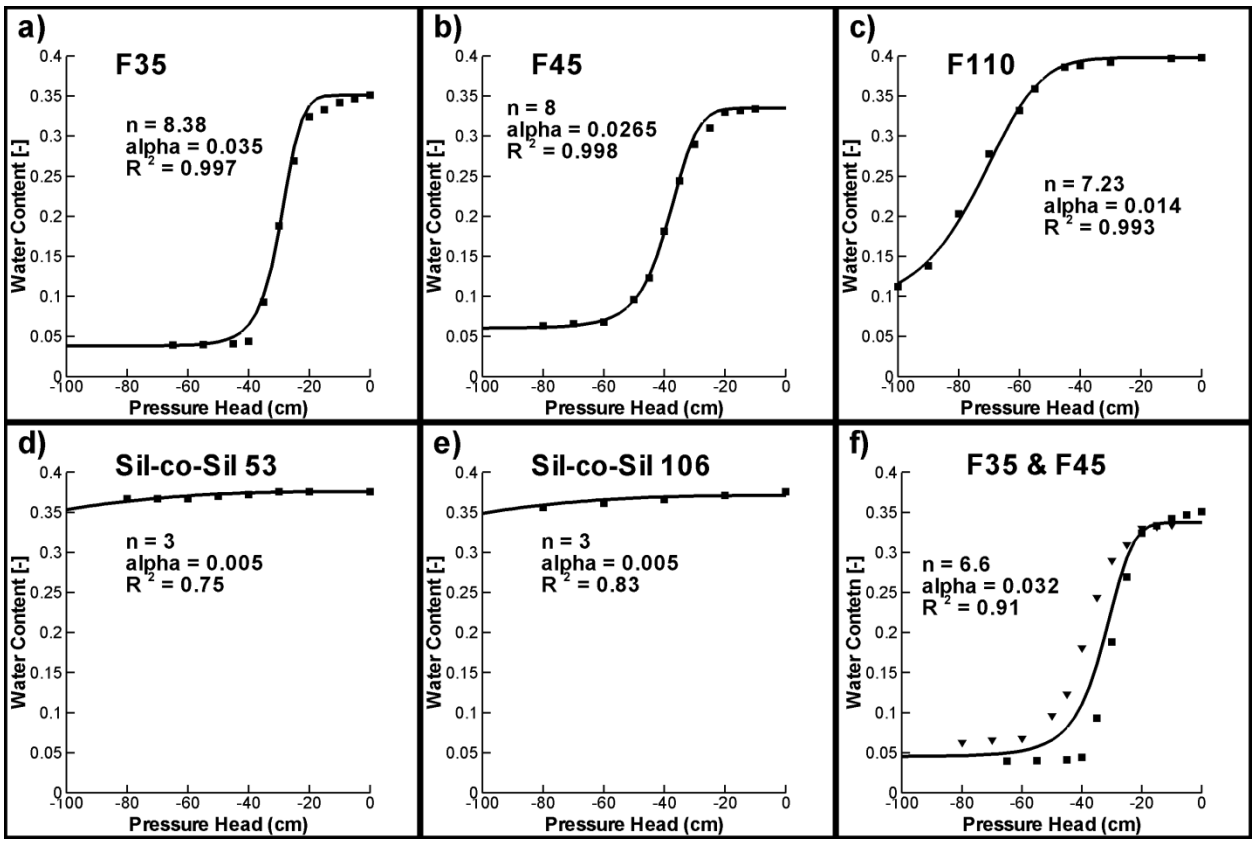


Figure C6: Moisture characteristic curves determined through the hanging column method of: a) F-35; b) F-45; c) F110; d) Sil-co-Sil 53; e) Sil-co-Sil 106; and f) F35 and F45 matched simultaneously. The squares represent F35 and gradient symbols represent F35.

To obtain an effective α and n representing the entire aquifer domain, we fit the *van Genuchten* [1980] model to the primary drainage curves for both F35 and F45 simultaneously (Figure C5f). These two sands were selected because they comprise the majority of the material in the unsaturated zone and

will therefore largely control the drainage response. The other materials present in the unsaturated zone (F110, Sil-co-Sil 53, and Sil-co-Sil 106) were not included in this simultaneous match as they are only present in minor portions of the unsaturated zone and are present as discontinuous lenses. Since they are present as discontinuous lenses, it is possible that when the coarser F35 and F45 material drains, these lenses will become disconnected from the main aquifer and will not be able to effectively release their water. Thus, only the two dominant material types in the upper portion of the sandbox are selected for this effective parameter estimation.

C4. Unconfined Pumping Test

After the sandbox was characterized using cross-hole tests while fully saturated the water table was allowed to fall under unconfined conditions. The objective of this test was to collect data from both the saturated and unsaturated zones during a pumping test so that we can compare the observations to variably saturated flow simulations constructed from our characterization efforts. Figure C7 is a schematic diagram showing the locations of pressure transducers, tensiometers, and water content sensors installed in the sandbox as well as the port pumped for this test.

The pumping test started with the sandbox fully saturated and the water table at the same elevation as the top of the sand. This was confirmed visually using a manometer to get an independent reading of the static water level. Prior to the beginning of the test all sensors were monitored for several minutes to check for sensor stability and to record baseline data. This baseline data was then use as the t = 0 reference to which the data collected during the test could be compared. Port 3 (located at the bottom middle of the sandbox) was then pumped at a rate 60 ml/min for 7.5 hrs. The signal from all of the sensors (pressure transducers, tensiometers, and water content sensors) was recorded for the duration of the test.

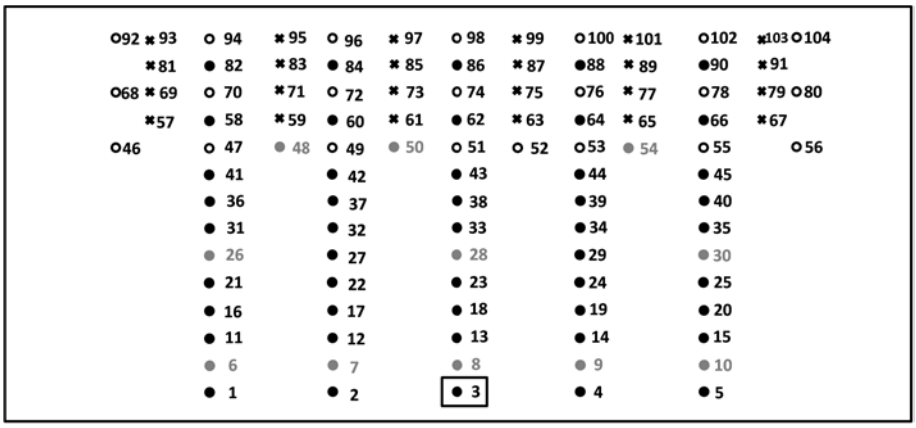


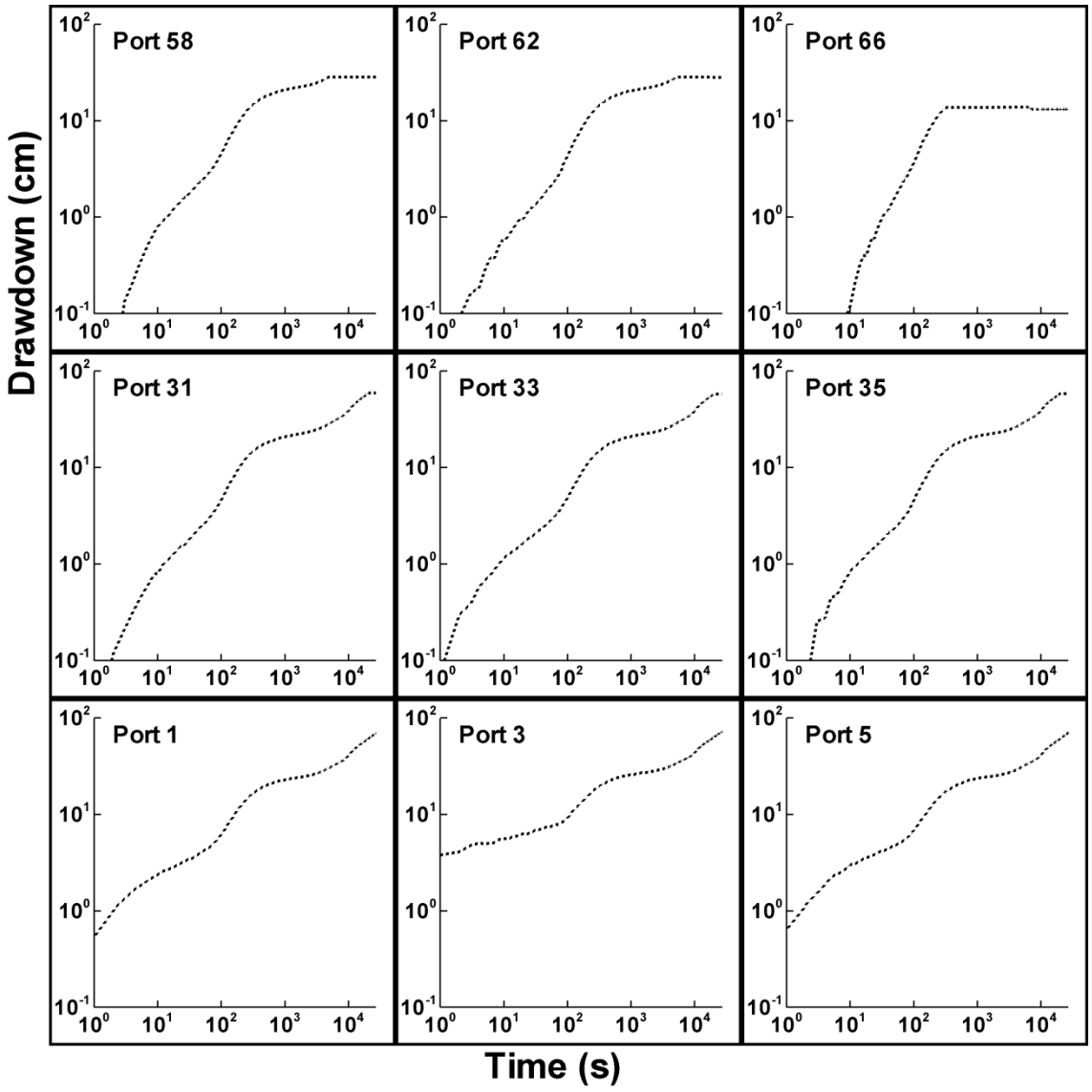
Figure C7: Schematic diagram of the sandbox showing an array of sensors (● = pressure transducers; ● = port without pressure transducers; ○ = tensiometers; and **x** = water content sensors) used to monitor the pumping test in a heterogeneous unconfined aquifer. The box indicates the port at which the unconfined pumping test was performed.

Figures C8 shows the drawdown as measured at selected pressure transducers (Figure C8a) and tensiometers (Figure C8b). The ports in these figures are organized such that their positions relative to each other reflect their relative positions within the sandbox. For example, Port 58 is located left of Port 62 and above Port 31. Figure C7 can be used as a reference to determine the absolute location of each port within the sandbox. The response measured at all ports is available online as supplementary Figures SC1 (note: for the purposes of this appendix the supplementary figures are provided at the end of this appendix).

The pressure transducers (Figure C8a) are only able to measure the pressure response within the saturated zone. As the test progresses and the water table lowers, some of the upper wells drain and the pressure transducer is no longer able to measure a change in pressure. The plateau in the signal for the upper ports (Ports 58, 62, and 66) in Figure C8a reflects this process. The shape of the drawdown curves are very different from what would be expected for a fully saturated system and reflect the unconfined nature of this test. The early portion of the curves display characteristic S-shaped behaviour of pumping tests performed in an unconfined aquifer. The S-shape is more pronounced for deep ports (further below the water table) than for shallow ports. This observation is consistent with that of *Nwankwor et al.,* [1992] and *Bevan et al.,* [2005] . At late time, the observed drawdown increases again, deviating from the

classical S-shaped unconfined drawdown curve. We attribute this to the lack of a constant head boundary which means at late time the no flow boundaries dominate the observed response and additional drawdown occurs.

a)



b)

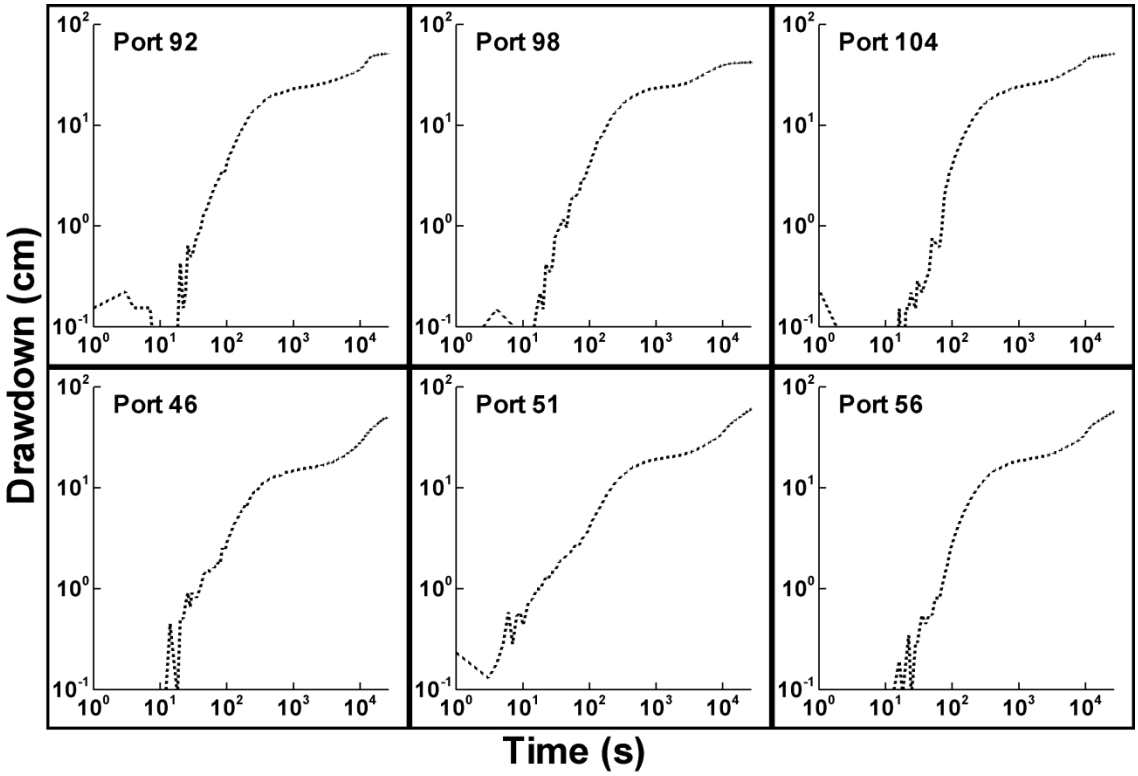


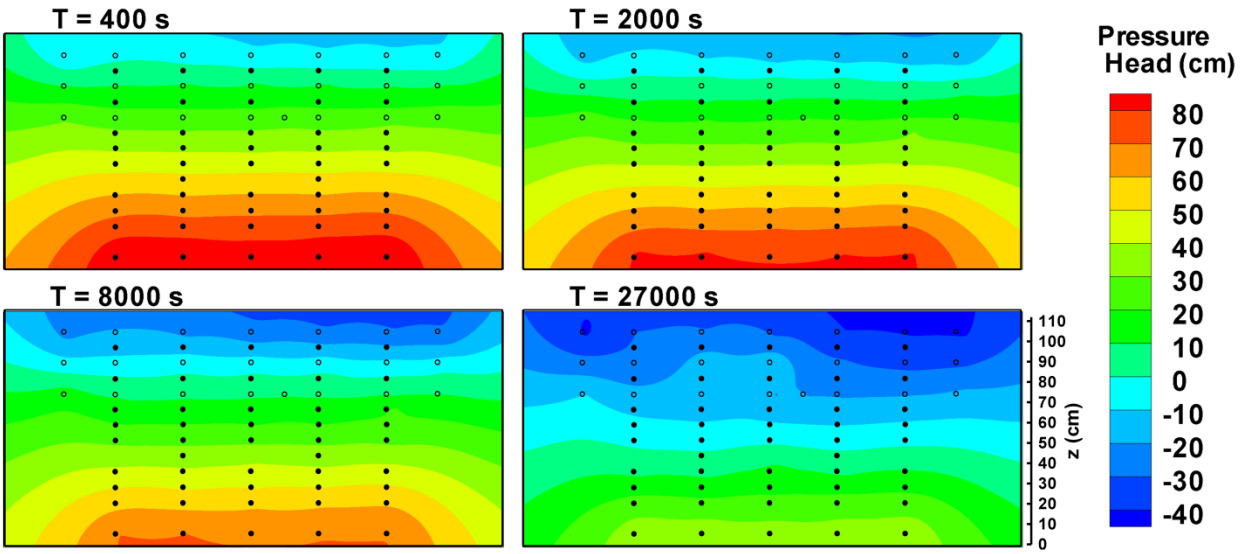
Figure C8: Observed drawdown during the unconfined pumping test at a) 9 pressure transducer ports; and, b) 6 tensiometers. The responses measured at all ports are available online as Supplementary Figures SC1a and SC1b.

The tensiometer responses measured during this test are shown in Figure C8b. Since the water table at the start of the test was coincident with the top of the sand in the sandbox the tensiometers were initially located below the water table. By calibrating the tensiometers for this range, these tensiometers record a continuous signal as the water table drops past them (i.e. they record both positive and negative pressure head). Thus, unlike the pressure transducers installed in wells (Figure C8a), the tensiometers are able to record a signal for the entire duration of the test. Initial inspection of Figure C8b seems to suggest that the S-shaped curve is present in the tensiometer response. However, upon closer inspection the S-shape at early time is missing and it is the additional late time drawdown caused by the no flow boundaries that creates the impression of an S-shape at late time. Some of the bottom tensiometers (Ports 46 to 52) appear to have a slight S-shape at early time, however, it is not very prominent and sensor noise

makes it difficult to make any concrete statements about this. Pressure transducers at similar elevations (Figure C8a: Ports 58 - 66) do not have a very pronounced S-shape at early time. Tensiometer data collected by *Nwankwor et al.,* [1992] during a pumping test at CFB Borden shows a very slight S-shaped response at early time only for ports near the pumping well. Thus, the absence of an S-shaped response in the tensiometer data here is not surprising. Ports 92, 98, and 104 (located in the top row of the sandbox) appear to plateau at late time. This likely indicates that the soil they are in is now disconnected from the main body of water, thus, the continued drop of the water table no longer influences the soil in this region. Figure C9 shows spatial distributions of pressure head (Figure C9a) and water content (Figure C9b) at t = 400, 2,000, 8,000, and 27,000 seconds after pumping began at port 3. The pressure and water content data were interpolated through kriging. We point out that the curvature of pressure head seen along the side boundaries is an artefact of the interpolation procedure. Figure C4a reveals that based on the pressure head distributions, an unsaturated zone begins to develop as early as t = 400 seconds. Subsequent pressure head distributions reveal that the water table drops with time and at t = 27,000 seconds, we notice the impact of the heterogeneity (i.e., the low *K* zone) on the contours of the pressure head. Based on this figure the water table is approximately 65 cm from the top of the aquifer.

Figure C9b shows the spatial distribution of the volumetric water content profiles in the upper half of the sandbox. Volumetric water content is generally uniform at the beginning of the test, but we see the impacts of pumping at t = 400 seconds in which the water content begins to decrease near the top of the aquifer. A slightly higher region of water content is visible near the top of the aquifer indicating the presence of a low *K*, high porosity layer. As the pumping test continues, the pore space drains throughout the aquifer except for the low *K* lens which retains some of its water due to capillarity. By t = 27,000 seconds the upper portion of the aquifer is at or near residual water content (shown as dark blue) although some water is still retained in the low *K* lens.

a)



b)

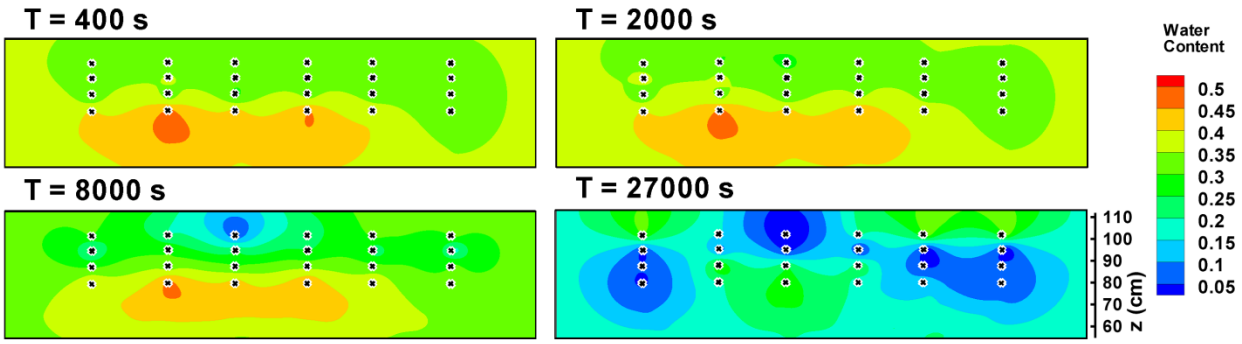


Figure C9: Spatial distribution of: a) pressure head and b) volumetric water content in the upper half of the sandbox with time during the unconfined aquifer pumping test at port 3. Symbols indicate the position of various sensors used to monitor the pumping test (● = pressure transducers; ○ = tensiometers; and **x** = water content sensors)).

C5. Variably saturated flow modeling of pumping test in an unconfined aquifer

We next utilize the homogeneous and heterogeneous values of K, S_s, n, and α for predicting the response of the unsaturated and saturated zones during the unconfined pumping test that we described in section 4. A finite-element numerical model, MMOC3 developed by *Yeh et al*. [1993] is used for these calculations. This program solves the partial differential equation that describes flow in 3-D, variably saturated geologic media:

$$\nabla \cdot [K(\psi, \mathbf{x}) \nabla(\psi + z)] = \omega S_s(\mathbf{x}) \frac{\partial \psi}{\partial t} + \frac{\partial \theta}{\partial t} = \left(\omega S_s(\mathbf{x}) + C(\psi, \mathbf{x})\right) \frac{\partial \psi}{\partial t} \quad (1)$$

where ∇ is the spatial gradient, $K(\psi, \mathbf{x})$ is the unsaturated hydraulic conductivity function, ψ is the pressure head where it is positive below the water table, while negative above the water table, $\mathbf{x}$ is spatial coordinates, z is the elevation, ω is the saturation index, $S_s(\psi, \mathbf{x})$ is specific storage, t is time, θ is volumetric moisture content, and $C(\psi, \mathbf{x})$ is the soil moisture capacity derived from the moisture/pressure constitutive relationship. The saturation index is equal to 1 when the porous medium is fully saturated and zero when it is unsaturated. The program employs the Newton-Raphson iteration scheme to solve the nonlinear finite element approximation of equation (1).

We use MMOC3 [*Yeh et al.*, 1993] to simulate flow to a well due to pumping in a 2-D unconfined aquifer. The dimension of the aquifer is 2.44 m × 0.094 m in the horizontal plane and 1.12 m in the vertical and is discretized into 3,645 finite elements and 7,544 nodes. A non-uniform mesh is used for the discretization in which some elements are larger than others so that the elements line up properly with the ports.

A no-flux boundary is assigned to the top (no infiltration or evaporation) and bottom of the aquifer; no-flux boundaries are imposed on the other five sides of the aquifer. The initial pressure head distribution in the aquifer is set to be hydrostatic with the water table at z = 1.12 m, representing a static condition. Pumping takes place at Port 3, located 7.3 cm from the bottom center of the aquifer with a constant pumping rate 60 cm^3/min. Because pumping takes place at a point, we neglected borehole storage which we consider to be a minor effect.

The hydraulic conductivity-pressure head and moisture-pressure head constitutive relationship of the aquifer are described by the model by *van-Genuchten* [1980]:

$$\theta(\psi) = \theta_r + (\theta_s - \theta_r)(1 + \alpha|\psi|^n)^{-m} \quad (2)$$

where | | is absolute value, θ_s is the saturated moisture content, θ_r is the moisture content at residual saturation and α, *n*, and *m* are shape-fitting parameters with *m*=1-1/*n*. We further assume that the *K*(*h*) follows *Mualem's* [1976] pore-size distribution model expressed as:

$$K_r(\psi) = \frac{\left\{1-\left(\alpha|\psi|\right)^{n-1}\left[1+\left(\alpha|\psi|\right)^{n}\right]^{-m}\right\}^2}{\left[1+\left(\alpha|\psi|\right)^{n}\right]^{m/2}} \qquad (3)$$

where K_s is the locally isotropic saturated hydraulic conductivity; and α, *n*, and *m* are assumed to be the same as those in equation (2). The K_s and S_s were estimated earlier from the cross-hole pumping tests described in section 3.1.2., while the unsaturated zone parameters for the van Genuchten-Mualem model were obtained through the hanging water column (3.1.3).

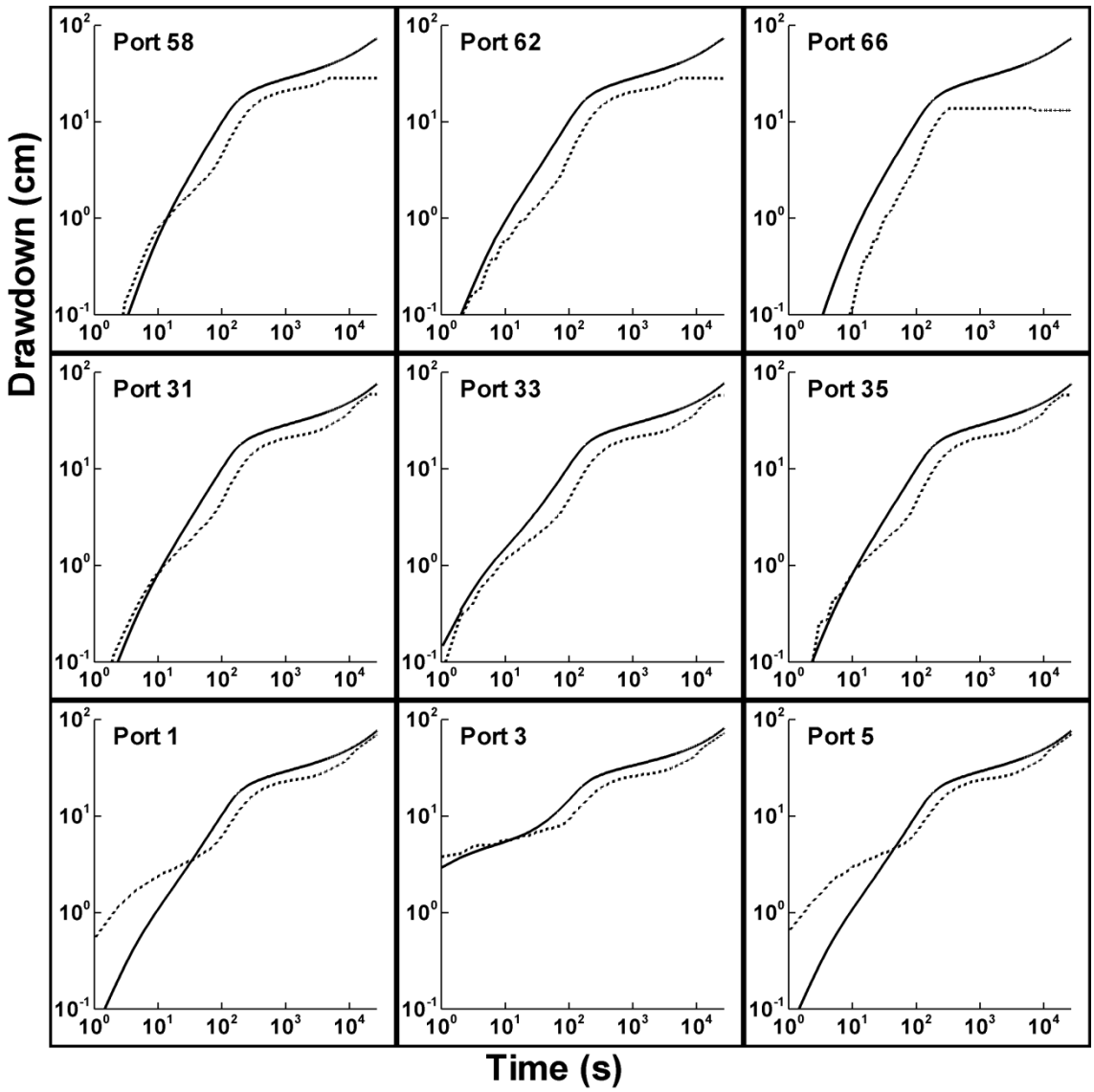
5.1 Case 1: Homogeneous saturated and unsaturated zone parameters

In Case 1, we obtained effective *K* and S_s estimates for the cross-hole test performed at Port 22 by coupling MMOC3 with the parameter estimation PEST. The resulting PEST estimated *K* and S_s values were 1.85 x 10^{-2} cm/s and 3.94 x 10^{-5} cm^{-1} respectively. We also used the van-Genuchten model parameters ($\alpha = 0.032$ and $n = 6.6$) estimated from a simultaneous match of sands F35 and F45 (Figure C5f) to obtain a homogeneous estimate for the unsaturated zone parameters.

Forward simulation results for the pumping test under unconfined conditions are shown on Figure C10. In particular, Figure 10a shows the simulated (solid line) versus the observed (dashed line) drawdowns from selected pressure transducers in the saturated zone, while Figure C10b shows a similar comparison for pressure transducers installed in the unsaturated zone. Figures C10a and C10b reveal that the comparison between the simulated and observed drawdowns in both the saturated (Figure C10a) and unsaturated (Figure C10b) zones are poor. The simulated drawdown is generally greater than the observed, especially at intermediate and late time. Early time drawdown at pressure transducers are overestimated at some ports and underestimated at others. This is attributed to the homogeneous *K* and S_s

field which, captures the heterogeneity in an average sense, and thus, is not expected to accurately simulate the response at all ports. Supplementary Figure SC2 shows the observed and simulated drawdown response for all pressure transducers and tensiometers.

a)



b)

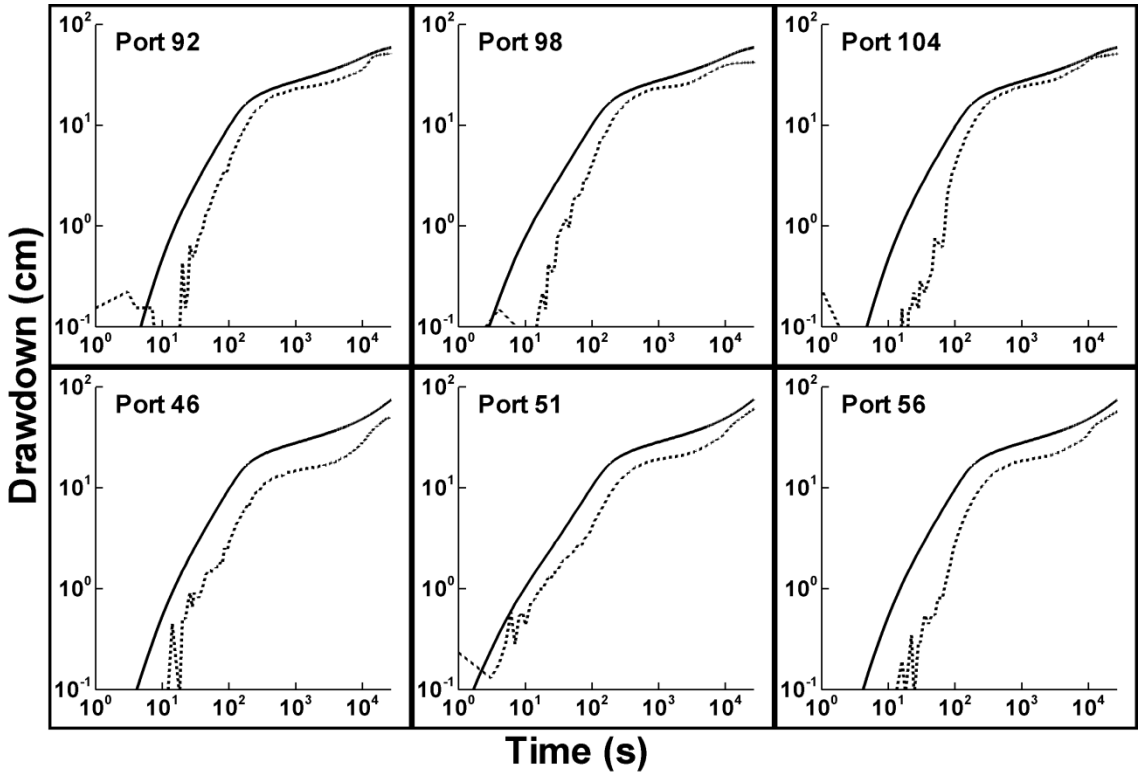
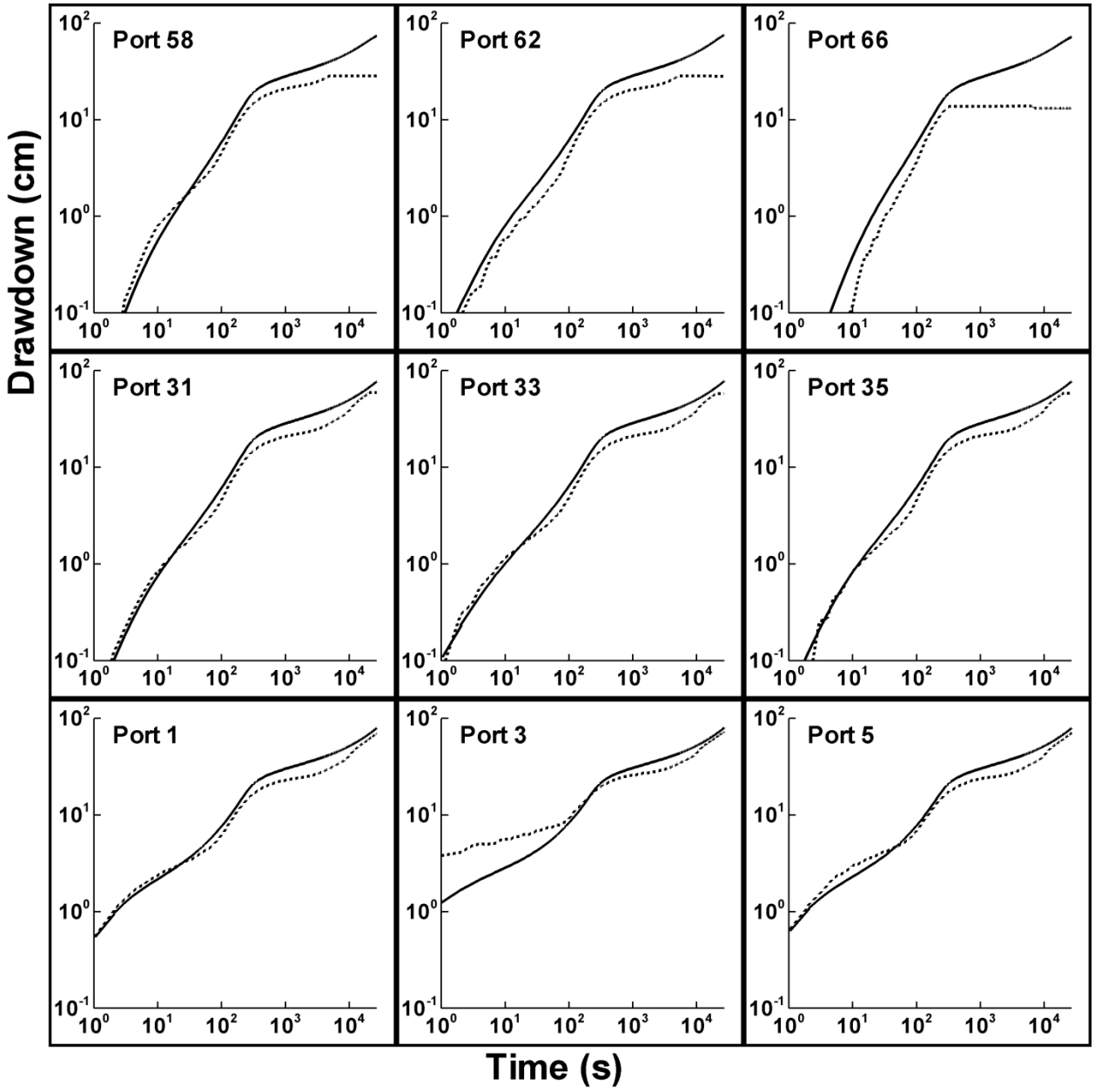


Figure C10: Simulated (solid line) and observed (dashed line) drawdown from selected: a) pressure transducers; and, b) tensiometers during the unconfined aquifer pumping test (Case 1). A comparison at all ports is available online as Figure SC2.

5.2 Case 2: Heterogeneous saturated parameters and homogeneous unsaturated zone parameters

In Case 2, we obtained the K and S_s fields through transient hydraulic tomography of 8 separate cross-hole tests. The treatment of unsaturated zone parameters is identical to Case 1 (effective homogeneous values). Forward simulation results shown on Figures C11a for the saturated zone reveals a marked improvement in the prediction of drawdowns at all times with the variably saturated flow model. On the other hand, we observe that the drawdown predictions in the unsaturated zone (Figure C11b) are slightly improved but still not very good showing that the simulated values generally exceeding the observed values especially at early time. This discrepancy could be attributed to a number of factors such as heterogeneity in unsaturated zone parameters and also due to a delay in response of the tensiometers.

a)


Port 58
Port 62
Port 66
Port 31
Port 33
Port 35
Port 1
Port 3
Port 5
Drawdown (cm)
Time (s)

b)

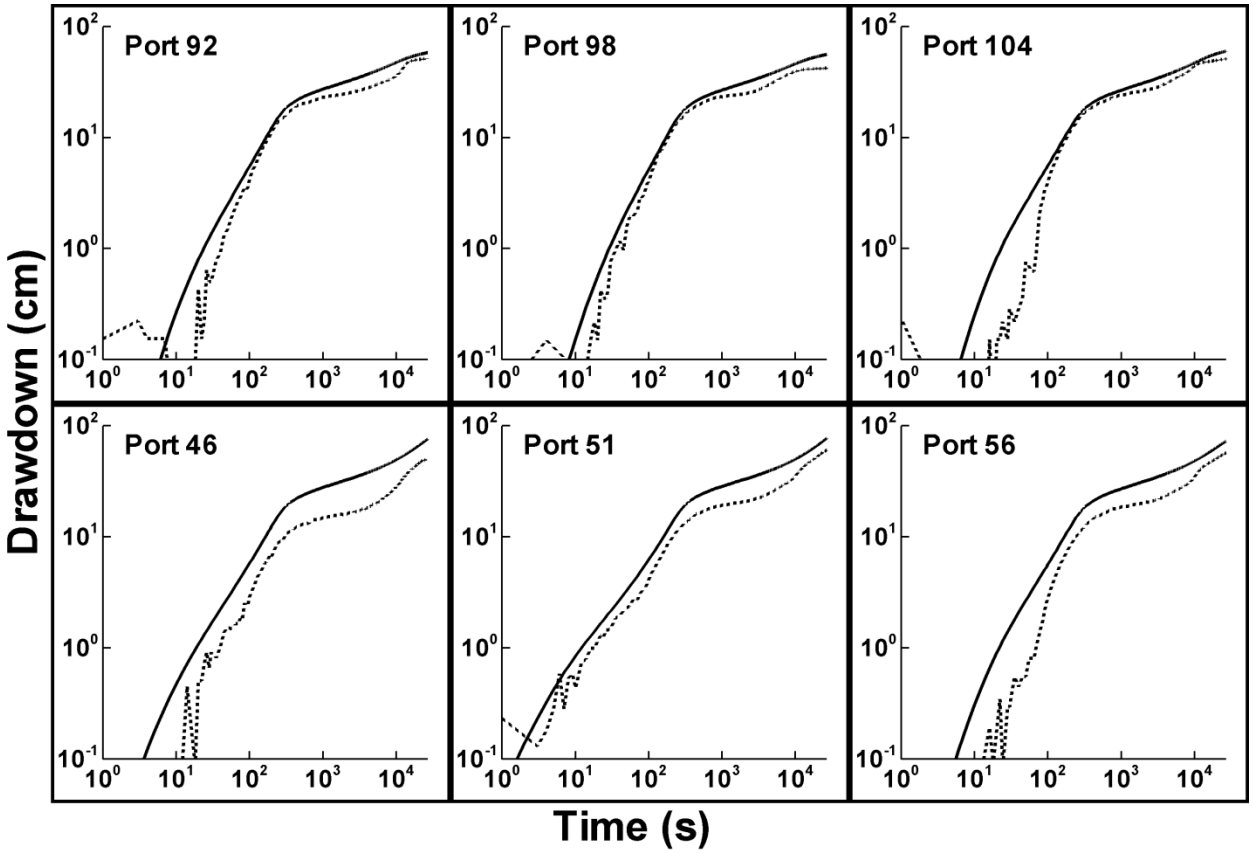


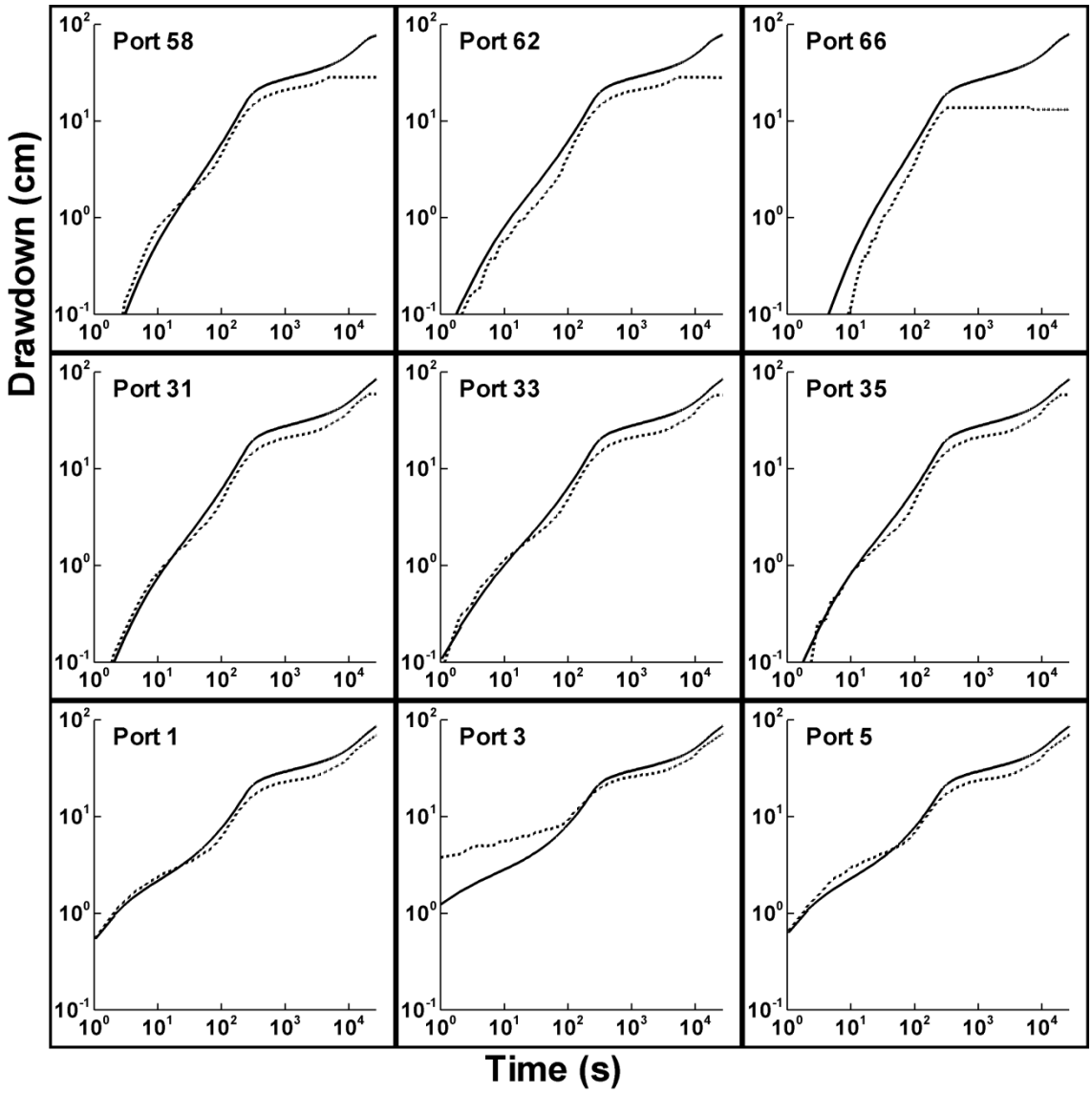
Figure C11: Simulated (solid line) and observed (dashed line) drawdown from selected: a) pressure transducers; and, b) tensiometers during the unconfined aquifer pumping test (Case 2). A comparison at all ports is available online as Figure SC3.

5.3 Case 3: Heterogeneous saturated parameters and heterogeneous unsaturated zone parameters

To investigate the cause of discrepancy of drawdowns observed in the unsaturated zone, we also considered in Case 3, the variability of the unsaturated zone parameters by assigning homogeneous values for each layer. Figure C12a reveals that the quality of matches has virtually not changed in comparison to Figure C11a perhaps suggesting that heterogeneity in unsaturated zone parameters may not be that critical in responses in the saturated zone. This observation is in agreement with the findings from a numerical study conducted by *Mao et al*. [in review]. Predictions of drawdown in the unsaturated zone show only a slight improvement, suggesting that as long as the K and S_s are accurately know and the unsaturated

parameters are represented accurately (in an effective sense), it is possible to predict unconfined pumping tests using a variably saturated flow model.

a)



b)

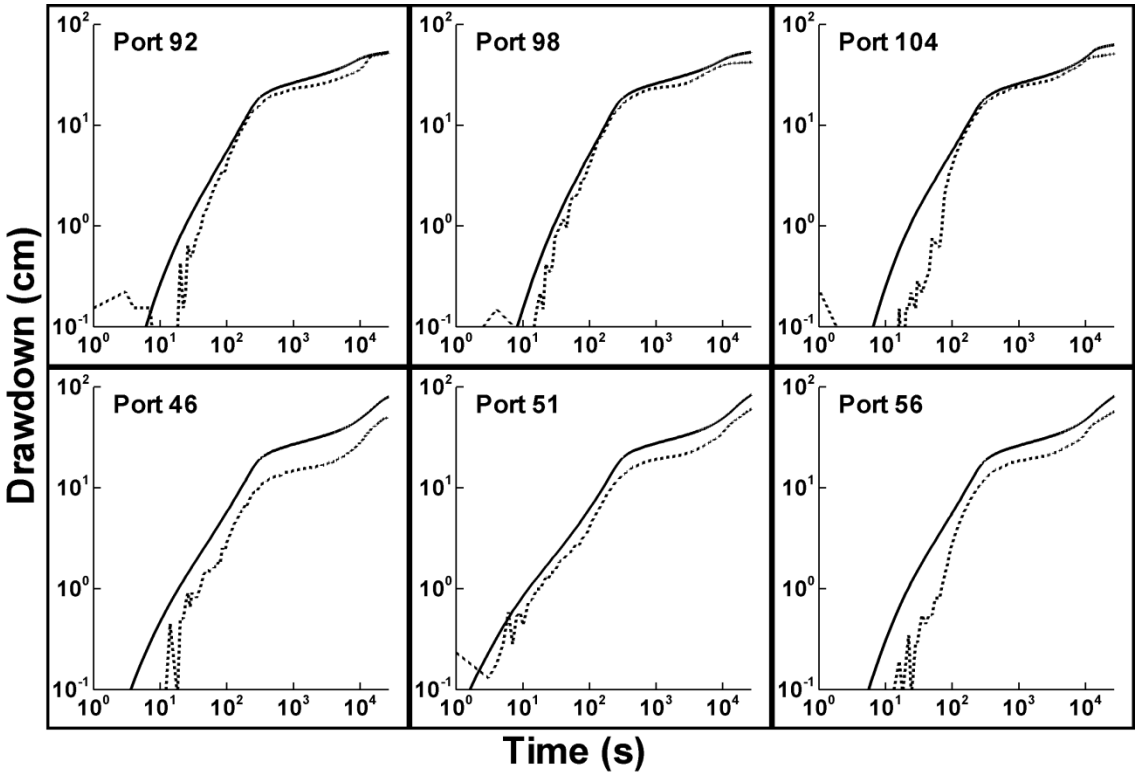


Figure C12: Simulated (solid line) and observed (dashed line) drawdown from selected: a) pressure transducers; and, b) tensiometers during the unconfined aquifer pumping test (Case 3). A comparison at all ports is available online as Figure SC4.

C6. discussion and Summary

Flow to wells due to the pumping of unconfined aquifers has been a topic of great interest for many decades. The majority of the studies described in the literature have focused on analytical solutions which treat the medium to be homogeneous. However, it is the rule rather than the exception that aquifers are heterogeneous.

We study the impact of the unsaturated zone and effects of heterogeneity on unconfined aquifer flow in a synthetic aquifer packed in an intermediate-scale laboratory sandbox. The synthetic aquifer was characterized initially with eight cross-hole pumping tests under fully saturated conditions and the hydraulic data are interpreted using the transient hydraulic tomography code developed by *Zhu and Yeh* [2005]. The resulting hydraulic conductivity (K) and specific storage (S_s) tomograms were validated using

eight additional tests not used in the inverse modeling. Effective parameters of K and S_s were also estimated through the inverse modeling of individual cross-hole pumping tests. Laboratory hanging column experiments of sands in the unsaturated zone were also to determine the van Genuchten parameters. We then conducted another cross-hole pumping test in which a port near the bottom of the aquifer was pumped at a constant rate. During this test, the water table was allowed to freely move downwards in response to pumping. The saturated zone was monitored via pressure transducers and the unsaturated zone with tensiometers and water content sensors.

Forward models of various complexities in saturated and unsaturated parameters were then built using the variably saturated code, MMOC3 [*Yeh et al*., 1993] to examine the sensitivity of the homogeneous and heterogeneous parameter estimates on the predictability of the unconfined aquifer test. Results show that an average or effective K and S_s determined through the averaging of equivalent K and S_s obtained from the inverse modeling of a pumping test cannot yield accurate predictions of drawdown responses in the saturated and unsaturated zones. Our predictions of drawdown responses in the saturated zone improved dramatically when the K and S_s distributions were used in the forward simulation of the unconfined aquifer pumping test. Because we monitored the pressure changes in the unsaturated zone during this experiment, we also attempted to evaluate the predictions of pressure responses in the unsaturated zone. Results show that slightly improved predictions of drawdowns in the unsaturated zone could be achieved if heterogeneous zones were assigned to the model. However, our results showed that predictions of drawdowns in the saturated zone were relatively insensitive to whether we conceptualized the unsaturated zone to be homogeneous or heterogeneous which confirms a conclusion reached by *Mao et al*. [in review] using numerical simulations.

Based on this study it is possible to accurately predict the response of a heterogeneous unconfined aquifer to pumping as long as the saturated parameters (K and S_s) are accurately characterized and an accurate effective value of the unsaturated parameters is known. This study was unique in that the heterogeneity pattern was exactly known and it was possible to select effective unsaturated parameters by assessing would material would dominate the drainage response. In the field however, this is not as

straight forward and the selection of unsaturated parameters could pose a challenge. As such, interpreting pumping tests in heterogeneous aquifers will benefit from inverse modelling of both the unsaturated and saturated zones. This will require the use of a variably saturated model coupled with an inverse algorithm such as the Sequential Successive Linear Estimator [e.g., *Hughson and Yeh*, 2000] for the proper interpretation of flow to wells in a heterogeneous unconfined aquifer.

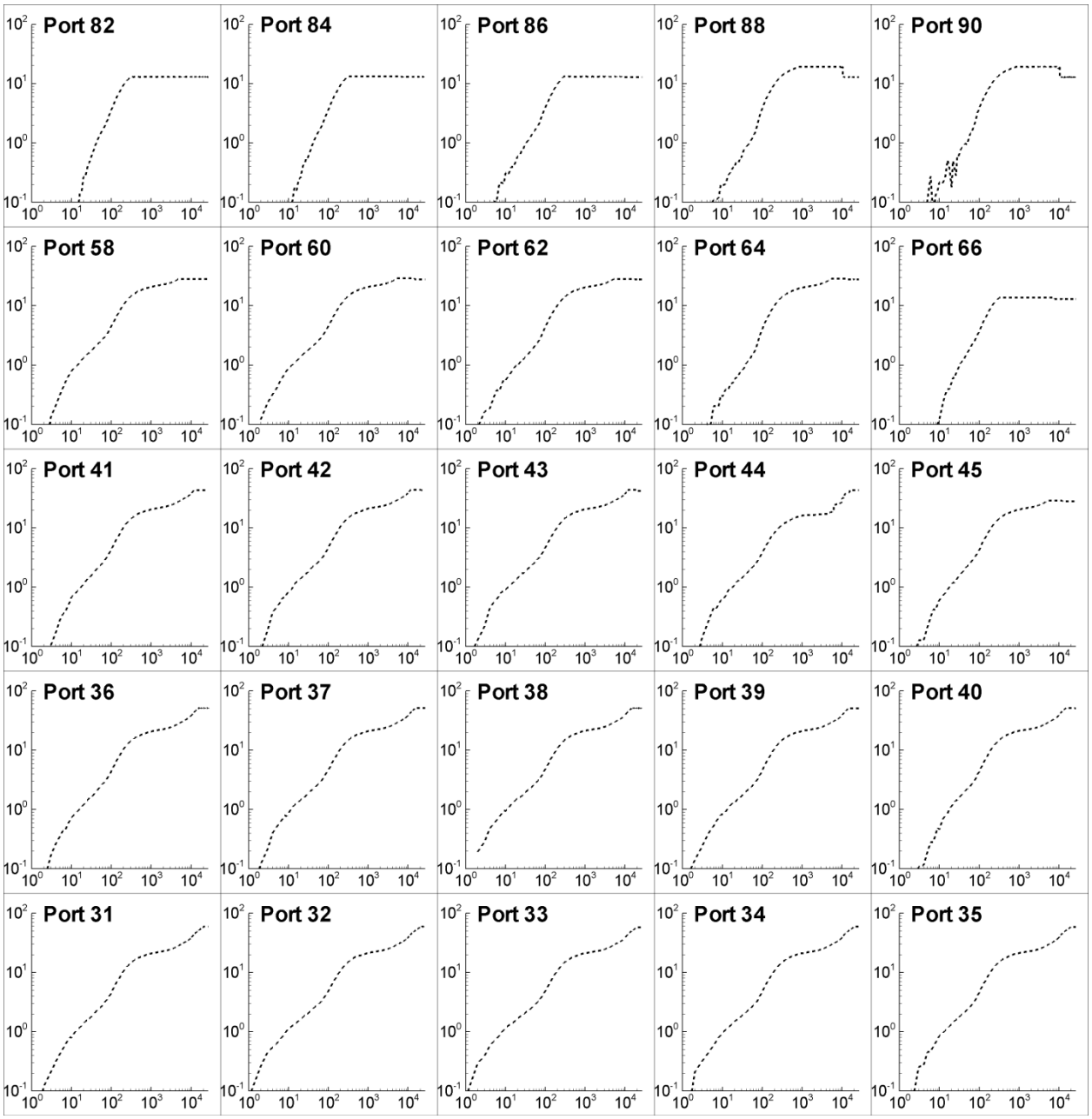


Figure SC1a - Observed pressure transducer response during an unconfined pumping test performed at Port 3. Relative port positions in this figure match the real position in the sandbox. (Top half of sandbox)

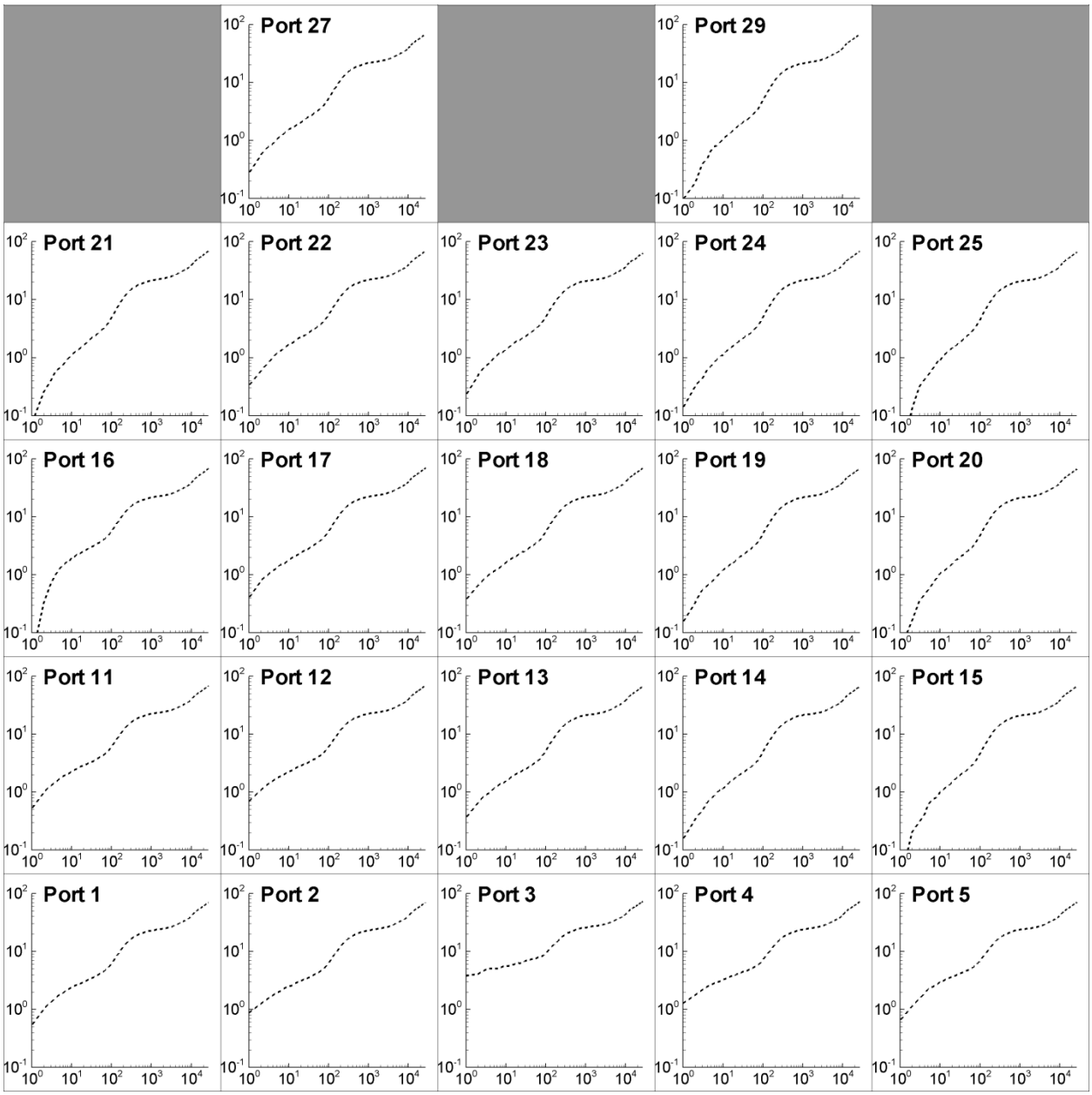


Figure SC1a (continued) - Observed pressure transducer response during an unconfined pumping test performed at Port 3. Relative port positions in this figure match the real position in the sandbox. (bottom half of sandbox)

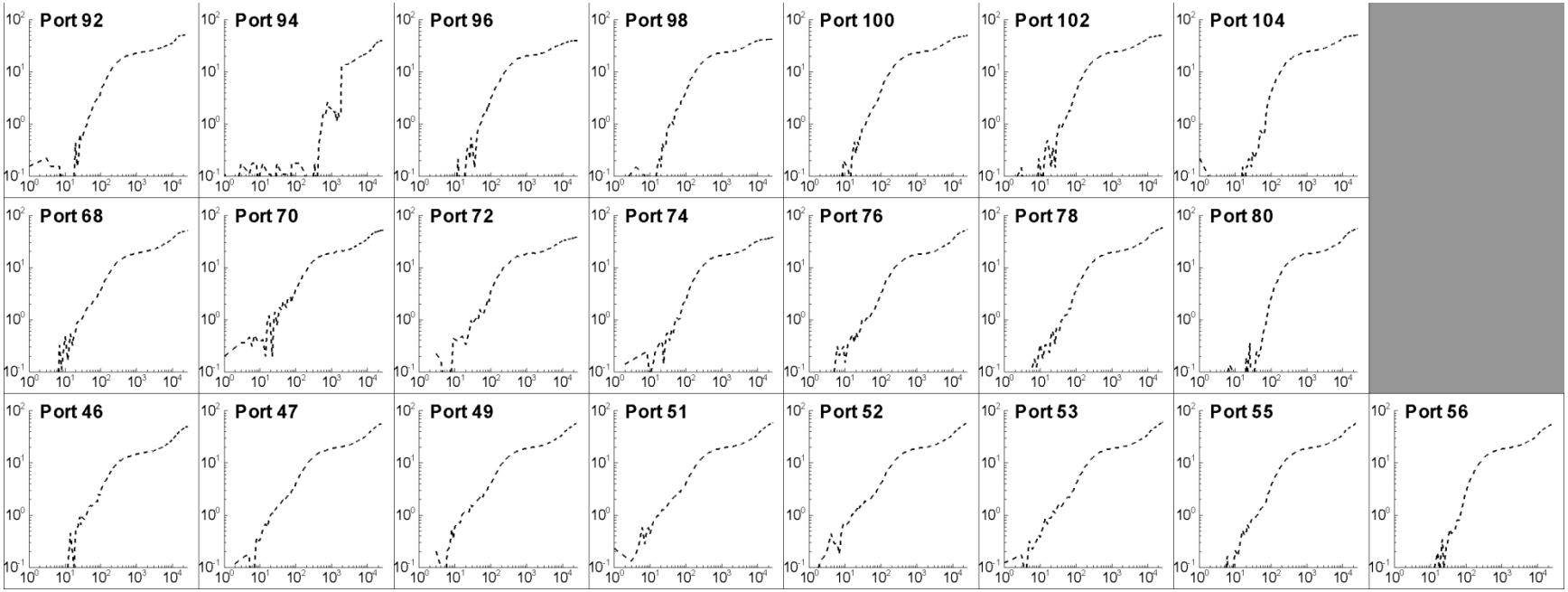


Figure SC1b - Observed tensiometer response during an unconfined pumping test performed at Port 3 compared to the simulated results from Case 1. Relative port positions in this figure match the real position in the sandbox.

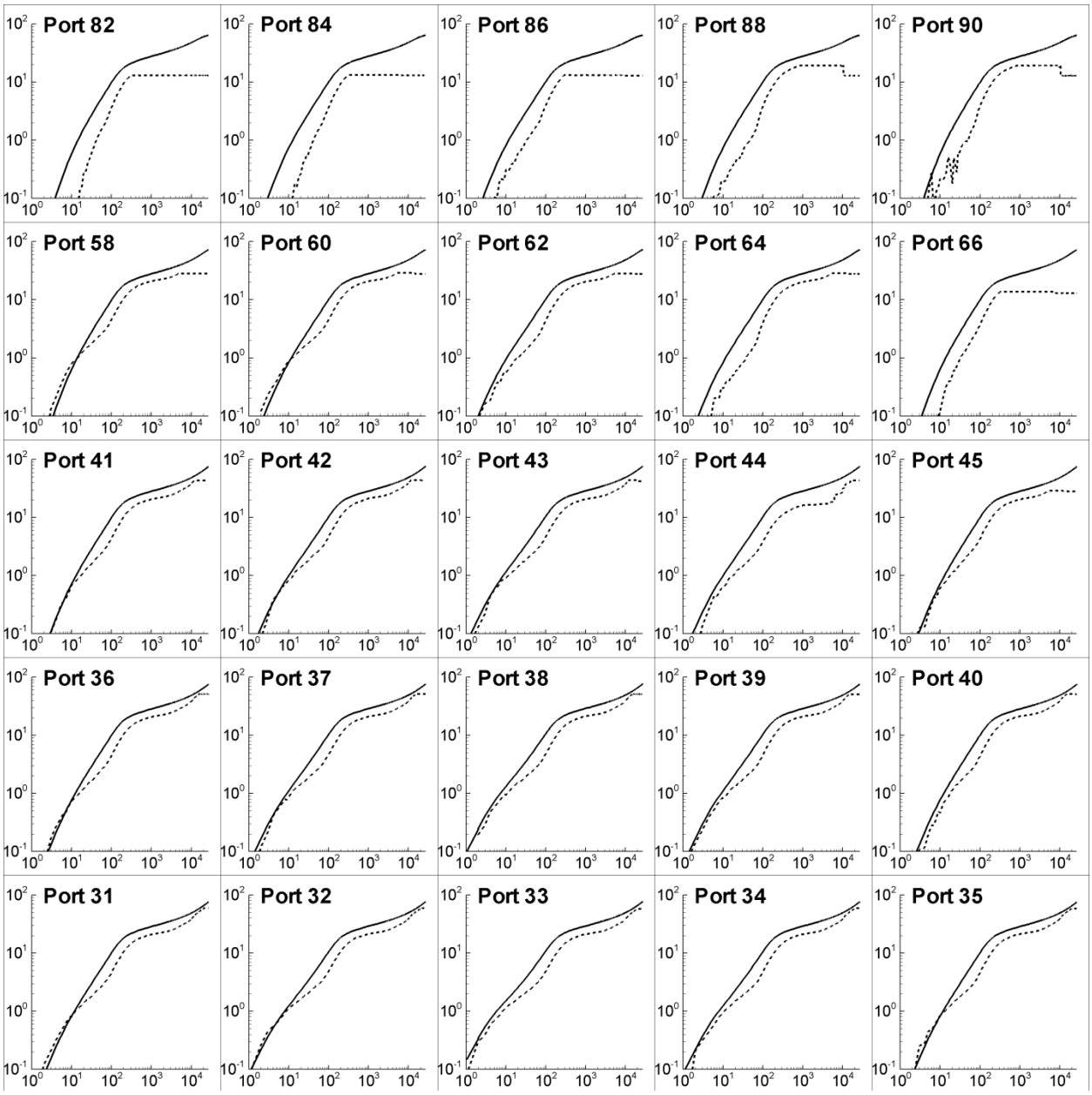


Figure SC2a - Observed pressure transducer response during an unconfined pumping test performed at Port 3 compared to the simulated results from Case 1. Relative port positions in this figure match the real position in the sandbox. (Top half of sandbox)

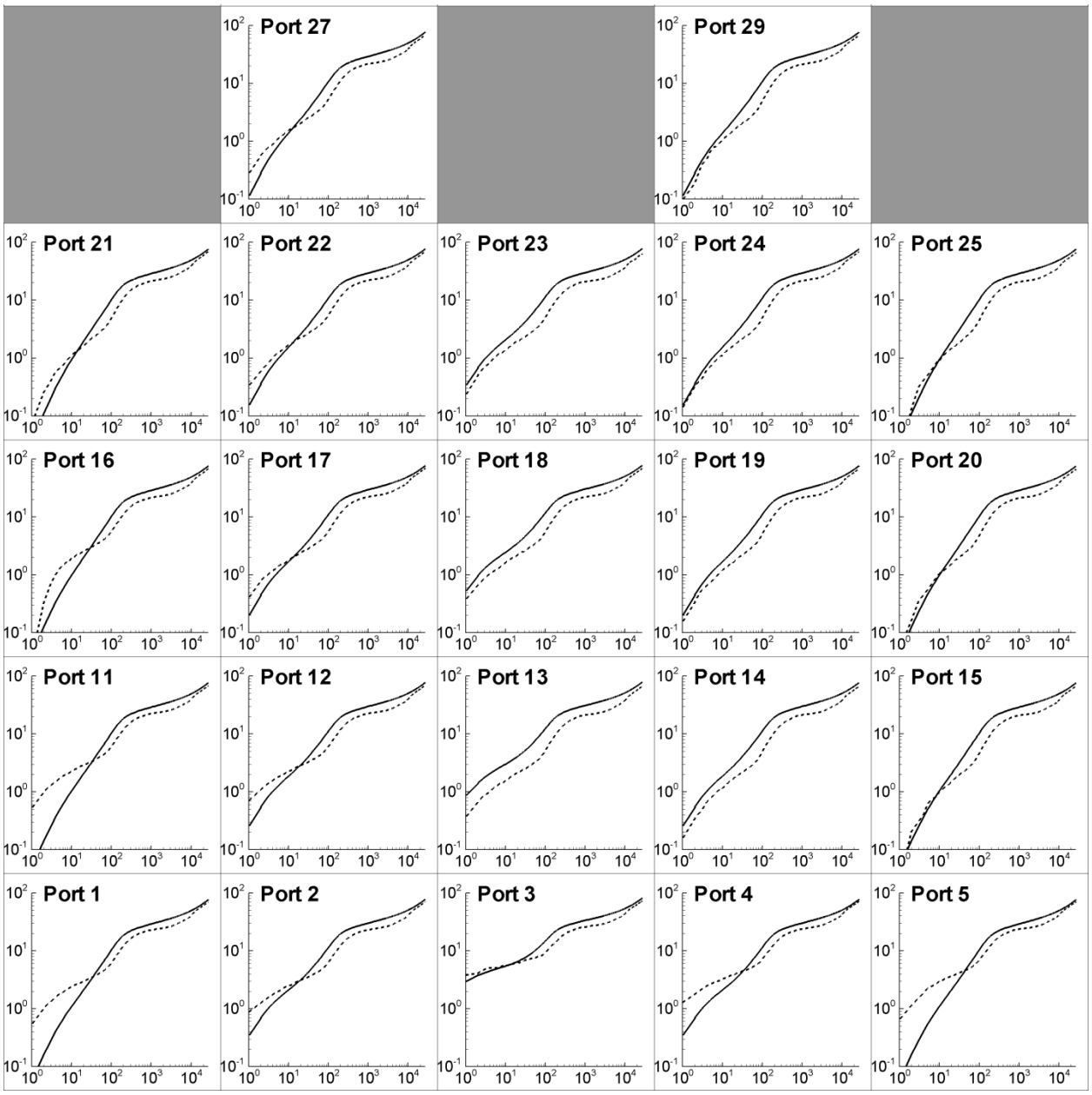


Figure SC2a (continued) - Observed pressure transducer response during an unconfined pumping test performed at Port 3 compared to the simulated results from Case 1. Relative port positions in this figure match the real position in the sandbox. (bottom half of sandbox)

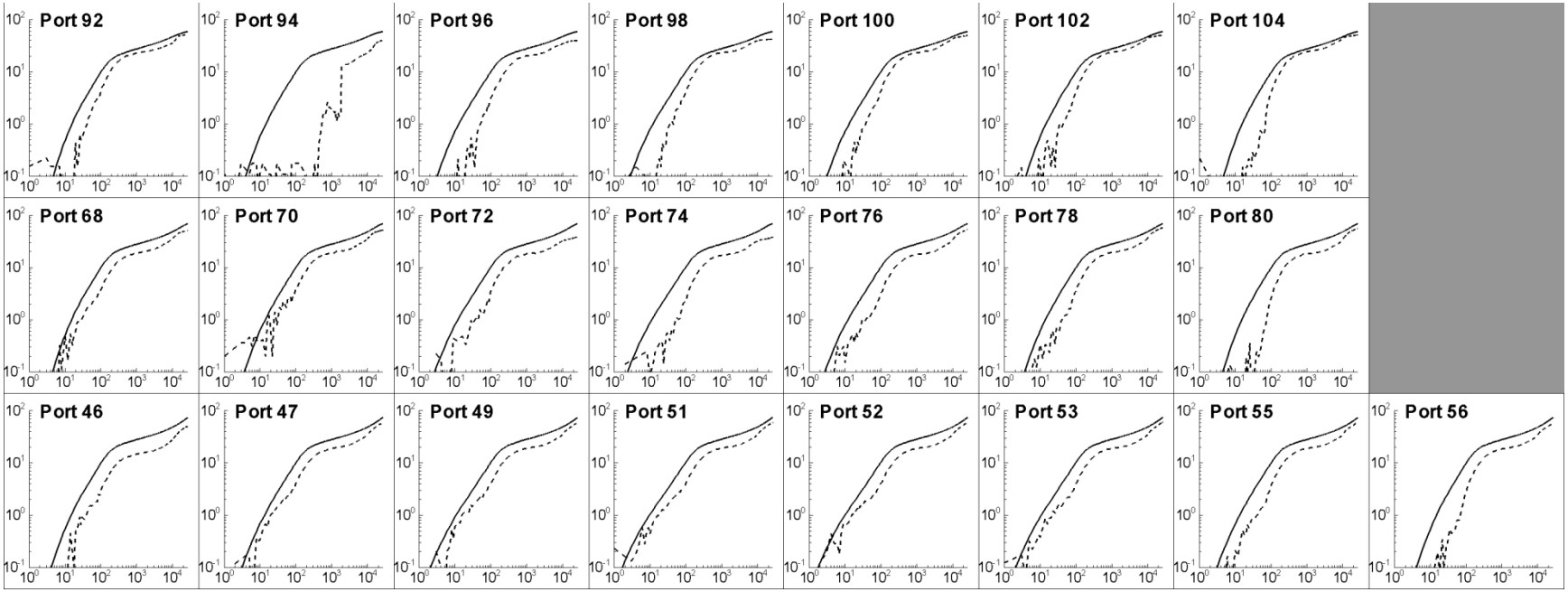


Figure SC2b - Observed tensiometer response during an unconfined pumping test performed at Port 3 compared to the simulated results from Case 1. Relative port positions in this figure match the real position in the sandbox.

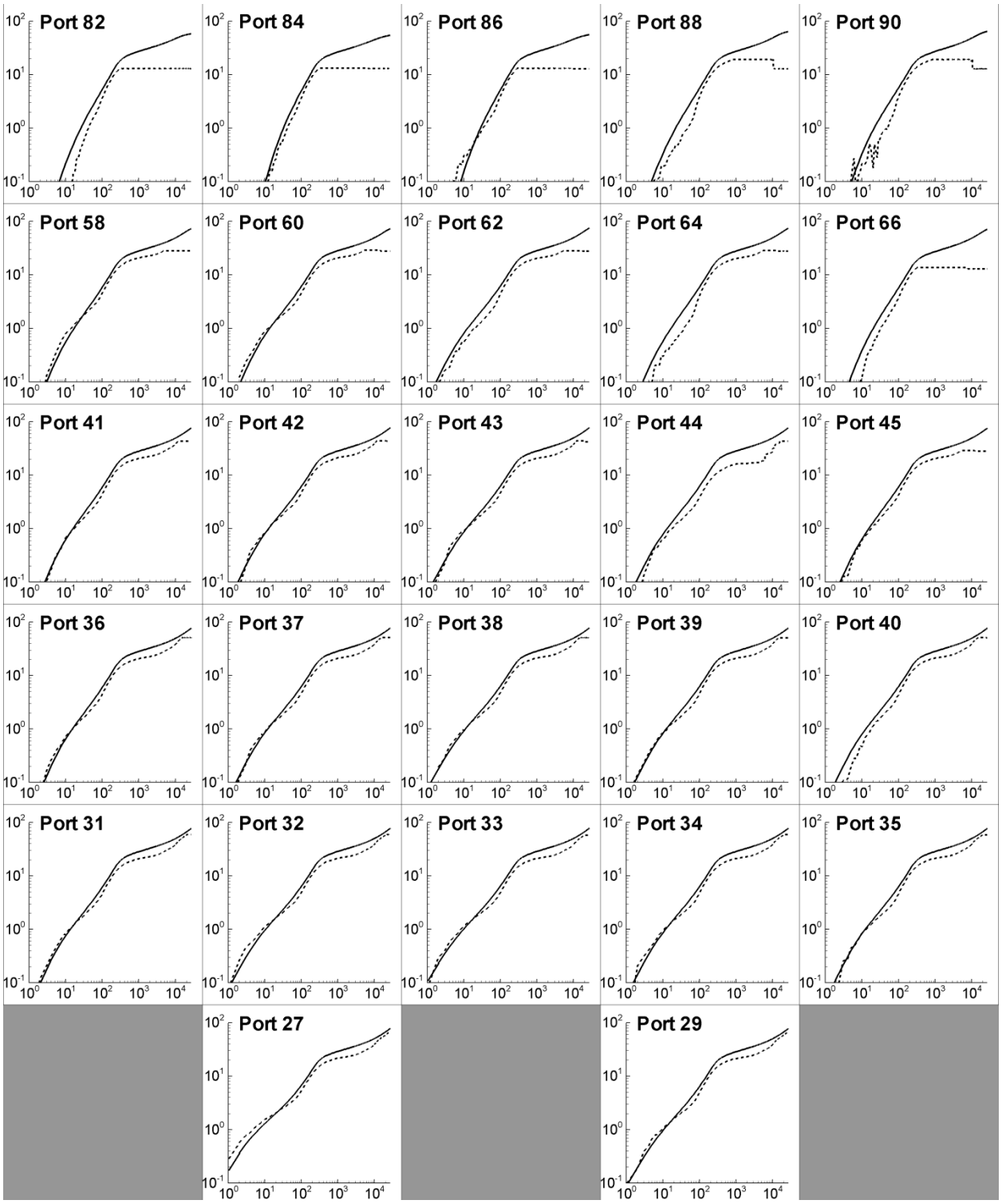


Figure SC3a - Observed pressure transducer response during an unconfined pumping test performed at Port 3 compared to the simulated results from Case 2. Relative port positions in this figure match the real position in the sandbox. (top half of sandbox)

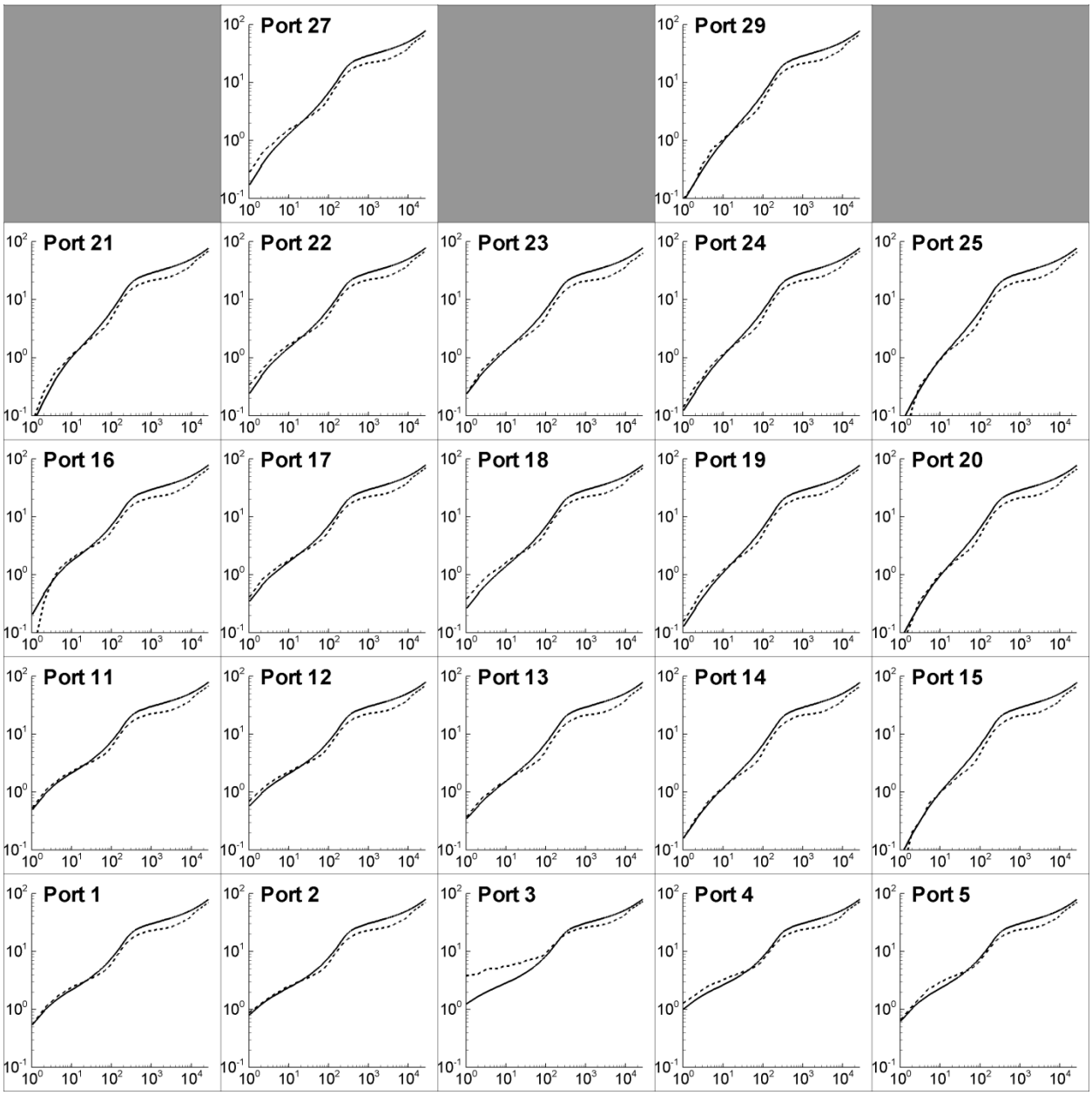


Figure SC3a (continued) - Observed pressure transducer response during an unconfined pumping test performed at Port 3 compared to the simulated results from Case 2. Relative port positions in this figure match the real position in the sandbox. (bottom half of sandbox)

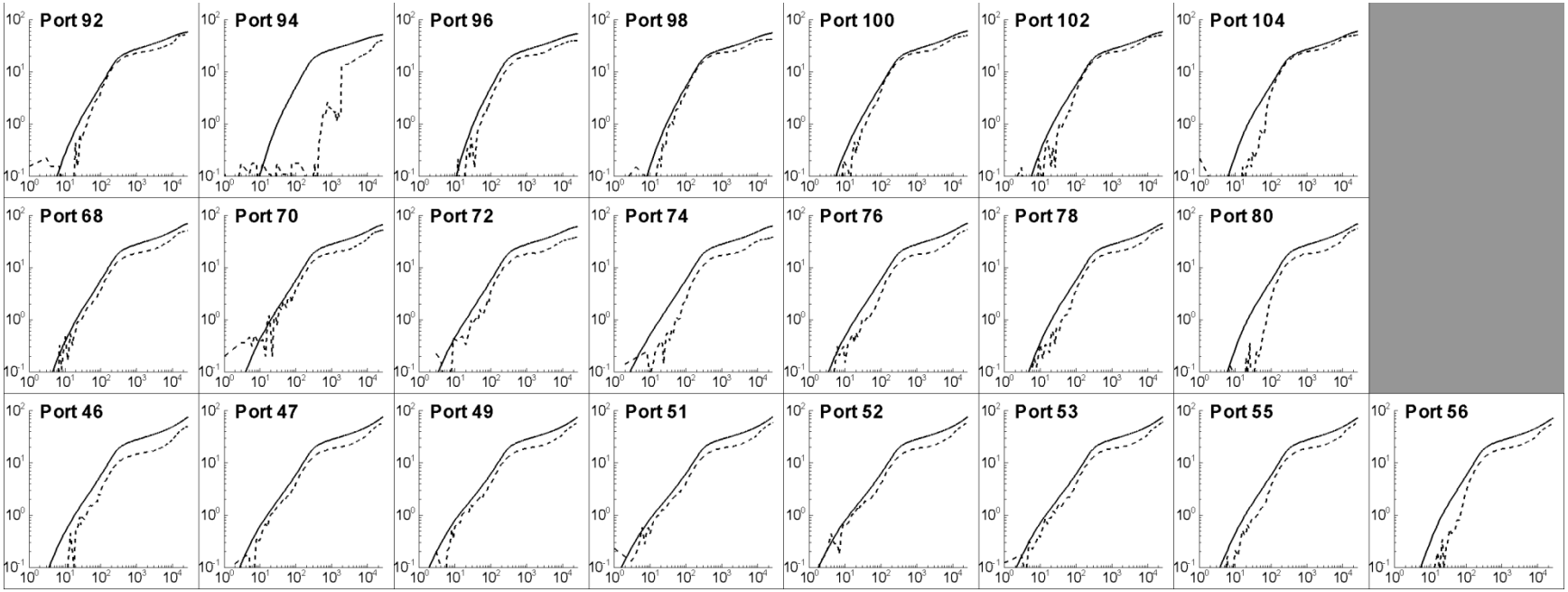


Figure SC3b - Observed tensiometer response during an unconfined pumping test performed at Port 3 compared to the simulated results from Case 2. Relative port positions in this figure match the real position in the sandbox.

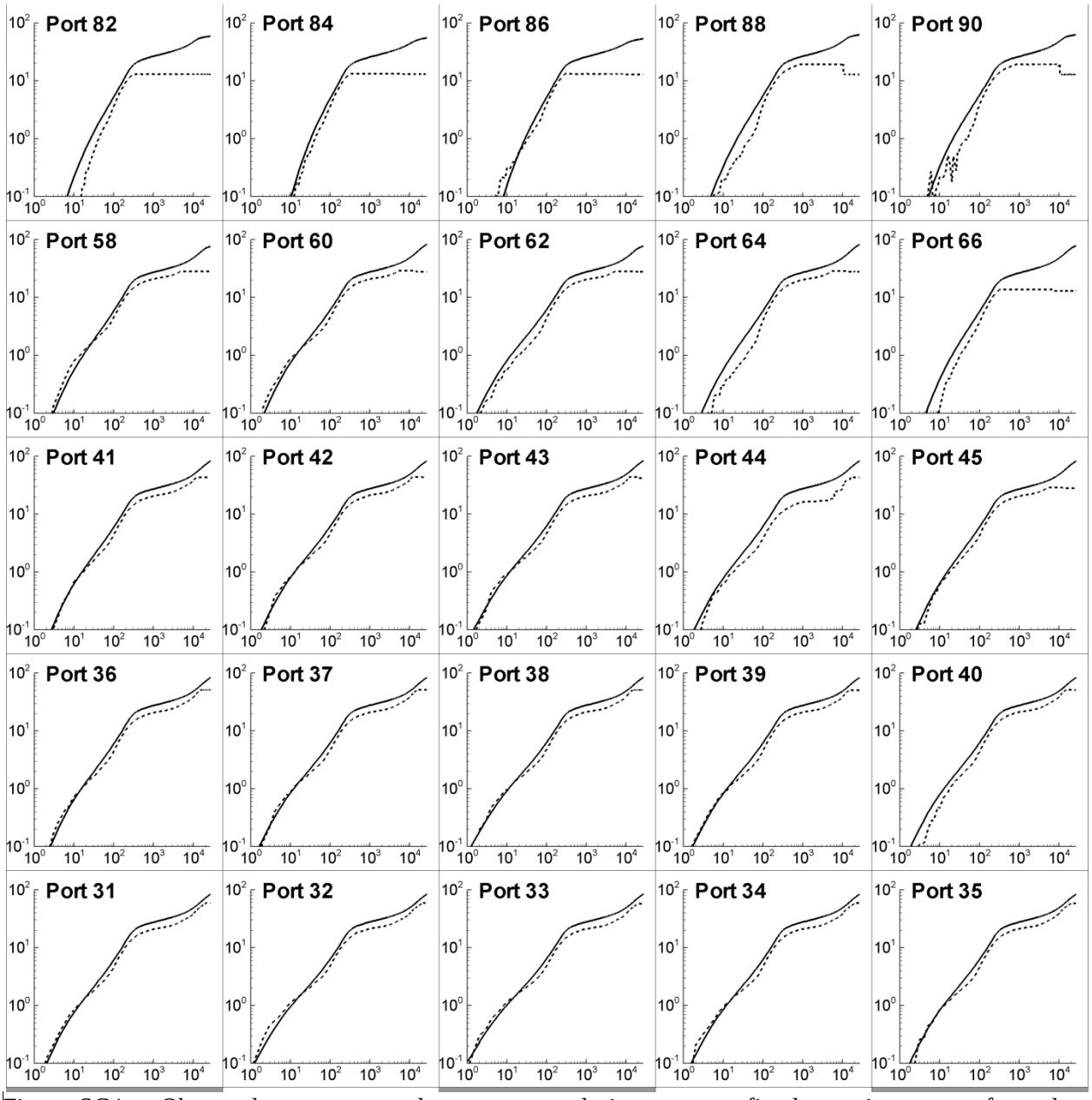


Figure SC4a - Observed pressure transducer response during an unconfined pumping test performed at Port 3 compared to the simulated results from Case 3. Relative port positions in this figure match the real position in the sandbox. (top half of sandbox).

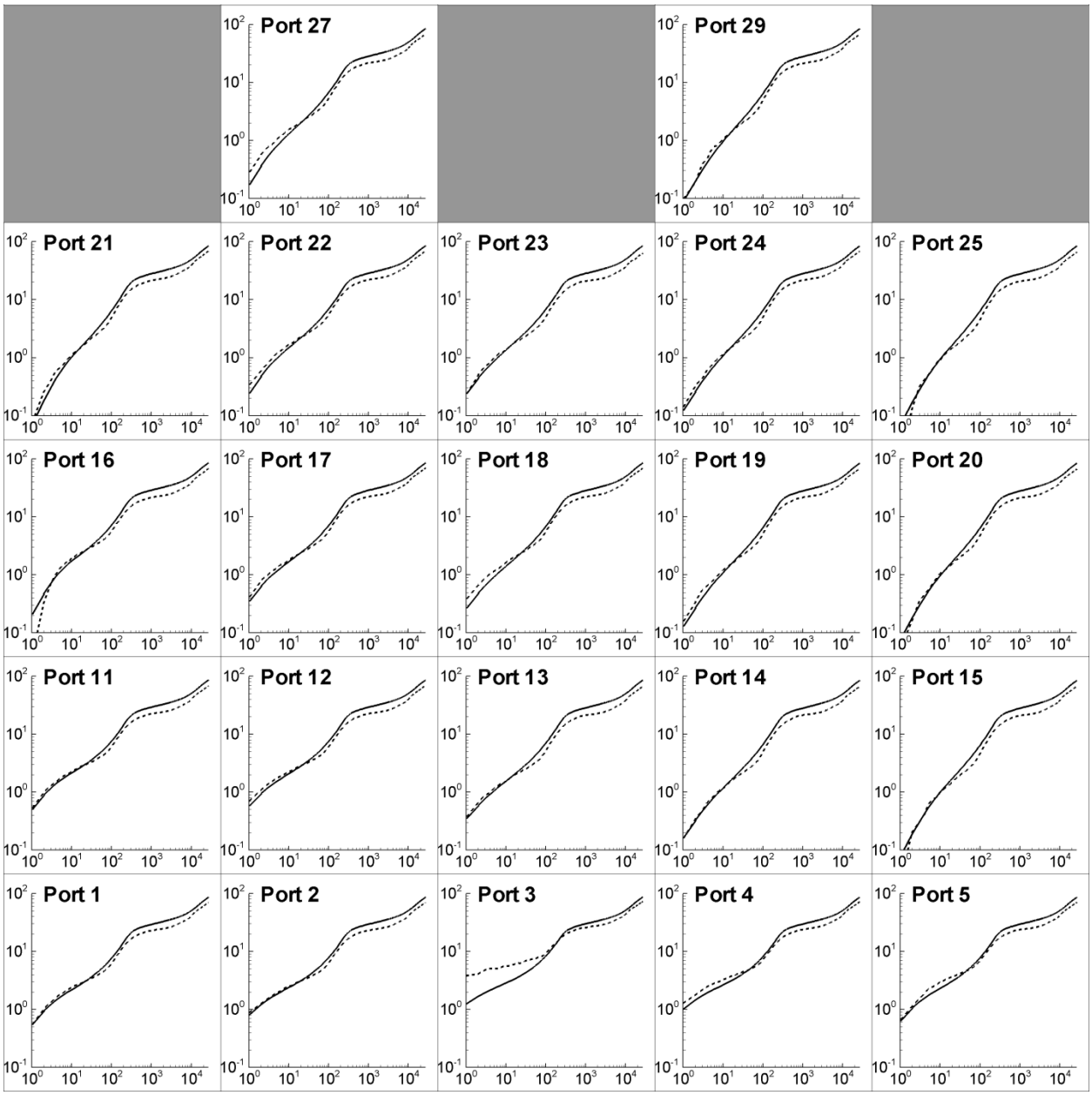


Figure SC4a (continued) - Observed pressure transducer response during an unconfined pumping test performed at Port 3 compared to the simulated results from Case 3. Relative port positions in this figure match the real position in the sandbox. (top half of sandbox).

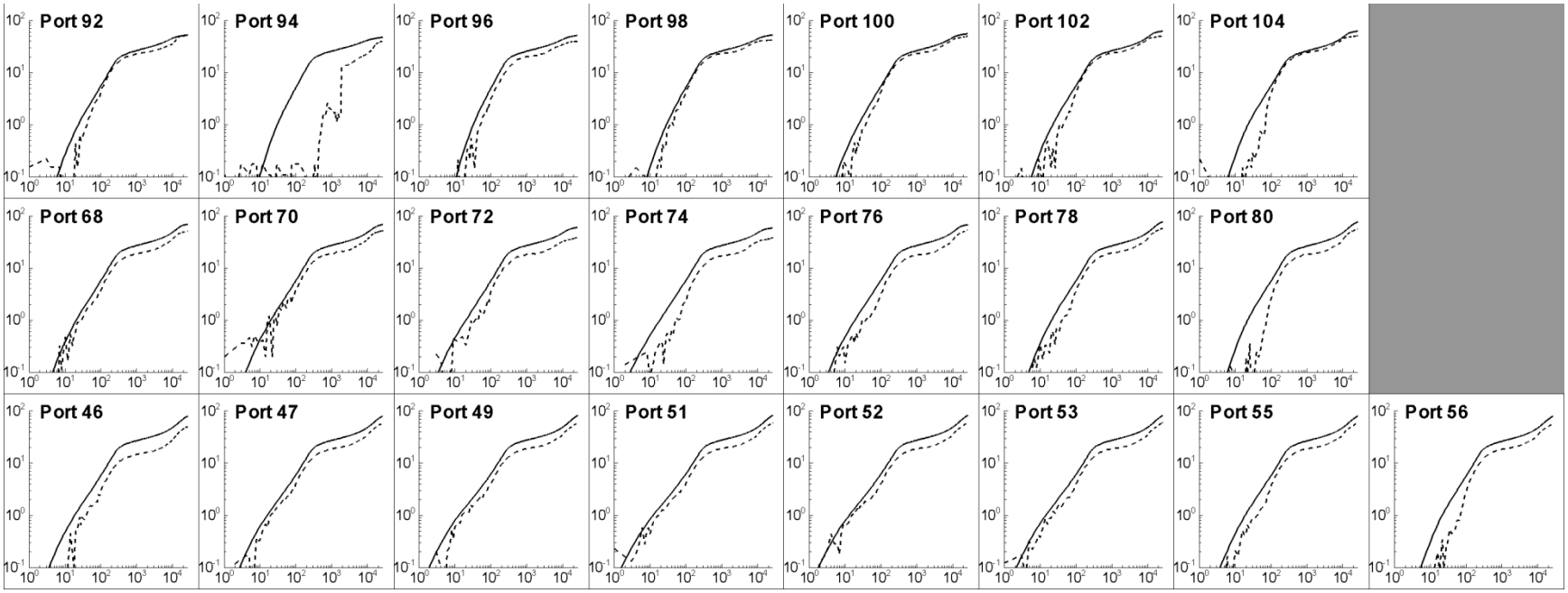


Figure SC4b - Observed tensiometer response during an unconfined pumping test performed at Port 3 compared to the simulated results from Case 3. Relative port positions in this figure match the real position in the sandbox.

Appendix D

Hydraulic/Partitioning Tracer Tomography for DNAPL Source Zone Characterization: Small-Scale Sandbox Experiments

This manuscript by *Illman, W.A., S.J. Berg, X. Liu, and A. Massi* [2010b],was published in the journal *Environmental Science and Technology*. Reproduced with permission from [Illman et al. *Environ. Sci. Tehcnol*., 2010, 44 (22), pp 8609-8614, DOI:10.1021/es101654j] Copyright [2010] American Chemical Society.

EXECUTIVE SUMMARY

Dense nonaqueous phase liquids (DNAPL) are prevalent at a large number of sites throughout the world. The variable release history, unstable flow, and geologic heterogeneity make the spatial distribution of DNAPLs complex. This causes difficulties in site remediation contributing to long-term groundwater contamination for decades to centuries. We present laboratory experiments to demonstrate the efficacy of Sequential Successive Linear Estimator (SSLE) algorithm that images DNAPL source zones. The algorithm relies on the fusion of hydraulic and partitioning tracer tomography (HPTT) to derive the best estimate of the K heterogeneity, DNAPL saturation (S_N) distribution and their uncertainty. The approach is non-destructive and can be applied repeatedly. Results from our laboratory experiments show that S_N distributions compare favorably with DNAPL distributions observed in the sandbox, but not so with local saturation estimates from core samples. We also found that the delineation of K heterogeneity can have a large impact on computed S_N distributions emphasizing the importance of accurate delineation of hydraulic heterogeneity.

D1. INTRODUCTION

Dense Nonaqueous Phase Liquids (DNAPLs) are prevalent at a large number of sites throughout the world. The high densities, low interfacial tensions, and low viscosities of halogenated solvents can lead to deep DNAPL penetration creating persistent source zones [*Pankow and Cherry,* 1996]. The variable release history, unstable flow, and geologic heterogeneity make the distribution of DNAPL in the source zone complex, where DNAPLs exist as residuals or as pools of pure phase. To determine the extent of DNAPL contamination, the most common methods are the use of groundwater monitoring wells and by obtaining cores using split spoon samplers or direct push methods [*Pankow and Cherry,* 1996; *Kram et al.,* 2001]. Other techniques to characterize DNAPL source zones include visual inspections of drill cores, soil vapor analysis, geophysical surveys, use of radon abundance data, and partitioning tracer tests [*Kram et al.,* 2001]. In particular, partitioning tracer tests have been used to provide estimates of residual DNAPL volume [*Jim et al.,* 1995; *Nelson and Brusseau,* 1996]]. The method relies on the retardation of partitioning tracers relative to a conservative tracer caused by the partitioning of tracers with DNAPLs. The resulting breakthrough curves (BTCs) are then typically analyzed using a temporal moment method to compute a spatially integrated estimate of DNAPL saturation (S_N) [*Annable et al.,* 1998].

To our knowledge, there are only few studies that have estimated the spatial distribution of DNAPL S_N through interpretation of partitioning tracer tests [*James et al.,*2000; *Datta-Gupta et al.,*2002]. However, these studies do not incorporate potentially useful hydraulic head data during the test and prior knowledge of hydraulic heterogeneity. Recently, Hydraulic/Partitioning Tracer Tomography (HPTT) was developed to characterize DNAPL source zones [*Yeh and Zhu,* 2007]. HPTT relies on the fusion of data from hydraulic tomography (HT) and partitioning tracer tomography (PTT). HT is similar in concept to the Computerized Axial Tomography (CAT) scan technology, but the energy source is a change (decrease or increase) in hydraulic head achieved

through water pumping or injection at multiple locations, and pressure transducers detect the arrival of head change throughout the source zone. To date, HT has been tested using synthetic simulations [*Yeh and Liu*, 2000; *Zhu and Yeh*, 2005], laboratory sandbox experiments [*Illman et al.*, 2007; *Liu et al.*, 2007; *Illman et al.*, 2010]] and in the field [*Straface et al.*, 2007; *Illman et al.*, 2009].

PTT is similar in concept to HT, but the source is a partitioning tracer that is introduced at some injection point and the signals are the corresponding BTCs obtained at multiple sampling points. Additional partitioning tracer tests are then sequentially conducted by injecting tracers at other locations and the corresponding BTCs are obtained at various locations. Because the tracers sweep through the DNAPL source zone from several directions, PTT yields additional information on S_N distributions that cannot be obtained from a single partitioning tracer test or through its interpretation using a temporal moment method.

The main advantage of conducting HT first and then conducting PTT is that the hydraulic conductivity (K) heterogeneity can first be captured through HT. This then allows one to conduct a tracer test that can be better controlled hydraulically leading to an improved sweep of the tracers through the source zone. In particular, synthetic simulations have shown [*Yeh and Zhu*, 2007] that the accurate knowledge of K distribution within the source zone is critical in obtaining accurate estimates of S_N distributions. In addition, when the tracer tests are conducted in a tomographic manner, more detailed information on DNAPL S_N can be obtained in comparison to a traditional partitioning tracer test.

Our ultimate goal is to validate the HPTT approach in the field to image DNAPL source zones for hydraulic heterogeneity and S_N as well as their uncertainties. However, the approach should be tested in the laboratory first in which the source zone characteristics are well known and experimental conditions can be fully controlled. Therefore, the main goal of this study is to

independently evaluate the performance of the HPTT algorithm (8) in a synthetic aquifer constructed in a laboratory sandbox.

D2. CONSTRUCTION OF HETEROGENEOUS AQUIFER AND ITS HYDRAULIC CHARACTERIZATION

D2.1 Sandbox design and construction

A synthetic DNAPL source zone consisting of trichloroethene (TCE) was constructed in a small sandbox to validate the HPTT algorithm. The sandbox is 91.44 cm in length, 61 cm in height, and has a thickness of 1.9 cm. The height of the sand in the tank is 45.7 cm. The sandbox was constructed with ¾" thick Plexiglas. The sandbox has solid "no-flow" boundaries on the left and right ends and at the bottom. The top boundary is open to the atmosphere in which water was allowed to pond on top of the sand. The sandbox was equipped with 30 water sampling and pressure monitoring ports that could also be used for water injection and extraction (Figure D1a).

D2.2 Porous media used to pack the sandbox

Porous media chosen for packing the sandbox was based on two criteria: 1) to create a simplified heterogeneity pattern with sufficient contrast in K to allow HT to image the various K zones and 2) to create a low K zone in the central part of the source zone to cause infiltrating TCE from the top to pool on this layer. Based on these criteria, we selected three commercially available sands (F-75, #14, and 20/30). The F-75 sand is the finest sand (d_{50} = 0.174 mm) among the three and was carefully packed as a single horizontal lens within a coarse matrix of 20/30 sand (d_{50} = 0.750 mm) (Figure D1a). The #14 sand is a medium grained sand (d_{50} = 0.457 mm) and was packed as short layer above the F-75 layer. The K of each sand (F-75: $K = 1.73 \times 10^{-2}$ cm/s; #14: $K = 1.21 \times 10^{-1}$ cm/s; 20/30: $K = 3.12 \times 10^{-1}$ cm/s) was estimated using a permeameter [*Klute and Dirksen*, 1986]. The synthetic

aquifer has a total volume of 7,985.02 cm^3 and a corresponding void volume of 2,635.06 cm^3 was obtained by a comparing the fully saturated and oven dried sample mass yielding an average porosity (ϕ) of 0.33. In the subsequent analyses of partitioning tracer tests, we treat the ϕ to be a constant as its spatial variability is considerably smaller than that of K and S_N.

D2.3 Hydraulic characterization prior to TCE spill

To characterize the fully water-saturated aquifer in the sandbox, we conducted dipole cross-hole tests prior to spilling the TCE. In the dipole cross-hole test, a mass balance of water injected and extracted was maintained using a peristaltic pump by connecting the injection and extraction ports in a single loop. Nine pairs of ports consisting of tests 1 - 9 (e.g., test 1: injection at port 1 and extraction at port 30, from now on I1/E30; test 2: I2/E29; test 3: I3/E28; test 4: I4/E27; test 5: I5/E26; test 6: I6/E25; test 7: I7/E24; test 8: I12/E19; and test 9: I13/E18) were chosen for these dipole cross-hole tests based on symmetry and coverage of the aquifer features (Figure D1a). Injection and extraction rates for all dipole cross-hole tests averaged 150 mL/min. We calibrated the pressure transducers prior to each dipole cross-hole test and collected head data for several minutes in all pressure transducers to establish a static, initial condition. We then pumped from one port while injecting to another using a peristaltic pump, while taking head measurements in all 30 ports. For each test, pumping continued until the development of steady state conditions, which was determined by observing the stabilization of all head measurements on the data logger connected to a computer.

D3. CREATION OF TCE SOURCE ZONE

A TCE source zone was then created in the synthetic aquifer by injecting 125 mL of TCE (Sigma-Aldrich) dyed with Sudan IV (Sigma-Aldrich) at a rate of 12.5 ml/min over 10 minutes using a syringe attached to a peristaltic pump approximately 5 mm beneath the soil surface in the central part

of the aquifer. After commencing the injection, the TCE quickly infiltrated through the 20/30 sand, moved through the #14 sand layer, and ponded on the F-75 sand as shown on Figure D1b. Based on our observations in the laboratory, most of the TCE appeared to rest on top of the F75 sand, while small amounts of TCE were left as residual phase above the F75 sand where the sand is dyed (Figure D1b). The presence of the #14 sand layer apparently caused some spreading of the TCE resulting in a wider zone of contamination than initially anticipated. The injection of TCE in this sandbox resulted in a sandbox aquifer-averaged S_N of approximately 5% by assuming a porosity of 0.33, which is equivalent to a volumetric TCE content of 0.0165.

D4. CHARACTERIZATION OF TCE SOURCE ZONE

D4.1 Determination of partitioning coefficient through batch tests

In order to determine S_N from partitioning tracer tests, the partitioning coefficient (K_{NW}) of the tracer and TCE is needed. The partitioning tracer selected for this study was 2-methyl-1-butanol (2M1B; Sigma-Aldrich). Here, we estimated the K_{NW} through batch tests (see Supplementary Information (SI) section for details) resulting in a K_{NW} of 3.54, a value which is comparable to those determined previously [*Dugan et al.*, 2003].

D4.2 Partitioning tracer tests

Upon settlement of the TCE in the source area, we conducted 4 partitioning tracer tests. In each tracer test, a dipole flow field was created by injecting tap water at one port at a rate of 500 mL/min and extracting at another port at 450 mL/min. A water mass balance was then achieved by setting 12 sampling lines at a cumulative rate of 50 mL/min. The selected flow rates for these experiments were found to have no impact on the experiments and on the integrity of the porous medium. Table D1 summarizes the injection, extraction, and sampling ports for each tracer test.

Once steady state flow conditions were reached in which the injection, extraction, and sampling rates all stabilized, a valve was switched on the injection line from water to a solution containing 2M1B as the partitioning tracer with a concentration of 500mg/L. Bromide (Fisher Scientific) was used as a conservative tracer with a concentration of 100 mg/L. The tracer concentrations were sufficiently low such that density effects could be ignored in the subsequent analyses of the partitioning tracer tests. The tracer solution was injected for 10 minutes and then the valve was switched back to tap water. Separate sample sets were collected for bromide and 2M1B throughout the duration of the test from the 12 sampling ports. Thirty 2M1B samples and 20 bromide samples were collected from each of the 12 sampling lines during each tracer test for a total of 600 samples. Each tracer test lasted approximately 2 hours.

Each of the tracer tests was performed with different injection/extraction and water sampling schemes to sweep the tracers through the source zone from different directions in a tomographic fashion. During these experiments, nonaqueous phase TCE was observed to be stable.

Water samples taken during each tracer test were separated and chilled in a refrigerator at 4^{o}C until sample analysis. Concentrations of 2M1B were determined using a procedure (see Supplementary Information (SI) section) for the analysis of the batch test samples using the GC-FID while the bromide concentrations were determined using an ion chromatograph (IC) (Dionex ICS 2000) equipped with an AS40 auto sampler.

D4.3 Direct estimates of S_N through coring

As soon as the tracer tests were completed, 6 core sampling tubes were inserted from the top of the sandbox to the bottom to obtain discrete samples for independent estimates of S_N to facilitate comparison with S_N from HPTT. The sampling tubes were extracted by applying a vacuum at the top

to avoid sample loss, then discrete samples were obtained from the cores and analyzed for S_N (see SI for further details).

D5. INVERSE MODELING OF DIPOLE CROSS-HOLE AND PARTITIONING TRACER TESTS

The hydraulic and partitioning tracer tomography analysis of the dipole cross-hole and partitioning tracer tests were conducted using the Sequential Successive Linear Estimator (SSLE) developed for HPTT [*Yeh and Zhu,* 2007]. SSLE is a geostatistical inverse modeling code which rigorously considers the governing flow and transport principles. As it is based on stochastic methods, it can also provide uncertainty estimates which can be crucial to decision making. SSLE first analyzes data from the dipole cross-hole tests to estimate the *K* tomogram and then the dipole partitioning tracer test data sequentially to delineate both the volumetric water content (θ_w) and TCE content (θ_N) tomograms. Note that the porosity is the sum of θ_w and θ_N. Details to the algorithm are provided in the original publication [*Yeh and Zhu*,2007], while we provide a brief description of the algorithm in the SI section.

D5.1 Inverse modeling of dipole cross-hole tests

To obtain a *K* tomogram from the available dipole cross-hole tests, we solve a 2D inverse problem for steady-state flow conditions. The synthetic aquifer was discretized into 612 elements and 1332 nodes with element dimensions averaging 2.54 cm × 1.91 cm × 2.69 cm. The top boundary was set to be a constant head boundary, while the other 3 sides were set as no flow boundaries. Therefore, despite the small size of the sandbox, we emphasize that boundary effects were properly accounted for in the forward and inverse modeling effort.

Input data to the inverse model include initial guesses for the effective hydraulic conductivity (K_{eff} = 0.3 cm/s), variance ($\sigma^2_{\ln K}$ = 0.1) and the correlation scales (λ_x = 40.0 cm, λ_y = 1.91 cm and λ_z = 5.0 cm), and injection/extraction rates (Q) during each dipole cross-hole test. We selected 9 dipole cross-hole tests and the corresponding steady-state head observations at the rest of 28 ports during each test as data sets. We elected to not use the head data from the injection and extraction ports from each test because those ports could be affected by skin effects (11).

D5.2 Inverse modeling of dipole partitioning tracer tests

After the computation of the K tomogram, we used the HPTT code to compute the θ_N tomogram within the laboratory aquifer. The θ_N tomogram was converted to a TCE saturation tomogram by dividing the θ_N tomogram by the porosity (0.33).

The model domain used for the inverse modeling of the partitioning tracers was identical to that for the HT analysis. Boundary conditions were also identical with the exception that the upper boundary was treated as a no flow boundary as opposed to a constant head boundary. The selection of no flow boundaries for all sides ensured that no mass was lost across a boundary during the simulations.

Input data to the inverse model for PTT include the K distribution, dispersivity (set to 0 uniformly within the model domain), the partitioning coefficient (K_{NW} = 3.54), variance (for TCE content) ($\sigma^2_{\ln K}$ = 0.01) and the correlation scales for TCE content (λ_x = 20.0 cm, λ_y = 1.91 cm and λ_z = 1.0 cm), injection/extraction, water sampling rates (Q) during each dipole partitioning tracer test, as well as concentrations of the partitioning tracer. Details to the selection of the concentration data are provided in the SI section.

D6. RESULTS

D6.1 Computation of *K* tomogram

Figure D2 is the *K* tomogram obtained by inverting the steady state head data from 9 dipole cross-hole pumping tests. The *K* values range between 5.3×10^{-1} and 9.0×10^{-3} cm/s, has a geometric mean of 2.1×10^{-1} cm/s cm/s and a $\sigma^2_{\ln K}$ of 0.91. For comparison purposes, we include the outline of the two blocks embedded within the 20/30 sand. Comparison of the *K* values from the tomogram to those from the permeameter analysis of core samples reported earlier shows that the values are quite comparable. Given the small difference between the *K* values of #14 (1.21×10^{-1} cm/s) and 20/30 sands (3.12×10^{-1} cm/s), it is not surprising that the block consisting of #14 sands is not detected. In contrast, the block consisting of F75 appears quite clearly due to the larger contrast in the *K* values between the 20/30 (3.12×10^{-1} cm/s) and F75 sands (1.73×10^{-2} cm/s), although the block appear smoother and larger than the outline indicated on the tomogram. This is a direct consequence of having only 30 ports available for pressure measurements. A previous study [*Straface et al.*, 2007] has shown that as the number of pressure measurement points increase, the resolution of the *K* heterogeneity increases.

D6.2 Computation of S_N tomograms using the *K* tomogram

We next computed the S_N tomogram using the *K* tomogram as the *K* field for the tracer inversion. Figure D3a - f show results from various cases in inverting the partitioning tracer test data. In particular, Figures D3a – d shows inversions of individual partitioning tracer tests 1 through 4, respectively. These results reveal that with a single partitioning tracer test, the existence of TCE in the source area is confirmed although the distribution is not very realistic compared to the photograph of the TCE distribution. This suggests that perhaps one tracer test with a limited number of sampling

ports (Test 1: Ports 2, 3, 6, 13, 25, 26, 27, 28; test 2 - Ports 4, 5, 12, 13, 18, 24, 27, 28, 29; test 3 - Ports 3, 4, 5, 6, 12, 24, 25, 26, 27; and test 4 - Ports 1, 5, 25, 26, 27, 28, 29) may not yield accurate S_N distributions. Interestingly, each S_N tomogram shows a different S_N distribution as well as mean S_N (Table D2) suggesting that the inversion of a single partitioning tracer test can yield a different mean S_N value. For example, the inversion of test 1 alone yields a mean S_N of 3.7% while test 4 yields 1.9%. Recalling that 125 ml of TCE was injected into the sandbox, the mean S_N is 5% suggesting that the mean estimate is close but there is definitely some error which could affect remediation designs. We attribute the differences in the mean S_N to the fact that each tracer test samples a different part of the aquifer.

We next sequentially inverted tests 1 through 4 in that order without the loop iteration scheme and the corresponding S_N tomogram is shown in Figure D3e. This figure shows a region of elevated S_N above the F75 block suggesting that after including 4 tests, HPTT is able to locate a higher S_N region in the central part of the aquifer, despite the exact position not being correct as it is above the F75 lens. We do note that the mean S_N is approximately 50% less of the true value (5%) at 2.2%.

The sequential interpretation of HT or PTT tests in the SSLE algorithm may not fully exploit the data as they are inverted [*Zhu and Yeh*, 2005]. A loop iteration scheme described in the SI section, was also applied to examine whether an improvement in S_N estimates could be obtained. Figure D3f is a S_N tomogram with a mean S_N of 4.7% that is very close to the true mean S_N of 5.0%. We also note that the S_N tomogram looks more realistic with higher S_N in the expected area.

D7. DISCUSSION

Despite the encouraging results, it is natural for one to question whether the S_N distribution can be further improved if we had a better estimate of K heterogeneity. A better estimate of K heterogeneity can be obtained with a larger number of measurement points during HT [*Yeh and Liu*, 2000]. As we

do not have additional sampling ports, we instead deterministically assign K values to the 3 sand types based on K measurements from the permeameter tests and then use this K field for the inversion of the partitioning tracer tests. The K tomogram that we computed earlier can be considered to be an unbiased conditional mean field and because it is based on geostatistical methods, the variability in the heterogeneity is relatively smooth with a limited number of pressure measurement points. Assigning K values as we do here is tantamount to knowing precisely the location of the stratification and its morphology as well as the locations of sharp boundaries between zones of different K values.

Figures D4a through D4f are the estimated S_N tomograms for the various cases that we examined before. As in Figure D3a-d, the inversion of individual partitioning tracer tests shows that the mean S_N estimates are not that much better than the case when the K tomogram was used. However, as Figures D4e and D4f show the position of the high S_N zone is considerably better constrained when the positions of the strata are known. This suggests that improved knowledge of K heterogeneity and in particular, precise knowledge of boundaries between strata can be critical in imaging S_N in heterogeneous aquifers. Undoubtedly, such accurate knowledge of K heterogeneity is difficult to obtain. However, our laboratory and field scale research on hydraulic tomography (e.g., *Illman et al.*, 2010; *Illman et al.*, 2009; *Berg and Illman*, 2009] suggests that such information can be obtained and improvements to the details to heterogeneity may be obtained by including: 1) more monitoring points; 2) conditioning data; and other types of information such as geological and geophysical data into the SSLE algorithm.

The visual comparison of Figure D1b and the S_N tomograms reveals that HPTT is promising in locating areas of higher S_N. Obviously, a better method to quantify S_N is needed. For example, researchers have used X-ray [*Moreno-Barbero and Illangasekare*, 2006] to image DNAPL S_N in their flow cells with high effectiveness. But to our knowledge, X-ray devices for obtaining S_N estimates in the field do not exist at this time, therefore, the development of HPTT technology is essential.

Aside from the visual comparison, a direct comparison of S_N estimates to those from HPTT is more desirable. To achieve this, we next compare the S_N tomograms to those obtained from the interpolation of TCE S_N data that we have obtained through coring. In particular, Figure D5 is the kriged S_N distribution showing regions of elevated S_N above and within the upper portion of the F75 layer. A grid by 1 cm by 1cm was used for kriging. A slightly higher S_N region is also visible to the left side of the F75 block that corresponds with the contaminated area evident in the photograph (Figure D1b).

There are two important features to note in Figure D5. One is that the S_N is considerably lower than those estimated by the HPTT technique. The mean S_N estimated through interpolation is 0.088%, which is significantly lower than the actual mean S_N of 5% or those estimated by HPTT. This suggests that direct sampling through coring could potentially contribute to S_N underestimation. Another observation is that the higher S_N zone is partially located within the F75 block. This is highly unlikely as seen also in Figure D1b that nonaqueous phase of TCE cannot penetrate the F75 sands due to the higher displacement pressure to overcome in order for the nonaqueous phase TCE to penetrate into the F75 sand. One likely explanation for observing nonaqueous phase TCE within the block is that the driving of the coring tube may have consolidated the sands causing the location of the high S_N zone to be pushed deeper. This observation highlights the problem with direct sampling in the source zone - that drilling can contribute to mobilization of DNAPLs to greater depths exacerbating the contamination. In addition, coring of unconsolidated sediments can result in sample loss which can additionally contribute to inaccuracies in S_N estimates.

This study showed the effectiveness of HPTT technology to image K heterogeneity and DNAPL S_N in a 2D sandbox. While its field application is our ultimate goal, a sandbox study was necessary to demonstrate the technology in a controlled environment. We note that the HPTT

algorithm in its current form [*Yeh and Liu*, 2000] can handle 3D cases and field studies on hydraulic tomography are currently underway at the University of Waterloo [*Berg and Illman*, 2009].

Table D1: Injection, extraction, and water sampling locations during tracer tests 1 - 4.

Partitioning tracer test	Injection/extration ports	Sampling port
1	30/1	2, 3, 4, 5, 6, 7, 13, 19, 25, 26, 27, 28
2	26/6	1, 2, 3, 4, 5, 12, 13, 18, 24, 27, 28, 29
3	1/30	3, 4, 5, 6, 12, 18, 24, 25, 26, 27, 28, 29
4	18/13	1, 2, 3, 4, 5, 7, 19, 25, 26, 27, 28, 29

Table D2: Mean S_N and volume estimated through the inversion of various partitioning tracer tests using the *K* tomogram from steady state hydraulic tomography and a known *K* distribution.

	K tomogram from SSHT		Known *K* distribution	
Test	Mean TCE saturation	V_{TCE} (mL)	Mean TCE saturation	V_{TCE} (mL)
1	0.0372	98.10	0.0277	71.99
2	0.0496	128.55	0.0476	126.09
3	0.0317	81.39	0.0309	81.09
4	0.0189	46.53	0.0202	51.48
1-4	0.0218	55.46	0.0491	128.73
1-4 (L)	0.0473	130.06	0.0593	156.94
4-1	0.0360	97.09	0.0521	139.60
4-1 (L)	0.0475	128.12	0.0699	189.58

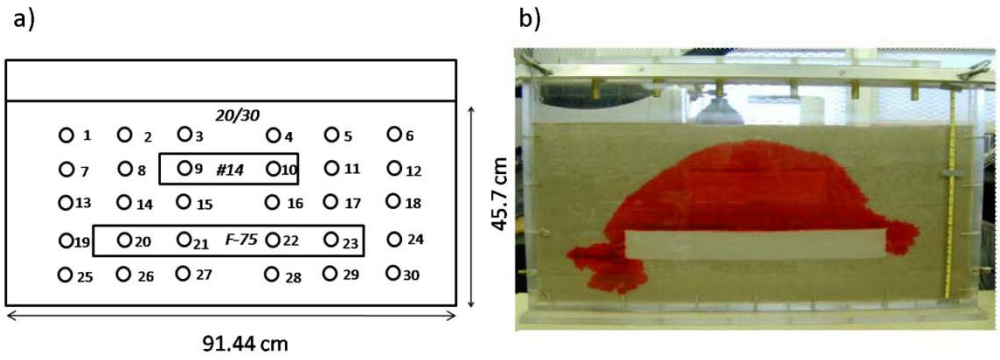


Figure D1: a) Schematic diagram of the sandbox showing sand types, water sampling, and pressure monitoring ports and b) photograph of the TCE source zone after TCE migration has stopped.

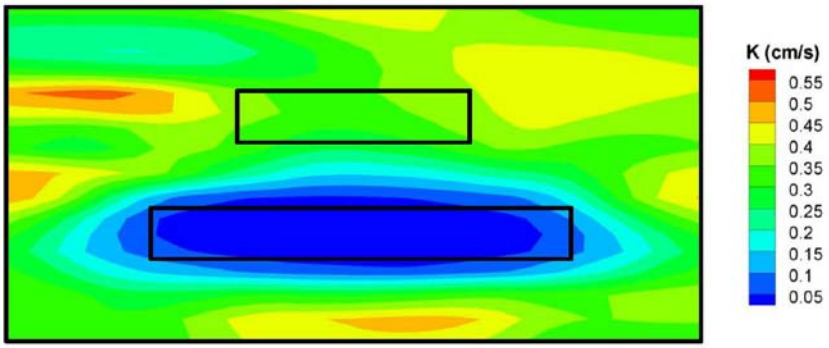


Figure D2: *K* tomogram generated using drawdown data from 9 dipole tests.

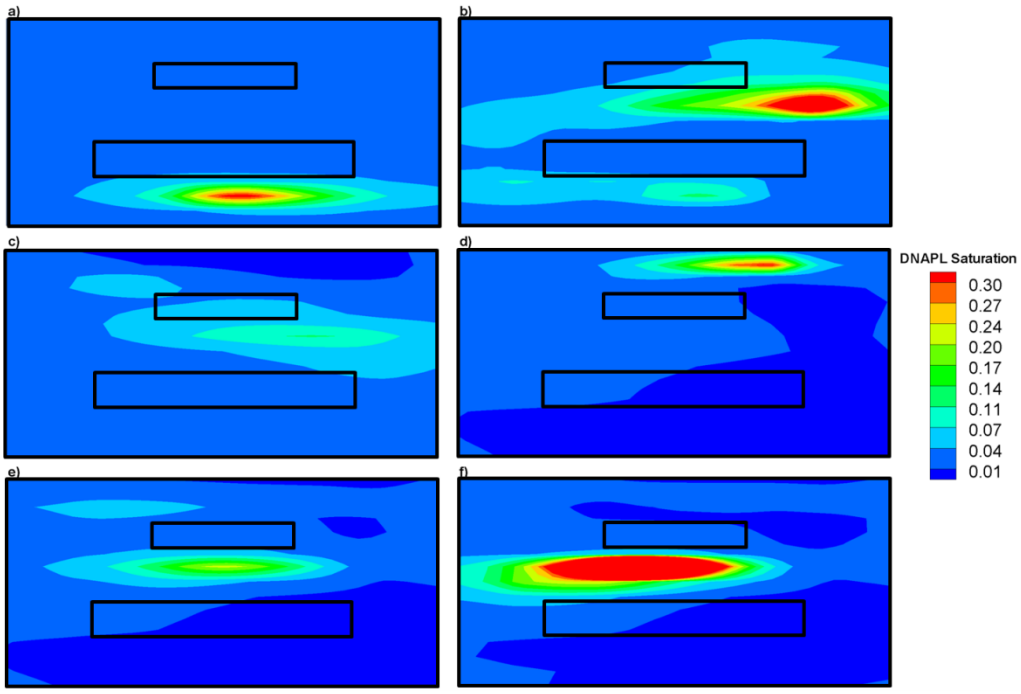


Figure D3: Estimated S_N tomograms with inversion of: a) test 1 only; b) test 2 only; c) test 3 only; d) test 4 only; e) tests 1 - 4; and f) tests 1 - 4 with loop iteration. In this case, the *K* tomogram estimated using steady state hydraulic tomography (Figure 3) was used for partitioning tracer tomography.

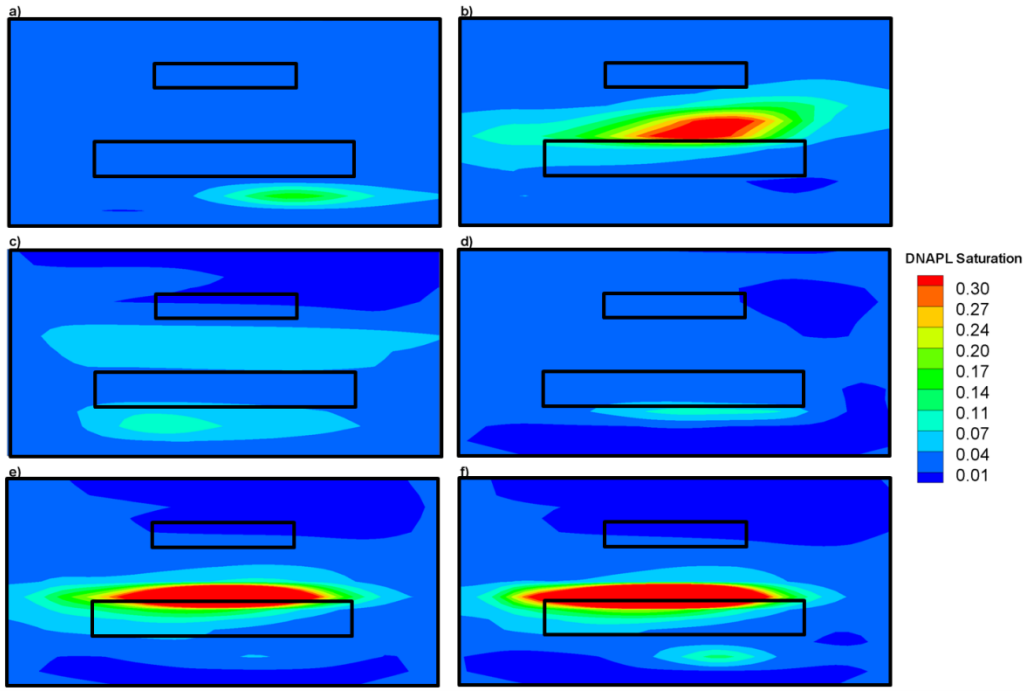


Figure D4: Estimated S_N tomograms with inversion of: a) test 1 only; b) test 2 only; c) test 3 only; d) test 4 only; e) tests 1 - 4; and f) tests 1 - 4 with loop iteration. In this case, the *K* values of the 3 sand types were obtained from permeameter measurements and deterministically assigned for PTT.

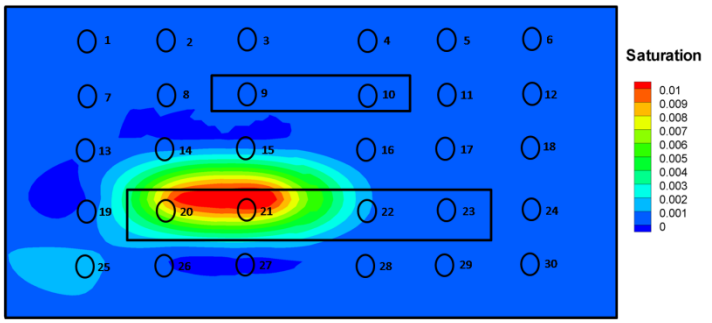


Figure D5: Kriged S_N based on the kriging of 30 core samples. Solid circles show the approximate centroid location of the core samples.

D. SUPPLEMENTARY INFORMATION

Batch tests to determine the partitioning coefficient

Six partitioning-tracer stock solutions of varying concentrations were prepared volumetrically with Hamilton gas-tight syringes in 250-mL volumetric flasks. Purified water (Barnstead Nanopure) was used to dilute the aqueous tracer stock solutions to the six different concentrations: 50, 100, 200, 400, 800, and 1600 mg/L of 2M1B. The 6 different levels were prepared in pre-cleaned 20mL EPA/ VOA glass vials with Teflon-lined septa (EnviroTech). Each sample set consisted of triplicate vials containing TCE and the aqueous 2M1B tracer solution and a control group of triplicate vials containing aqueous tracer solution only. Twenty mL of aqueous tracer solution was dispensed into each vial using a 10 mL pipette followed by 2.5 mL injection of TCE with a gas-tight syringe. The vials were immediately capped to avoid volatilization to the atmosphere.

The 25mL vials were allowed to equilibrate on a platform shaker (Innova 2000) at 150 rpm and room temperature for 24 hours, sufficient to obtain equilibrium. Aliquots with an aqueous phase volume of 1.5 mL were withdrawn from the 25mL vials with minimal disturbance using a gas-tight syringe and transferred to 1.5 mL glass GC auto sampler vials with Teflon septa and analyzed for 2M1B concentrations using a gas chromatograph (Shimadzu GC 2010) with an auto sampler (AOC-20i auto injector) equipped with a flame ionization detector (FID). The GC-FID method consisted of a Restek Stabilwax® capillary column (30 m, 0.32 mm ID, with film thickness of 1µm), helium as the carrier gas, zero-grade air, injection volume of 1 µL, and injector and FID temperatures of 200°C. Also, a split ratio of 35:1 to vent excess water (steam) in the sample and a linear velocity of 50 cm/s were incorporated into the method. Four external calibration standards were prepared for 2M1B in deionized water at levels of 1, 10, 100, and 1000 mg/L. The calibration curve ($R^2 > 0.99$) was created by injecting triplicates of water samples containing different concentrations of 2M1B. The batch test

results were plotted on an arithmetic plot and the partitioning coefficient (K_{NW}) for TCE-2M1B was estimated to be 3.54.

Details to core sample preparation and analysis for TCE saturation

Six brass tubes were pushed vertically downward slightly offset from each of the six columns of sampling ports under fully saturated conditions until they reached the sandbox bottom. Then, bentonite was added to the top of the core to eliminate the empty space and the core tubes were plugged at the end with epoxy putty. Once the epoxy putty was dried, the cores were then removed individually and cut into segments with a pipe cutter. Each of the segments was capped on each end with a silicone plug and cooled in the refrigerator at 4°C until the samples were ready for sample analysis.

The core samples were collected from the core tubing segments and deposited into pre-labeled, pre-weighed 40 mL vials and capped with Teflon-lined septa caps. After this step, the vials were then weighed again and the mass of the vial containing the cap and sample was recorded. The vials were then filled to the top with methylene chloride (Sigma Aldrich) and weighed again. We then recorded the mass of the vial, along with the cap, sample and the methylene chloride. The volume of methylene chloride was calculated based on the mass differences and density of methylene chloride. The vials were placed on an Innova 2000 orbital shaker at 150 rpm for 24 hours. After shaking, the vials were weighed again and the volume of methylene chloride was corrected if any mass loss was observed. Approximately 5 mL of the solution was extracted from each 40 mL vial and transferred to three 1.5 mL vials for GC analysis.

The GC samples were then analyzed via GC-FID with a 7-level triplicate calibration method consisting of standards of TCE in methylene chloride at levels of 1, 10, 100, 500, 5,000, and 10,000 mg/L. The sample TCE concentrations were recorded and the average and standard deviations from

each set of 3 samples were calculated. The mass of TCE in each core sample was calculated by multiplying the average sample by volume of liquid in the sample (assumed to be the volume of methylene chloride). The mass was then converted to volume by dividing by the density of TCE. The core samples were then dried thoroughly by baking at 50° C to evaporate the methylene chloride and then at 105° C for 24 hours to remove any remaining water in the core sample. Total volume of the dried sample was determined by measuring in a 25 mL graduated cylinder with 0.2 mL graduations.

The volume of solids in each sample was determined by placing the sample in a 25 mL graduated cylinder pre-filled with 12 mL of de-ionized water and measuring the displaced volume. The void volume is then calculated by the total volume minus the solid volume. The TCE saturation of each sample was calculated using:

$$S_{Ni} = \frac{V_{NAPL}}{V_{voids}} \qquad [1]$$

Description of the SSLE algorithm

The inverse model assumes a steady flow field and the natural logarithm of K, θ_w, and θ_N (ln K, ln θ_w, and ln θ_N, respectively) are treated as a stationary stochastic process. The model additionally assumes that the mean and correlation structure of the K, θ_w, and θ_N fields are known, *a priori*. The algorithm is composed of two parts. First, the Successive Linear Estimator (SLE) is used for each dipole cross-hole test to estimate a K tomogram. The estimator begins by cokriging the initial estimate of effective hydraulic conductivity (K_{eff}) and observed heads collected in one pumping test during the tomographic sequence to create a cokriged, mean removed ln K (f, i.e., perturbation of ln K) map for the hydraulic tomography analysis. For the partitioning tracer tomography, the cokriging is done instead using the initial estimates of effective θ_w, and θ_N with the concentration data.

Cokriging does not take full advantage of the observed head values because it assumes a linear relationship (20) between head and K, while the true relationship is nonlinear. To circumvent

this problem, a linear estimator based on the differences between the simulated and observed head values is successively used to improve the estimate. Analogous to inversion of head data, the partitioning tracer tomography utilizes the differences between the concentration measurements in a successive fashion to overcome the nonlinear relationships between concentrations, θ_w, and θ_N.

The second step is to use the hydraulic head data sets (or concentration data for PTT) sequentially instead of including them simultaneously in the inverse model thus the term, "Sequential Successive Linear Estimator" (SSLE) is used to describe the inverse algorithm hereon. In essence, the sequential approach uses the estimated K (or θ_w, and θ_N for PTT) field and covariances, conditioned on previous sets of head measurements (or concentration measurements for PTT) as prior information for the next estimation based on a new set of pumping test data for HT and tracer test data for PTT. This process continues until all the data sets are fully used. Modifications made to the code for this study include its ability to account for variations in the boundary conditions with each pumping test as they are sequentially included and implementing the modified loop scheme (10).

Details to the selection of concentration data for inverse modeling

Since bromide was included in the injection solution as a conservative tracer, it was possible to assess the ability of the K-tomogram to reproduce the transport of the conservative tracer. The observed breakthrough curves at each sample location, for each tracer test were visually compared to a synthetic simulation of the test. In general, there was good agreement between the synthetic and observed cases. Using this comparison, any port that did not show good agreement between the observed and synthetic bromide BTCs were not included in the inverse model for the estimation of TCE saturation. The sampling ports included in the TCE saturation estimation were: test 1-Ports 2, 3, 6, 13, 25, 26, 27, 28; test 2 - Ports 4, 5, 12, 13, 18, 24, 27, 28, 29; test 3 - Ports 3, 4, 5, 6, 12, 24, 25, 26, 27; and test 4 - Ports 1, 5, 25, 26, 27, 28, 29. We selected an average of 5 data points to define the

breakthrough curves at each port. The first point selected was at approximately 1% of the injected concentration. The second point was approximately half the peak concentration. The third point was the peak concentration and the fourth point was half the peak concentration after the passing of the peak and the fifth point represented approximately 1% of the injected concentration after the passing of the peak. The total number of data points used for each test was 40, 55, 60, and 45 for tests 1, 2, 3 and 4, respectively.

Details to the SSLE looping scheme

A looping scheme [*Zhu and Yeh*, 2005]) was developed to improve the SSLE algorithm to fully exploit the data as they are included sequentially. In this scheme, as new data sets are added into SSLE, the algorithm iterates until the data set meets a converge criterion. Then, instead of going straight to the next new data set, the scheme goes back to check the convergence for the first data set. If the converge criterion is not met, the program starts a loop iteration in which the iteration involves both the first and second data sets. That is, the first data set is iterated once, and then the second data set is incorporated and iterated once also in a loop. The loop iteration continues until both data sets meet the converge criterion within one loop. Then, the next new data set is added. The algorithm treats this new data set similarly to the second data set, except the loop iteration now involves three data sets. Additional data sets are added in a similar way. As a consequence, the SSLE approach improves estimates throughout the loops, maximizes the usefulness of data sets, and alleviates the problems associated with the previous SSLE approach.

Appendix E

Estimating Hydraulic Parameters When Poroelastic Effects are Significant

This manuscript by *Berg, S.J., P.A. Hsieh, and W.A. Illman* [2011],was published by the journal *Ground Water.*

EXECUTIVE SUMMARY

For almost 80 years, deformation-induced head changes caused by poroelastic effects have been observed during pumping tests in multilayered aquifer-aquitard systems. As water in the aquifer is released from compressive storage during pumping, the aquifer is deformed both in the horizontal and vertical directions. This deformation in the pumped aquifer causes deformation in the adjacent layers, resulting in changes in pore pressure that may produce drawdown curves that differ significantly from those predicted by traditional groundwater theory. Although these deformation-induced head changes have been analyzed in several studies by poroelasticity theory, there are at present no practical guidelines for the interpretation of pumping test data influenced by these effects. To investigate the impact that poroelastic effects during pumping tests have on the estimation of hydraulic parameters, we generate synthetic data for three different aquifer-aquitard settings using a poroelasticity model, and then analyze the synthetic data using type curves and parameter estimation techniques, both of which are based on traditional groundwater theory and do not account for poroelastic effects. Results show that even when poroelastic effects result in significant deformation-induced head changes, it is possible to obtain reasonable estimates of hydraulic parameters using methods based on traditional groundwater theory, as long as pumping is sufficiently long so that deformation-induced effects have largely dissipated.

Introduction

For close to 80 years, unexpected water-level fluctuations caused by poroelastic effects have been reported during pumping tests in multilayered aquifer-aquitard systems. Two such examples are the Noordbergum effect (Verruijt 1969; Rodrigues 1983) and the Rhade effect (Langguth and Treskatis 1989; Kim and Parizek 1997, 2005). These effects manifest as "reverse" water-level fluctuations (generally on the order of centimeters) when pumping commences at the start of the drawdown phase (Noordbergum effect) and when pumping terminates at the start of the recovery phase (Rhade effect). These water-level responses can be considered "reverse" because water levels in aquitard observation ports adjacent to the pumped aquifer rise temporarily when water is extracted from the ground (Noordbergum effect) and drop briefly when extraction is terminated (Rhade effect). The first rigorous quantitative analysis of the Noordbergum effect was presented by Verruijt (1969) and named after the village in the Netherlands where it was observed (Noordbergum).

The Noordbergum effect is commonly attributed to three-dimensional deformation induced by pumping (Verruijt 1969; Hsieh 1996). In particular, once pumping begins in the aquifer, the reduction in pressure causes the expansion of water and compression of the aquifer skeleton. The latter causes a decrease in aquifer volume, which results in straining of the aquifer in both the horizontal and vertical directions. These strains will, in turn, induce deformation in adjacent layers (Hsieh 1996). It is important to note that deformation-induced effects do not manifest solely as water-level rises. Strains in the pumped aquifer can also induce water-level drops in adjacent aquitards. However, water-level rises at the start of pumping is much easier to recognize because it is not possible to produce these effects with traditional groundwater theory, and thus they stand out as unusual. Deformation-induced drops in water levels are more subtle, and thus they might not be recognized.

In traditional groundwater applications, the horizontal deformation of aquifers is ignored, and it is only the vertical compression of the aquifer that is accounted for (Meinzer 1928; Jacob 1940). For most applications, this simplification is sufficient and allows important hydraulic parameters such as hydraulic conductivity (K) and specific storage (S_s) to be estimated. However, when pumping test data exhibit induced-deformation effects, the analyst may be uncertain in selecting the appropriate method for data analysis, and unsure if the estimated hydraulic parameters are accurate. To our knowledge, only one paper (Rodrigues 1983) has been published that explicitly attempts to estimate hydraulic properties when significant deformation-induced effects are present in the pumping test data. While acknowledging the presence of the deformation-induced head changes (in this case the Noordbergum effect), Rodrigues ignored the rising portion of the water-level data, and used the Neuman and Witherspoon (1972) ratio method to estimate aquitard properties. In his analysis, Rodrigues (1983) noted that using the *"highest water level caused by the Noordbergum effect as the reference level for drawdown computations seems acceptable because leakage seems to start at this point"*. However, since poroelastic effects can manifest in a variety of complex ways and not always as the classical Noordbergum response, the technique described by Rodrigues (1983), where the highest water-level is used as a reference for drawdown is not used in this study. Rather, taking advantage of the transient nature of poroelastic effects, our paper focuses on using water-level data collected after the poroelastic effects have largely dissipated to estimate hydraulic properties.

More recently, the significance of the poroelastic nature of aquifers has been highlighted in the literature. For example, Burbey (2003, 2006) and Burbey et al. (2006) demonstrated that accurate measurements of aquifer deformation during long-duration pumping tests can be used to estimate with a high degree of accuracy the specific storage of the unit experiencing compaction. Moreover, Yin et al. (2007a,b) found that poroelastic effects are "far from negligible" when the compressibility

of the matrix exceeds that of the fluid. For near-surface groundwater investigations, it is not uncommon for the compressibility of the matrix to exceed that of the fluid.

Several numerical models are available that are capable of simulating the poroelastic behavior of aquifer-aquitard systems, including: 1) BIOT2 (Hsieh 1996), a 2-dimensional axisymmetric or plane-strain poroelastic model and one of the first numerical models to simulate the Noordbergum effect; 2) COWADE123D (Kim 1995), a one, two, or three dimensional saturated-unsaturated poroelastic model that has been used to simulate both the Noordbergum and Rhade effects; 3) ABAQUS (Hibbit and Sorenson Inc. 1998), a general purpose finite element model that can be adapted to simulate poroelastic processes and 4) DDFEM (Yin et al., 2007b), which couples the displacement continuity method (DDM) with the finite element method (FEM) to simulate large reservoirs while reducing the computational burden.

Despite the availability of poroelastic models, there is at present a knowledge gap in how one should analyze pumping test data affected by poroelastic effects, and whether the application of methods based on traditional groundwater theory yield reasonable estimates of hydraulic parameters. Therefore, the main objective of this paper is to investigate the ability of two traditional methods (type-curve analysis and the coupling of a groundwater model with a parameter estimation program) to estimate hydraulic parameters when poroelastic effects are significant.

This is accomplished by using BIOT2 (Hsieh 1996) to simulate pumping tests with poroelastic effects in three hypothetical aquifer-aquitard systems. When pumping begins, all three aquifer-aquitard systems exhibit deformation-induced head changes. The synthetic water-level data from these simulations are then analyzed to estimate hydraulic parameters using:

1) the Neuman-Witherspoon (1969) type curves, and

2) a groundwater model that does not account for poroelasticity coupled with the parameter estimation program PEST (Doherty 2005).

The resulting parameter estimates are compared to the values specified in the poroelastic simulations to assess whether these methods based on traditional groundwater theory are capable of producing reasonable estimates of hydraulic parameters when deformation-induced head changes are significant.

Linear Poroelasticity Theory and BIOT2

Pioneering work by Biot (1941) provided a mathematical description of the interaction between pore pressure and matrix deformation in three-dimensions. The solution for a fully three-dimensional system requires solving four separate governing equations (one to describe the change in hydraulic head and three to describe displacement, one for each coordinate direction). However, if one conceptualizes the domain as a two-dimensional axisymmetric system, the number of equations that need to be solved reduces to three (one that describes the change in hydraulic head and two that describe displacement, one for vertical displacement and one for radial displacement). By contrast, in traditional groundwater theory, only one equation needs to be solved. It is due to this additional complexity that poroelastic effects are often ignored in traditional groundwater investigations. In this study the following assumptions are made for all poroelastic simulations.

1) The entire model domain is initially at steady state (or hydraulic and mechanical equilibrium). The calculated hydraulic heads and displacements represent deviations from the initial state.
2) The fluid is compressible, but solid grains are incompressible. The latter assumption is valid as long as the grain compressibility is much lower than the matrix compressibility, which is often the case in sedimentary materials.
3) Each hydrogeologic layer (aquifer or aquitard) is isotropic and homogeneous.

Application of these assumptions to the linear poroelasticity theory of Biot (1941) simplifies the form of the equations to that presented in Verruijt (1969) and Hsieh (1996):

$$K\nabla^2 h = \frac{\partial}{\partial t}\left(\nabla \cdot \underline{u}\right) + \rho_f g n \beta_f \frac{\partial h}{\partial t} \quad (1)$$

$$G\nabla^2 \underline{u} + \frac{G}{1-2\nu}\nabla\left(\nabla \cdot \underline{u}\right) - \rho_f g \nabla h = 0 \quad (2)$$

where K is hydraulic conductivity, h is change in hydraulic head from the initial head, $\underline{u}$ is displacement (change in position from the initial position) of the skeletal matrix, ρ_f is fluid density, g is the gravitational constant, n is porosity, β_f is fluid compressibility, G is the shear modulus of the skeletal matrix, and ν is the drained Poisson's ratio of the skeletal matrix.

For an axisymmetric domain, Equations 1 and 2 reduce to three component equations with three unknowns: the change in hydraulic head (h) and the vertical and radial components of $\underline{u}$ (u_r and u_z). In order to model this axisymmetric system, the specification of three boundary conditions at each of the model boundaries is required (one for each unknown, h, u_r and u_z). BIOT2 allows for either constant-head or specified-flux boundaries to be assigned as hydraulic boundary conditions. The possible mechanical boundary conditions are specified displacement or specified change in boundary traction (applied stress) (Hsieh 1996).

Hypothetical aquifer-aquitard systems

Three hypothetical aquifer-aquitard systems are used to examine the effect of deformation-induced head changes on results of pumping-test analysis. Figure E1 presents a schematic diagram of the three different model domains, referred to as Case 1, Case 2, and Case 3, showing the pumped well location as well as monitoring locations (dots). In each case, the pumped well (located in the 'pumped aquifer') is screened across the entire thickness of the aquifer (100 m), has a radius of 0.1 m, is located at the inner boundary of the model, and is pumped at a rate of 50 L/s for a period of 50 days. The time step is initially 15 seconds, and increases by a multiple of 1.2 per time step until model

completion (a total of 60 time steps). A summary of the hydraulic properties assigned to each unit for the simulations is included in Table 1.

The following boundary conditions are assigned for all simulations. 1) The inner radial boundary is a specified-flux boundary and experiences no radial displacement and no change in the vertical component of boundary traction. 2) The outer radial boundary (located at a distance of 10km) is a no-flow boundary with no change in applied stress. 3) The top boundary is a water table and is represented by a constant-head boundary. This assumes that any drawdown that would have occurred at the water table is negligible. This boundary is free of applied stress and is free to deform. 4) The bottom boundary is a no-flow boundary and a no-displacement boundary.

Case 1

This simulation is designed to produce a Noordbergum effect that is similar to that shown in Figure 4 of Hsieh (1996). The model setting as shown in the top panel of Figure E1 is similar to the model setting shown in Figure 1 of Hsieh (1996), except in this case the lower aquitard hydraulic conductivity (1×10^{-9} m/s) is two orders of magnitude smaller than the upper aquitard hydraulic conductivity. This causes the lower aquitard to yield negligible flow to the pumped aquifer. The upper aquitard is 100 m thick, the pumped aquifer is 100 m thick, and the lower aquitard is 9800 m thick. This places both the bottom boundary and outer radial boundary at a distance of 10 km. These large distances were selected to minimize any interference that may be caused by boundary effects. The 10 km x 10 km domain is discretized into a rectangular mesh with 40 variably-sized elements in the horizontal direction and 100 variably-sized elements in the vertical direction. The accuracy of the simulation was verified by making additional runs using a refined mesh, which yielded similar results.

Case 2

This case simulates a 100 m thick pumped aquifer overlying a rigid and impermeable boundary and underlying a 100 m thick aquitard (Figure E1, middle panel). The outer radial boundary is at a distance of 10 km. The domain is discretized into a rectangular mesh with 40 variably-sized elements in the horizontal direction and 48 variably-sized elements in the vertical direction. This case is identical to Case 1, except that the lower aquitard in Case 1is replaced with a rigid and impermeable boundary in Case 2 to assess the impact of such a boundary on deformation-induced head change.

Case 3

Case 3 (Figure E1, bottom panel) includes a 50 m thick unpumped aquifer overlying the aquitard in Case 2. This aquifer has identical properties to the pumped aquifer. Adding this unpumped aquifer allows us to investigate the effect of the position of the top boundary on head changes in the aquitard overlying the pumped aquifer. As with Cases 1 and 2, the outer radial boundary is located at a distance of 10 km. The pumped aquifer and overlying aquitard are both 100 m thick. The domain is discretized into a rectangular mesh with 40 variably-sized elements in the horizontal direction and 79 variably-sized elements in the vertical direction.

For the purpose of interpreting the synthetic pumping tests, a total of 30 observation points are included in the domain. As shown in Figure E1, the observation points are divided into five columns of six points each, mimicking a set of multilevel observation wells. Within each group, the observation points are located at different depths below ground surface: 10 m, 30 m, 50 m, 70 m, 90 m, and 150 m for Cases 1 and 2, and 60 m, 80 m, 100 m, 120 m, 140 m, and 200 m for Case 3. This positioning means that for all cases, observation points are located at identical points within each system relative to the vertical center of the pumped aquifer. This allows for a direct comparison of the synthetic water-level data for the different cases. The upper 5 observation points are all contained

within the overlying aquitard, while the 6th (deepest) observation point is located in the vertical center of the pumped aquifer. The five multilevel observation wells are located at increasing radial distances from the pumped well: 10 m, 25 m, 50 m, 100 m, and 200 m. The observation points are placed at these locations to investigate how parameter estimates are affected by data collected at different radial distances from the pumped well. Additionally, observation points are only included in the pumped aquifer and overlying aquitard as the underlying aquitard (Case 1) is often not instrumented in field investigations, and hydraulic parameters in the near surface unpumped aquifer (Case 3) would be determined by tests performed directly within this unit.

Estimation of Aquifer Parameters Using the Neuman-Witherspoon Type-Curve Matching Method

The Neuman-Witherspoon (1969) model consists of two confined aquifers separated by an aquitard. The lower (pumped) aquifer rests on an impermeable boundary. The upper aquifer is not pumped. The Neuman-Witherspoon (1969) solution was selected because it provides a type-curve matching method that uses water-level data collected in the aquitard to estimate aquitard hydraulic parameters.

The assumptions behind the Neuman-Witherspoon (1969) model are that: 1) the aquifers and aquitard have infinite areal extent; 2) the aquifers and aquitard are homogeneous and isotropic with uniform thickness; 3) the aquifer-aquitard system is initially at hydraulic equilibrium; 4) the pumped well is fully penetrating; 5) flow in the aquifers is horizontal; 6) flow in the aquitard is vertical; 7) flow is transient; 8) groundwater is released instantaneously from storage with decline of hydraulic head due to pumping; 9) the pumped well is treated as a line sink so it neglects borehole storage.

The key between the Neuman-Witherspoon (1969) model and the three hypothetical cases presented here occurs at the top boundary. The upper (unpumped) aquifer in Neuman-Witherspoon (1969) model is a confined aquifer, and therefore the top boundary is a no-flow boundary. By

contrast, in Cases 1, 2, and 3, the top boundary is a constant-head boundary. Nonetheless, the Neuman-Witherspoon (1969) can be used to analyze the synthetic data generated for Cases 1, 2, and 3 by assuming negligible downdown in the upper aquifer. This can be implemented by setting high values of hydraulic conductivity and specific storage for the upper aquifer during type curve matching.

The Neuman-Witherspoon (1969) analytical solution, as implemented in the aquifer test analysis software Aqtesolv (Duffield 2007), is simultaneously matched to all the observation points at the same radial distance from the pumped well. This match is repeated for each of the 5 radial distances. Aqtesolv provides direct estimates of pumped aquifer hydraulic conductivity (K), pumped aquifer specific storage (S_s), and the hydraulic conductivity of the aquitard overlying the pumped aquifer (K'). However, it does not provide a direct estimate of specific storage for the aquitard overlying the pumped aquifer (S_s'). Rather it estimates the leakage parameter, β, which can be used along with the other estimated parameters to calculate S_s'. Specifically,

$$S_S' = \frac{\left[\left(\frac{\beta}{r}\right)(4b)\right]^2 (KS_S)}{K'} \qquad S_s' = \frac{\left[\left(\frac{\beta}{r}\right)(4b)\right]^2 (KS_s)}{K'}$$

(3)

where:

S_s' = the specific storage of the aquitard overlying the pumped aquifer;

$\frac{\beta}{r}$ = the leakage parameter β divided by the radial distance of the observation point from the pumped well (this parameter is estimated by Aqtesolv when fitting the water-level data);

b = the thickness of the pumped aquifer;

K = the hydraulic conductivity of the pumped aquifer;

S_s = the specific storage of the pumped aquifer; and

K' = the hydraulic conductivity of the overlying aquitard.

The type-curve matching procedure using Aqtesolv is based on the "Curve Matching Tips" from the Aqtesolv User's Guide (Duffield 2007) for the Neuman-Witherspoon (1969) type curves. This method involves initially estimating aquifer properties using the Cooper and Jacob (1946) solution. After this, the Neuman-Witherspoon (1969) solution is approximately fit using visual curve matching, followed by the automatic estimation feature to achieve a final fit. The automatic estimation feature uses a nonlinear least squares procedure to reduce the error between the type curve and the water-level data. This is accomplished iteratively by adjusting the hydraulic parameters until convergence criteria are satisfied (Duffield 2007).

Limitations of the Neuman-Witherspoon (1969) Type Curves

Before using the Neuman-Witherspoon (1969) type curves to analyze water-level data from pumping tests where deformation-induced head changes are significant, a series of non-poroelastic simulations (i.e., standard groundwater model runs) were performed for all three cases using BIOT2 with poroelastic effects disabled. The model parameters used in these non-poroelastic simulations are summarized in Table 1. The results of type-curve matching for non-poroelastic versions of Cases 1, 2, and 3 are shown in Figures E2a, E2b, and E2c respectively. The most striking features in Figure E2 are the large differences between the K' and S_s' values estimated using data near the pumped well and the corresponding 'true' K' and S_s' (parameters used in the non-poroelastic simulation). Although the estimated parameters converge towards the corresponding true parameters when data are used from observation points at larger radial distances, the large errors that result from matching to data near the pumped well suggest there may be limitations to the Neuman-Witherspoon (1969) type curves.

The Neuman-Witherspoon (1969) solution relies on many assumptions. Of particular note is the assumption that only vertical flow occurs in the aquitard. It is reasonable to expect some horizontal component of flow which would become less significant with increasing distance from the pumped well. Figure E3 shows the distribution of changes in hydraulic-head simulated by BIOT2 without poroelastic effects for Case 3 at the end of the model run, when drawdown in the pumped aquifer has fully propagated across the thickness of the aquitard to the unpumped aquifer. In the aquitard, horizontal gradients are greatest near the bottom of the aquitard and close to the pumped well. Therefore, error caused by the vertical flow assumption would be expected to be significant near the pumped well and decrease with increasing radial distance from the pumped well. This is similar to the trend in the error seen in Figure E2 when analyzing drawdown with the Neuman-Witherspoon (1969) solution for non-poroelastic simulations. As such, it is reasonable to suspect the radial flow component in the aquitard to be the source of the large error in K' and S_s' when these parameters are estimated using data close to the pumped well.

To further examine the Neuman-Witherspoon (1969) assumption of vertical flow in the aquitard, an additional simulation, referred to as Case 3a, is made. In the original Case 3, the hydraulic conductivity in the aquitard is assumed to be isotropic ($K'_r = K'_v$). In Case 3a, the radial component of the hydraulic conductivity in the aquitard is set to zero ($K'_r = 0$). This modification results in strictly vertical flow in the aquitard and is therefore consistent with the assumption in the Neuman-Witherspoon (1969) model. Figure E4 shows comparisons between the Neuman-Witherspoon (1969) solution and the simulated drawdowns in Case 3 and Case 3a. All the results are computed using the hydraulic parameters in Table 1. To account for the upper constant-head boundary of the numerical model (Cases 3 and 3a) the K and S_s of the overlying unpumped aquifer in the Neuman-Witherspoon (1969) solution are set to relatively high values (1×10^{-2} m/s and 1×10^{-4} m^{-1} respectively). Inspection of Figure E4 shows that the Neuman-Witherspoon (1969) solution

deviates from drawdowns simulated in Case 3 (with $K'_r = K'_v$) at small radial distances (e.g., r = 10 m, and less so at r = 25 m). However, the Neuman-Witherspoon (1969) solution is able to accurately reproduce the drawdowns simulated in Case 3a (with $K'_r = 0$) at all radial distances. This comparison confirms that, when curve matching to aquitard drawdown data close to the pumped well, the large errors in the estimated K_r' and S_s' can be attributed to the Neuman-Witherspoon (1969) assumption of vertical flow in the aquitard.

Poroelastic Simulations of Aquifer Tests

We next simulate Cases 1 - 3 with poroelastic effects enabled in BIOT2. For these simulations, instead of specifying S_s and S_s', the mechanical parameters G (shear modulus) and ν (drained Poisson's Ratio) are specified. Table 1 summarizes the hydraulic and mechanical parameters used in all three cases for the poroelastic simulations. We note that these values have been shown by Hsieh (1996), to be capable of producing strong deformation-induced head changes in the aquitard overlying the pumped aquifer for Case 1.

When examining the aquifer parameters obtained by type-curve analysis, it is straightforward to compare the hydraulic conductivities (K and K') estimated from the type-curve analysis with the corresponding hydraulic conductivities used in the poroelastic simulation. The latter conductivities are referred to as the "true" K and K'. By contrast, there is no rigorous way to evaluated the specific storages (S_s and S_s') estimated from the type-curve analysis, as specific storage is not used in BIOT2. Nonetheless, for the purpose of comparison, specific storage values are computed for the poroelasticity simulation as follows. First, the parameters G and ν from the poroelastic simulation are used to calculate the vertical matrix compressibility β_p, according to (Hsieh 1996 Equation 3):

$$\beta_p = \frac{1-2\nu}{2G(1-\nu)} \qquad (4)$$

Then, the specific storage is computed using (Domenico and Schwartz 1997 Equation 4.30):

$$S_s = \rho_f g\left(\beta_p + n\beta_f\right) \tag{5}$$

Equation 5 gives the standard specific storage in groundwater theory. In poroelasticity theory, this quantity is sometimes referred to as "uniaxial specific storage" (Wang 2000), because Equation 5 is based on the traditional groundwater assumption that deformation occurs only in the vertical direction and under constant vertical stress. Because deformation is generally three dimensional in the poroelasticity simulation, the S_s computed by Equation 5 is not necessarily the "true" value against which the specific storages estimated by type-curve analysis are evaluated. Instead, the S_s computed by Equation 5 is intended for comparison with the estimated S_s.

Figure E5 shows the changes in hydraulic head for all three cases. Solid lines represent head changes in the poroelastic simulation. Dashed lines represent head changes in the corresponding non-poroelastic simulation, using values of K and K' from the poroelastic simulation and values of S_s and S_s' computed by Equations 5. Inspection of the three cases (Figures E5a, E5b, and E5c) shows that deformation-induced head changes in the aquitard can manifest in different ways. Hydraulic head in the aquitard could initially rises after pumping begins (Noordbergum effect) and then falls as pumping continues. Alternatively, hydraulic head in the aquitard could start to fall earlier than the start of drawdown in the corresponding non-poroelastic simulation. However, in all cases, the head changes in the poroelastic simulation approach the corresponding head changes in the non-poroelastic simulation after poroelastic effects have largely dissipated.

The simulated head changes in Case 1 (Figure E5a) show that, with one exception, deformation-induced head rise occurs in all observation points in the aquitard. The magnitude of the head rise depends on vertical position within the aquitard and radial distance from the pumped well, with those observation points closest to the pumped well (r = 10 m) and in the upper part of the

aquitard (z = 30 m) experiencing the greatest head rise. Near the top of the aquitard (z = 10 m), the head rise is lessened due to proximity to the constant-head boundary. The exception to head rise occurs at the observation point at r = 10 m and z = 90 m. At this observation point, hydraulic head starts to fall earlier than the start of drawdown in the corresponding non-poroelastic simulation. This behavior is consistent with the finding by Hsieh (1996, p. 1085)—deformation caused by pumping can induce a small region of pore volume increase, and therefore head drop, in the aquitard near the pump interval.

The simulated head changes in Case 2 (Figure E5b) are similar to those of Case 1. The deformation-induced head rises in Case 2 are slightly lower than the corresponding head rises in Case 1. The deformation-induced head drop (at r = 10 m, z = 90 m) in Case 2 is more prominent than that of Case 1. The overall similarity in simulated head changes for Case 1 and for Case 2 (Figures E5a and E5b) suggests that deformation-induced head changes in the overlying aquitard is relatively insensitive to how the underlying aquitard is represented; whether as a thick, deformable layer (Case 1) or as a rigid, impermeable boundary (Case 2).

The simulated head changes for Case 3 (Figure E5c) are significantly different from those of Cases 1 and 2. In Case 3, there is no head rise at the aquitard observation points. Instead, the onset of drawdown in the aquitard occurs much earlier in the poroelastic case than in the non-poroelastic case. This behavior is more prominent at small radial distances and diminishes with distance from the pumped well (Figure E5c). The absence of a 'reverse' water-level response in Case 3 seems to be the result of including the overlying aquifer and the change in position of the top (traction-free) boundary. For Cases 1 and 2 (Figures E5a and E5b), the analyst would likely recognize that the rising portion of the head data cannot be matched by traditional type curves and possibly exclude it from the analysis. However, for Case 3, if the deformation-induce head drop is interpreted as hydraulic

propagation from the pumped aquifer into the aquitard, this could severely overestimate the aquitard hydraulic diffusivity (K'/S_s').

Neuman-Witherspoon (1969) Type-Curve Analysis of Data Showing Deformation-Induced Effects

Following the same procedure as for non-poroelastic cases (Figure E2), type-curve matching using the Neuman-Witherspoon (1969) solution was performed on the synthetic data from the poroelastic simulations to estimate the hydraulic parameters, K, K', S_s and S_s'. Figure E6 (a, b, and c) shows the type-curve matches for all Cases 1, 2, and 3 respectively. Note that negative drawdowns (head rises) are ignored by Aqtesolv and do not appear on the graphs presented in Figure E6. Despite the complex nature of the drawdown curves, the Neuman-Witherspoon (1969) solution is still able to achieve reasonable fits to the synthetic data after deformation-induced effects have largely dissipated.

The results of the type-curve matching for the poroelastic case are presented in Figure E7. The most striking features in Figure E7 occur when matching to data near the pumped well: (1) the estimated K' is significantly larger than the 'true' K' used in the poroelastic simulation, and (2) the estimated S_s' is significantly larger than the S_s' calculated with Equation 5, (with β_p computed by Equation 4 using G and v from the poroelastic simulation). The similarity of these features to those in Figure E2 suggests that the Neuman-Witherspoon (1969) assumption of horizontal flow in the overlying aquitard may also be the predominant source of error in the K' and S_s' estimates for the poroelastic cases. For the pumped aquifer, K is slightly lower than the true K when matching to data near the pumped well. When matching to data at a greater radial distance from the pumped well, the estimated K becomes slightly higher than the true K in Cases 1 and 2, and approaches the true K in Case 3. For S_s, the type-curve estimates are generally higher than the S_s value computed by Equation 5. These differences are discussed at greater length in the next section.

Estimation of Hydraulic Parameters by Traditional Groundwater Modeling Coupled with a Parameter Estimation Program

A commonly used numerical approach for estimating hydraulic parameters is to couple a groundwater model with a parameter estimation program. To evaluate this approach, BIOT2 is run as a traditional groundwater model by turning off its poroelasticity capability. It is then coupled with the parameter estimation code PEST (Doherty 2005) to estimate hydraulic properties. At each radial distance the synthetic head data generated from the poroelastic simulations for Cases 1, 2, and 3 are matched by PEST using BIOT2 without poroelastic effects. The starting parameter values used in the parameter estimations are those estimated at a radial distance of 10 m by the Neuman-Witherspoon (1969) solution.

For each case, two different matches are performed. The first match (denoted as PEST-1) uses all of the synthetic head data (including those that exhibit deformation-induced head changes). The second match (denoted as PEST-2) excludes head data that are influenced by deformation-induced effects. For each case, hydraulic conductivity and specific storage are estimated for the pumped aquifer (K and S_s) and for the overlying aquitard (K' and S_s'). For Case 1, the hydraulic conductivity and specific storage of the underlying aquitard could not be estimated due to the lack of head data in that layer. In the poroelastic simulation for Case 1, the hydraulic conductivity of the underlying aquitard is two orders of magnitude lower than that of the overlying aquitard. For parameter estimation, the hydraulic conductivity and specific storage of the underlying aquitard are set to very small values so that this layer yields negligible flow to the pumped aquifer. Similarly, for Case 3, the parameters for the upper (unpumped) aquifer could not be estimated due to lack of head data in that layer. Inspection of the poroelastic simulation for Case 3 shows that the drawdown in the unpumped aquifer is less than 0.1 cm at the end of the simulation. This is a result of the constant-head

boundary at the top of the domain. Therefore, as long as sufficiently high values of hydraulic conductivity and specific storage are specified for the upper aquifer, the parameter estimates for the pumped aquifer and aquitard should be largely unaffected. Several PEST runs were performed using a variety of fixed values for the unpumped aquifer and this indeed was the case. For the final PEST estimation for Case 3, the hydraulic conductivity and specific storage of the unpumped aquifer were fixed at 1×10^{-2} m/s and 1×10^{-4} m^{-1} respectively.

The results of the PEST estimation are presented on Figure E7 to allow comparison with the results of the Neuman-Witherspoon type-curve matching. In the overlying aquitard, the K' estimates obtained by PEST using data near the pumped well are significantly closer to the true K' than the K' estimates obtained by type-curve matching. This is because the numerical model used with PEST accounts for radial flow in the aquitard, whereas the Neuman-Witherspoon (1969) model does not. For the same reason, the S_s' estimates obtained by PEST using data near the pumped well are significantly closer to the specific storage value computed with Equation 5 than the S_s' estimates obtained by type-curve matching. When matching to data at greater distances from the pumped well, both PEST and type-curve matching yield similar estimates for K' and for S_s'.

For the pumped aquifer, the K estimates obtained by PEST and by type-curve matching are all within ±20% of the true K. However, the S_s estimates obtained by PEST and by type-curve matching are generally higher than the specific storage value computed using Equations 5. This difference is particularly noticeable in Case 1. This observation suggests that the volume of fluid released by a unit volume of porous medium per unit decline in hydraulic head in a poroelasticity simulation (with three-dimensional deformation) is somewhat larger than the fluid released in a traditional groundwater simulation using S_s computed by Equation 5. For the aquifer-aquitard systems examined in this study, the S_s estimated by well test analysis can be 10% to 60% higher than the S_s

computed by Equation 5. However, these differences are comparable to uncertainties associated with parameters estimated from typical pumping tests.

Figure E7 also shows that little is gained by the PEST-2 runs, which exclude head data that are influenced by deformation-induced effects, compared to the PEST-1 runs, which include all head data. This is likely due to the relatively small magnitude of the deformation-induced effects when compared to the increasing magnitude of drawdown as the pumping test progresses. The relatively large drawdowns during the latter part of the test dominate the estimation of the hydraulic parameters.

Summary and Conclusions

Deformation-induced head changes, which are most conspicuous in confining layers overlying or underlying the pumped aquifer, are often unobserved because confining layers are often not monitored during a pumping test. However, even when deformation-induced head changes are recorded, they are often overlooked, unrecognized, or ignored. Possible reasons for this are: 1) the effect is often small and therefore missed; 2) if observed, it may be attributed to other transient effects (diurnal fluctuations, sensor drift or signal noise); and 3) data may not be collected at a frequency high enough to record the effect often seen during a relatively brief period after pumping begins. However, the impact of ignoring the poroelastic effects on the estimation of aquifer parameters has not been addressed in the literature to date.

The poroelastic model BIOT2 (Hsieh, 1996) is used to simulate deformation-induced head changes during a pumping test in three different aquifer-aquitard settings. The deformation-induced head changes are produced when water in a pumped aquifer is released from compressive storage in the region close to the pumped well. This release from storage is accompanied by a volume reduction of the aquifer, both in the horizontal and vertical direction. This deformation of the pumped aquifer causes deformation in the adjacent layers, resulting in complex pore pressures that may manifest in a

variety of ways. As shown in Figures E5a, E5b, and E5c, aquifer deformation can induce head rise or head drop in adjacent layers.

Two methods based on traditional groundwater theory are used to match the head changes simulated by the poroelastic model. The type-curve matching method uses the analytical solution of Neuman and Witherspoon (1969) as implemented in the aquifer test analysis software Aqtesolv (Duffield 2007). The parameter estimation method uses BIOT2 with poroelasticity turned off (thus acting as a traditional groundwater model) and coupled with parameter estimation program PEST (Doherty 2005). Both methods are used to estimate the hydraulic conductivity and specific storage of the pumped aquifer (K and S_s) and of the overlying aquitard (K' and S_s').

A limitation of the Neuman-Witherspoon (1969) solution is the assumption of vertical flow in the aquitard overlying the pumped aquifer. In the region near the pumped well, horizontal flow in the aquitard might be significant. Consequently, aquifer parameters estimated by type-curve matching to aquitard head data collected close to the pumped well might suffer from significant errors. By contrast, parameter estimation using PEST coupled with a numerical groundwater model does not suffer from these errors because the numerical model can simulate both horizontal and vertical flow in aquitards.

While it is straightforward to evaluate the estimated hydraulic conductivities (K and K') by comparing them to the corresponding parameters used in the poroelastic simulation, there is no rigorous way to evaluate the estimated specific storages (S_s and S_s'). This is because specific storage is not a parameter used in the poroelasticity simulations. While it is possible to calculate a specific storage according to equation 5, such a computation assumes that deformation occurs only in the vertical direction. For the aquifer-aquitard systems examined in this study, the S_s estimated by well test analysis can be 10% to 60% higher than the S_s computed by Equation 5. However, these

differences are comparable to uncertainties associated with parameters estimated from typical pumping tests.

For the most part, both the Neuman-Witherspoon type-curve matching method and the PEST estimation method yield reasonable estimates for hydraulic parameters, as long as pumping is sufficiently long so that deformation-induced effects have largely dissipated. The exception occurs when applying the curve-matching method to estimate aquitard parameters using data collected in the aquitard near the pumped well. As noted above, the assumption of vertical flow in the aquitard can cause significant errors in these estimates. Excluding head data that exhibit deformation-induced effects does not yield significant improvements in the estimated parameters.

Table E1. Parameters used for model simulations (after Hsieh 1996).

Property	Aquifer Value	Aquitard Value
[1]S_s, specific storage (m^{-1})	1.2×10^{-5}	1.1×10^{-4}
K, hydraulic conductivity (m/s)	1×10^{-4}	1×10^{-7} (1×10^{-9})*
n, porosity (dimensionless)	0.3	0.4
B_w, fluid compressibility (m^2/N)	4.4×10^{-10}	4.4×10^{-10}
[2]G, shear modulus (N/m^2)	3×10^{8}	3×10^{7}
[2]v, drained Poisson's Ratio (dimensionless)	0.25	0.25

[1] Parameters used for simulation without poroelastic effects
[2] Parameters used for simulation with poroelastic effects
* Value of lower aquitard for Case 1

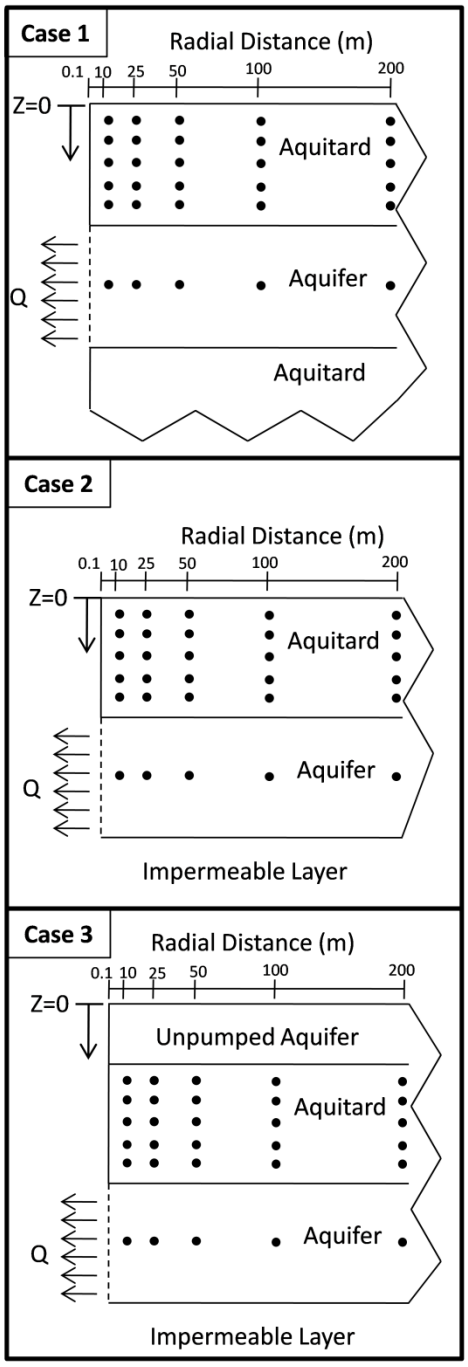


Figure E1. Schematic diagram of the 2 dimensional axisymmetric domain for Case 1 after (Hsieh 1996), Case 2, and Case 3. (Dots represent observations points).

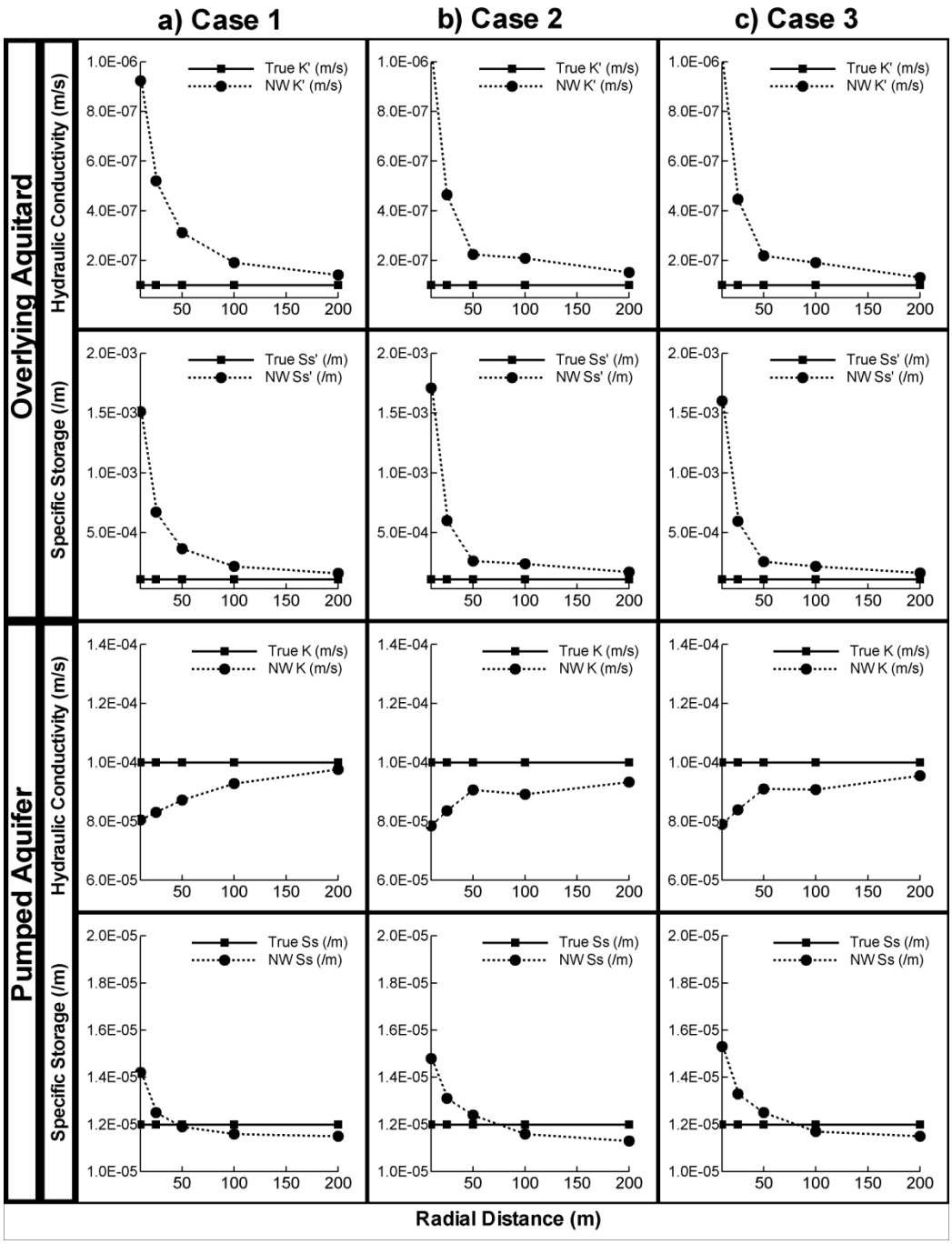


Figure E2. Hydraulic parameters estimated by the Neuman-Witherspoon (1969) type-curve method for Cases 1, 2 and 3 without poroelasticity effects. The solid line with squares represents the true value as defined in the numerical simulation. The dashed line with circles represents the estimated parameter using data at varying radial distances.

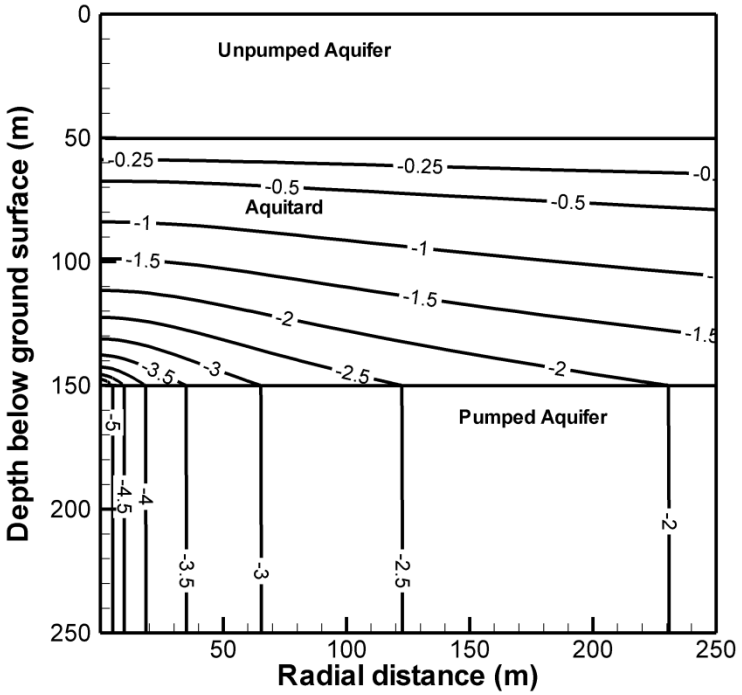


Figure E3. Distribution of changes in hydraulic-head (m) simulated by BIOT2 without poroelastic effects for Case 3 at the end of the model run.

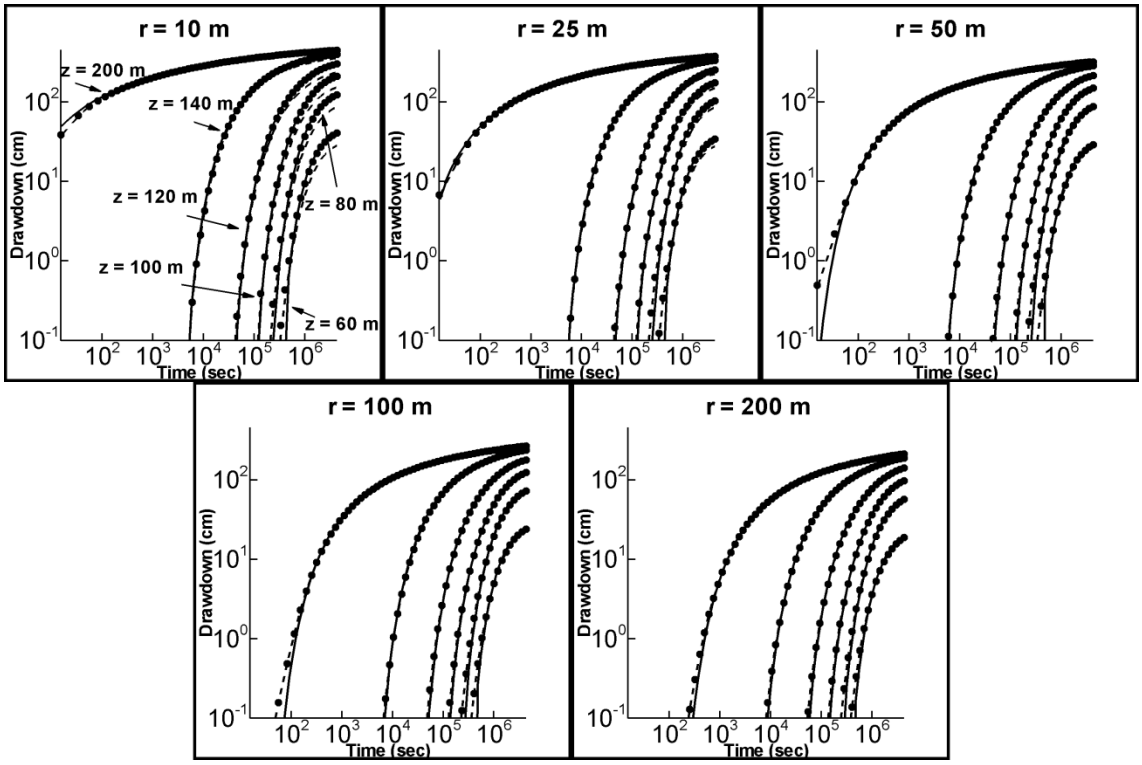


Figure E4. Comparison between the Neuman-Witherspoon (1969) solution (solid line) and drawdown simulated with BIOT2 without poroelasticity effects for Case 3 ($K'_r = K'_v$, dashed line) and Case 3a ($K'_r = 0$, dots).

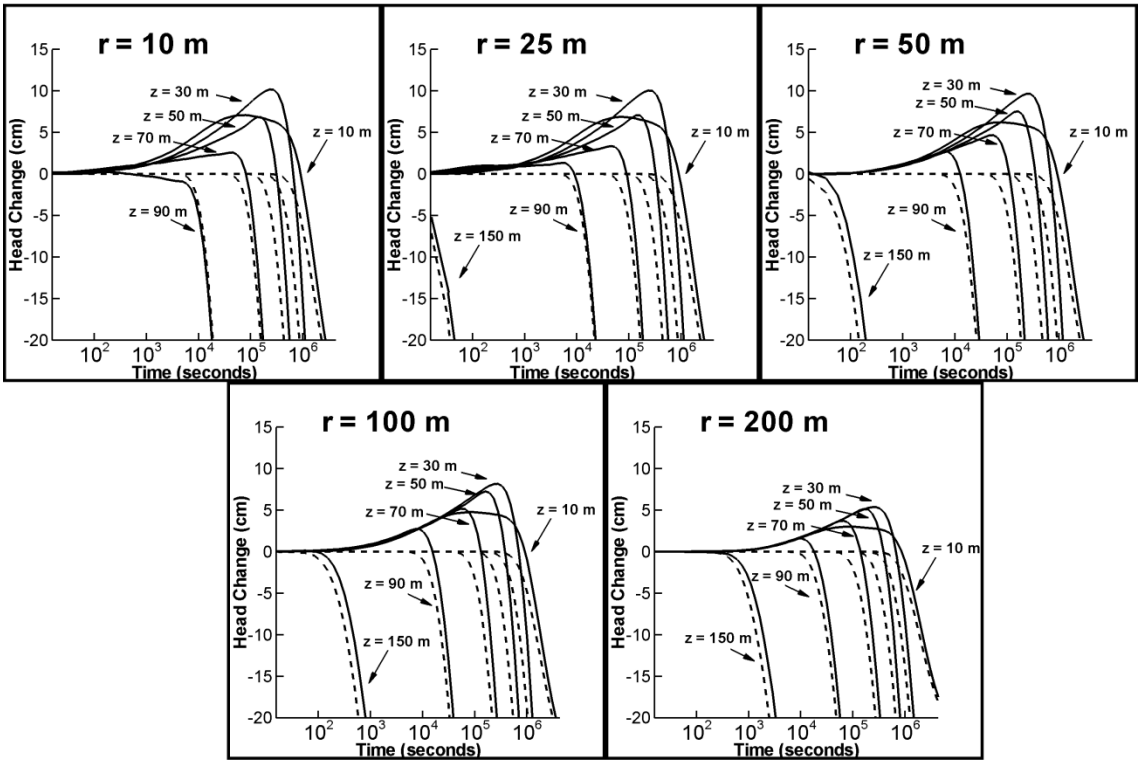


Figure E5a. Change in hydraulic head vs. time for Case 1. Solid line represents the simulation with poroelastic effects. Dashed line represents the simulation without poroelastic effects. Positive head change indicates head rise. Negative head change indicates head drop.

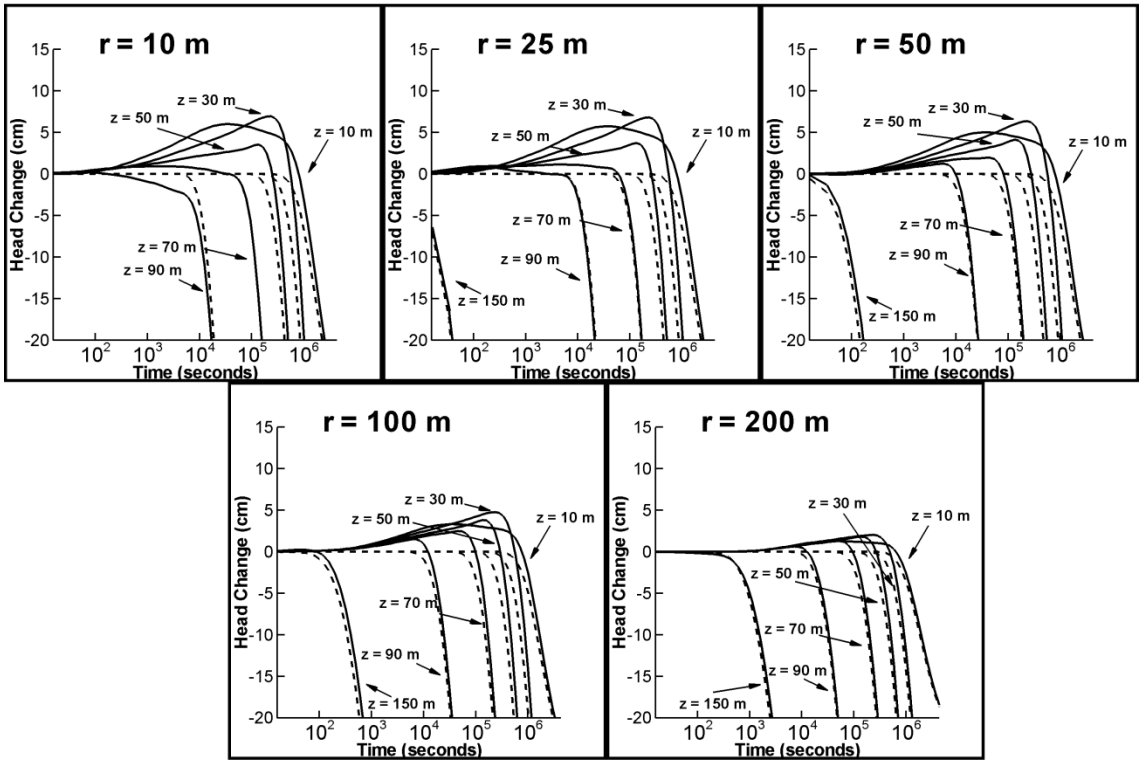


Figure E5b. Change in hydraulic head vs. time for Case 2. Solid line represents the simulation with poroelastic effects. Dashed line represents the simulation without poroelastic effects. Positive head change indicates head rise. Negative head change indicates head drop.

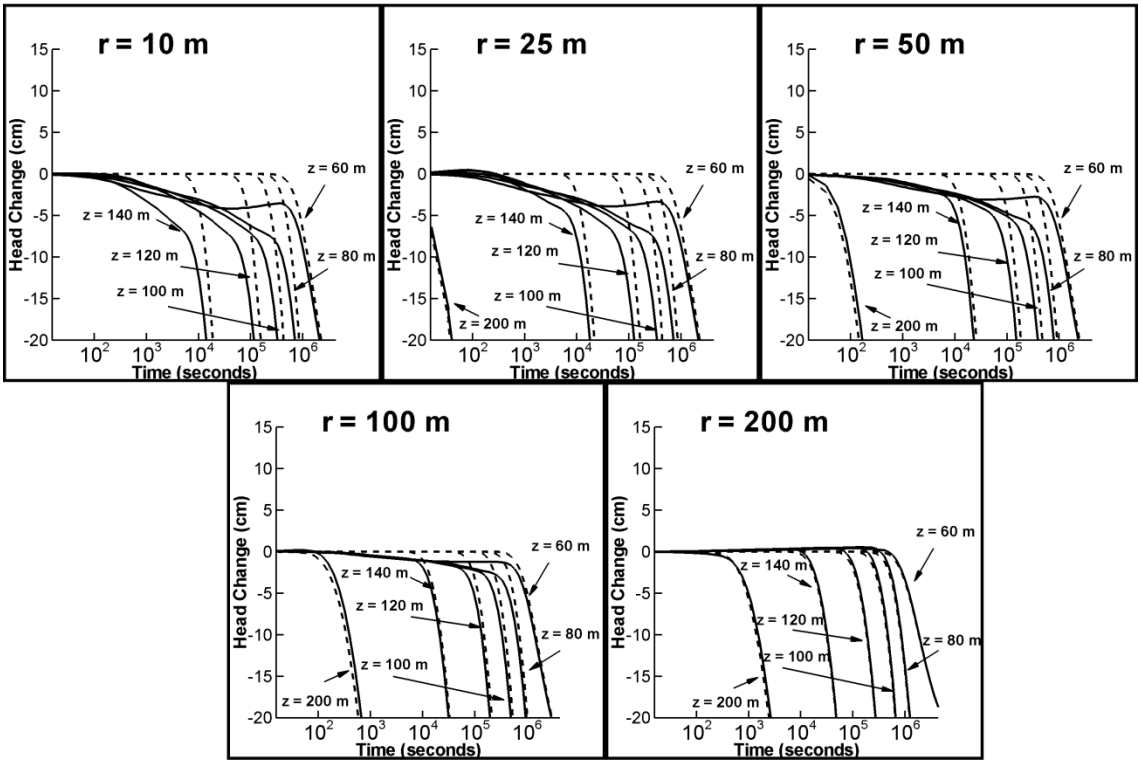


Figure E5c. Change in hydraulic head vs. time for Case 3. Solid line represents the simulation with poroelastic effects. Dashed line represents the simulation without poroelastic effects. Positive head change indicates head rise. Negative head change indicates head drop.

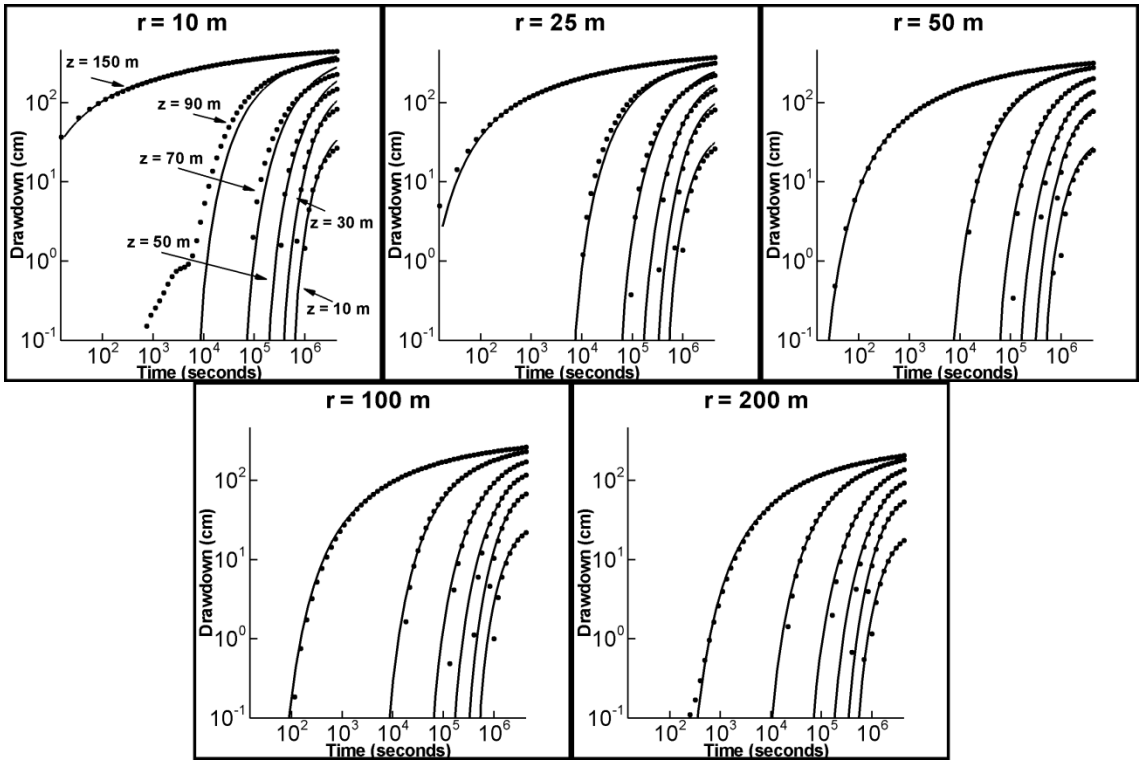


Figure E6a. Neuman-Witherspoon type-curve matches for synthetic data (dots) at varying radial distances for Case 1 with poroelastic effects. Solid lines represent the analytical solution.

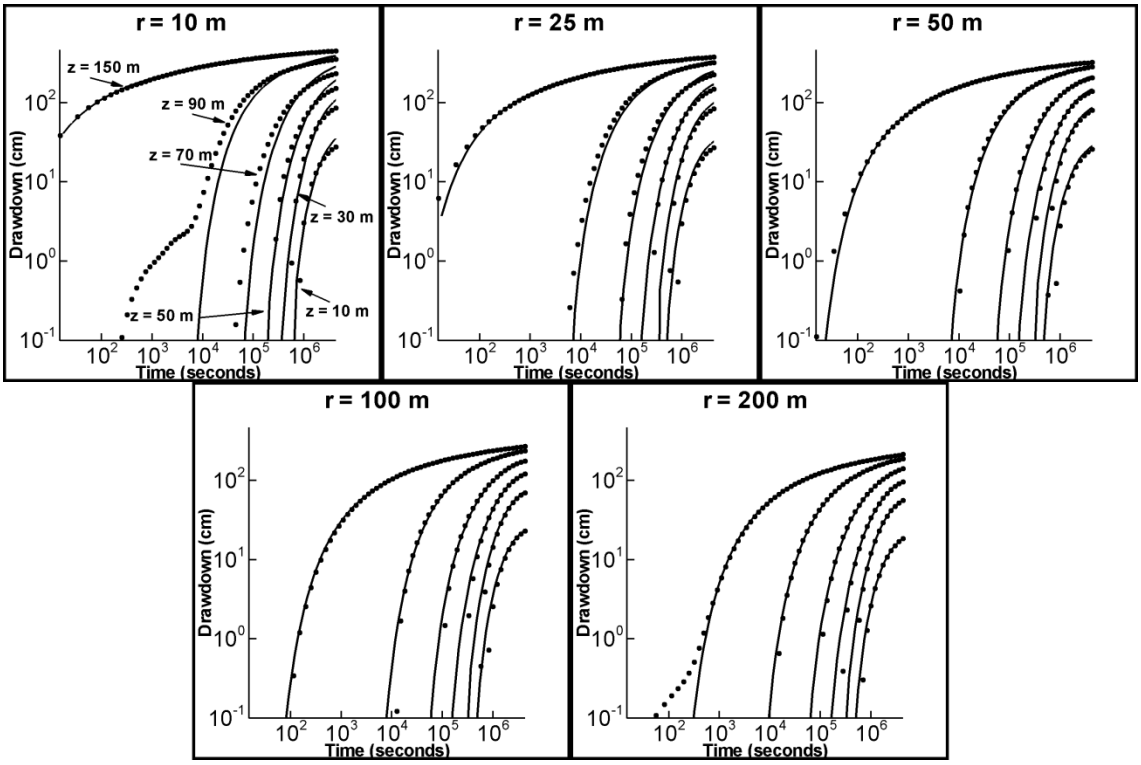


Figure E6b. Neuman-Witherspoon type-curve matches for synthetic (dots) at varying radial distances for Case 2 with poroelastic effects. Solid lines represent the analytical solution.

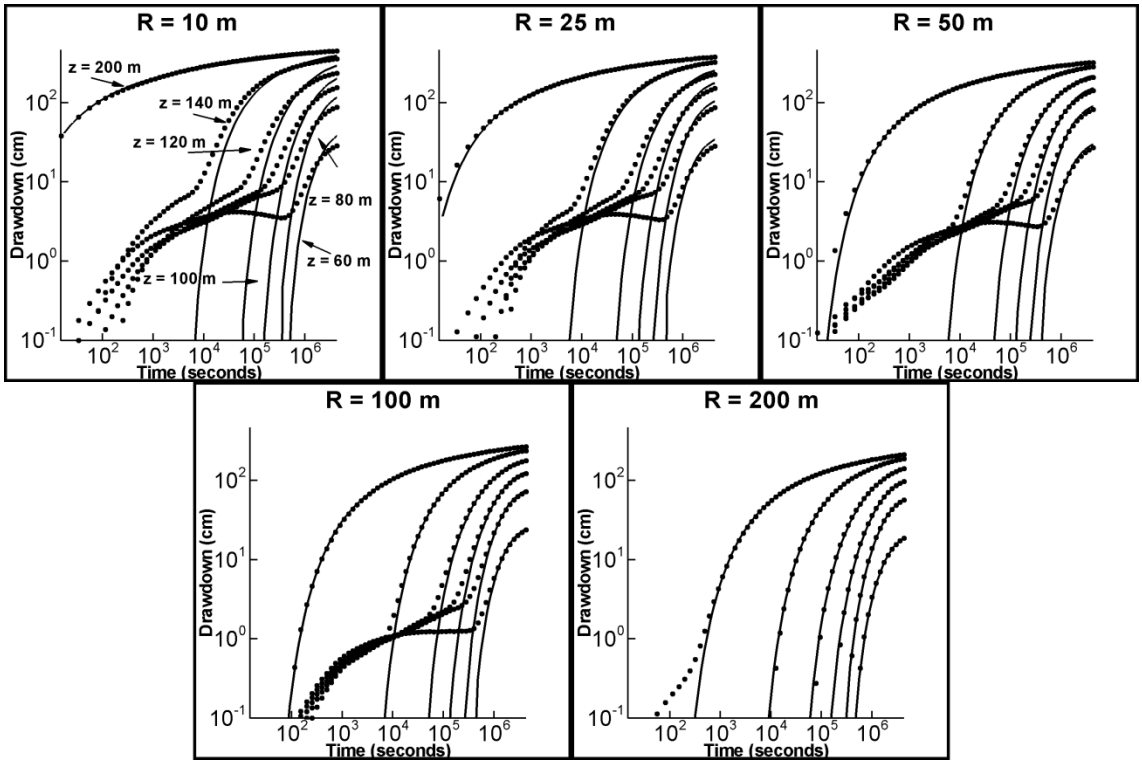


Figure E6c. Neuman-Witherspoon type-curve matches for synthetic data (dots) at varying radial distances for Case 3 with poroelastic effects. Solid lines represent the analytical solution.

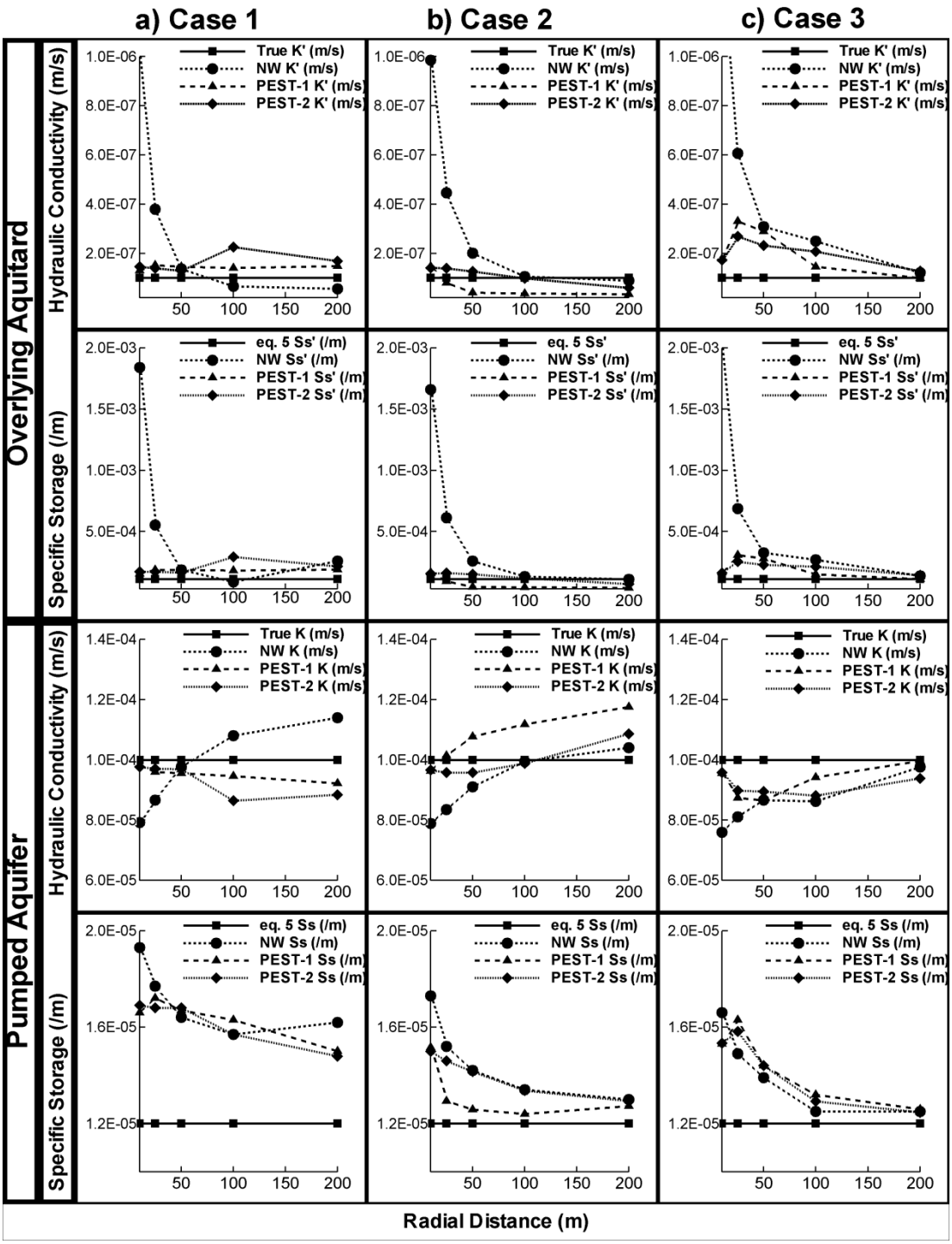


Figure E7. Estimated hydraulic parameters for Cases 1, 2 and 3 with poroelasticity effects, using data from varying radial distances. The solid line with squares represents the true value for K and K', and the value calculated from Equation 5 for S_s and S_s'. The dashed line with circles represents a

parameter estimated by the Neuman-Witherspoon (1969) type-curve method (NW). The dashed line with triangles represents a parameter estimated using PEST and including all synthetic data (PEST-1). The dotted line with diamonds represents a parameter estimated using PEST and excluding synthetic data that are influenced by deformation-induced effects (PEST-2).

Appendix F

Three-Dimensional Transient Hydraulic Tomography in a Highly Heterogeneous Glaciofluvial Aquifer-Aquitard

This manuscript by *Berg, S.J. and W.A. Illman* is under review.

EXECUTIVE SUMMARY

Hydraulic tomography has been proposed as an alternative site characterization method, however, relatively few field scale studies have been attempted. In this paper, we characterize the highly heterogeneous glaciofluvial aquifer-aquitard system at the North Campus Research Site, located at the University of Waterloo, Waterloo, Ontario, Canada using transient hydraulic tomography (THT). In particular, we performed 9 pumping tests in a network of wells to image the hydraulic conductivity (K) and specific storage (S_s) distributions (or tomograms) as well as their uncertainties in three-dimensions using the THT code of *Zhu and Yeh* [2005]. We first performed stochastic inverse modeling of the 9 pumping tests individually to gain insight into the level of detail that can be imaged. Then, we sequentially included 4 of the pumping tests in a THT analysis. The THT analyses were performed using 3 different types of *a priori* information: 1) effective K and S_s fields estimated from pumping test data; 2) 3-layer K and S_s fields with each layer being homogeneous based on parameter estimates from type-curve analysis using the *Hantush* [1960] solution; and 3) homogeneous K and S_s fields with the inverse model populated with local-scale K obtained via permeameter analysis of core samples along 5 boreholes. We find that all 3 cases produced very similar heterogeneity patterns for the main aquifer zone, however, the estimates were quite different for the upper and lower portions of the domain, where drawdown was minimal. The

resulting K and S_s tomograms were then validated visually by comparison to stratigraphy and permeameter K estimates, as well as through the simulation of all 9 pumping tests and comparing the simulated drawdown to the observed records.

F1. INTRODUCTION

All groundwater investigations require the estimation of hydraulic parameters such as hydraulic conductivity (K) and specific storage (S_s). However, in almost all investigations, K and S_s are commonly treated to be homogeneous despite the recognition that the aquifer is heterogeneous. *Wu et al*. [2005] recently found that the treatment of an aquifer to be homogeneous, when in reality it is heterogeneous, can lead to parameters estimates that are ambiguous or of questionable value. The conceptualization of an aquifer or an aquitard to be homogeneous and the problems associated with this has led to the development of various characterization methods of aquifer heterogeneity. However, the characterization of subsurface heterogeneity of hydraulic parameters is fraught with difficulties as numerous samples are required to delineate the variability of these parameters as well as their spatial correlations and connectivity.

Recent technological advances have led to increased efficiencies in drilling and *in situ* testing. Despite these recent advances, the traditional approaches for the characterization of subsurface heterogeneity in hydraulic parameters still requires a large number of boreholes which are cored or are tested *in situ*. For example, if a large number of core samples are obtained from boreholes, the samples have to be tested in the laboratory using a constant or a falling head permeameter to obtain high resolution records of K variability [e.g., *Sudicky*, 1986; *Sudicky et al*., 2010; *Alexander et al*., 2011]. Alternatively, samples obtained from cores are sieved to obtain grain size distributions, which can then be analyzed using various empirical relations that relate some characteristic grain size to K. A large number of small-scale K estimates have also been obtained

using flow meter tests [e.g., *Hufschmied*, 1986; *Molz et al*., 1989; *Boman et al*., 1997], steady-state dipole flow tests [e.g., *Kabala*, 1993; *Zlotnik et al*., 2001], slug tests [e.g., *Rehfeldt et al*., 1992; *McCarthy et al.*, 1996] and single-hole pumping or injection tests [e.g., *Chen et al*., 2000]. More recently, several researchers have developed direct push slug and permeameter test approaches [e.g., *Butler et al*., 2007; *Dietrich and Leven*, 2005; *Dietrich et al*., 2008 and *Liu et al*., 2009], which allow one to collect a large number of K measurements as the instrument is advanced into unconsolidated sediments. Because tests can be conducted at narrow intervals, high resolution characterization of vertical variations in K is possible. However, the application of direct push techniques at this time is impossible in consolidated media and severely limited in unconsolidated materials with pebble to cobble-sized particles.

Provided a sufficient number of samples are collected and analyzed accurately to obtain local-scale estimate of K, it may be possible to capture the salient features of heterogeneity in K without interpolation. However, this is often unachievable as collecting and analyzing the required number of samples is time consuming and expensive. For example, in order to confidently predict the transport of tracers in an alluvial aquifer at the Columbus research site, *Rehfeldt et al.* [1992] estimated that approximately 400,000 K measurements would be required. Clearly, this level of sampling is not practical or even possible. As such, various interpolation methods are often used to estimate the variability in K between boreholes [*de Marsily et al.*, 2005].

Recently, hydraulic and pneumatic tomography have been proposed as alternative site characterization methods to obtain information on heterogeneity in flow parameters between boreholes. The method relies on the inverse modeling of two or more hydraulic or pneumatic cross-hole pumping or injection tests conducted in geologic media. To our knowledge, *Neuman* [1987] first suggested this as an analogue to geophysical tomography. Since then, a number of inverse methods which analyze multiple pumping tests (either sequentially or simultaneously) have been developed

[e.g., *Gottlieb and Dietrich*, 1995; *Yeh and Liu*, 2000; *Vesselinov et al.*, 2001a-b; *Bohling et al.*, 2002; *Brauchler et al.*, 2003; *McDermott et al.*, 2003; *Zhu and Yeh*, 2005, 2006; *Li et al.*, 2005; *Fienen et al.*, 2008; *Ni and Yeh*, 2008; *Castagna and Bellin*, 2009; *Xiang et al.*, 2009; *Liu and Kitanidis*, 2011].

Initially, research on hydraulic tomography consisted largely of numerical studies [e.g., *Gottlieb and Dietrich*, 1995; *Yeh and Liu*, 2000], followed by laboratory sandbox studies [e.g., *Liu et al.*, 2002; *Illman et al.*, 2007; *Liu et al.*, 2007; *Yin and Illman*, 2009; *Illman et al.*, 2010] leading to much optimism in imaging heterogeneity between boreholes. Both the numerical and laboratory studies on hydraulic tomography were critical in validating the method under controlled conditions and laid the foundation for field applications of hydraulic tomography. Despite the encouraging results from synthetic studies and laboratory experiments, one could argue that these interpretations are based on data collected synthetically or in the laboratory under highly controlled conditions.

To date, relatively few field studies of hydraulic and pneumatic tomography have been published [e.g., *Vesselinov et al.*, 2001a-b; *Bohling et al.*, 2007; *Straface et al.*, 2007; *Li et al.*, 2008; *Cardiff et al.*, 2009; *Illman et al.*, 2009; *Brauchler et al.*, 2011]. In particular, *Vesselinov et al.* [2001a-b] conducted pneumatic tomography in unsaturated fractured tuffs at the Apache Leap Research Site (ALRS) in three-dimensions using 3 cross-hole pneumatic injection tests performed by *Illman and Neuman* [2001; 2003]. They used the pressure records from these tests to estimate equivalent permeability and porosity values, as well as their heterogeneous distributions. The results of the pneumatic tomography were compared to kriged permeability fields based on single-hole pneumatic injection tests [*Chen et al.*, 2000] and were found to share a similar internal structure, however, no other methods were employed to validate their results.

In a different study, *Bohling et al.* [2007] assessed steady shape hydraulic tomography [*Bohling et al.*, 2002] in an alluvial aquifer at the Geohydrologic Experimental and Monitoring Site (GEMS) of the Kansas Geological Survey. They analyzed a total of 23 pumping tests performed at

discrete intervals within 2 wells that were several meters apart. Between these 2 wells were 2 monitoring wells with 6 vertical observation points all aligned in a 2D plane, to record the drawdown responses. The tomographic analysis produced a 1D vertical profile of *K* between the 2 pumping wells which agreed reasonably with profiles obtained from an induced gradient tracer test, and direct push permeameter tests. *Bohling and Butler* [2010] followed up on their own study showing that heterogeneity in hydraulic parameters could not be obtained perpendicular to the planar configuration of wells in which they conducted the hydraulic tomography survey. That is, *Bohling and Butler* [2010] reported that heterogeneity in hydraulic parameters cannot be estimated accurately in areas where drawdown data or other information is not available.

Hydraulic tomography based on the stochastic or geostatistical inversion framework may be able to overcome some of the limitations cited by *Bohling and Butler* [2010] as it considers spatial correlations in the estimated parameters as well as cross-correlations between the parameter and heads. A field example of the use of a geostatistically based inverse model was first published by *Straface et al.* [2007] who analyzed 6 pumping tests performed sequentially within a 6-well network using the transient hydraulic tomography (THT) code developed by *Zhu and Yeh* [2005] to estimate the heterogeneous transmissivity (*T*) and storage coefficient (*S*) fields (or tomograms from now on) in two-dimensions. Despite the small number of wells, they concluded that the *T* and *S* tomograms were reasonable representations of the aquifer based on the geological setting of the site, however, no attempts were made to validate the tomograms.

At the Krauthausen test site in Germany, *Li et al.* [2008] employed a geostatistical inverse approach to jointly analyze steady-state drawdown and borehole flowmeter data from multiple pumping tests to estimate the *K* distribution in three-dimensions. They found that jointly inverting both steady-state drawdown and flow meter data produced an improved 3D structure when compared to just inverting pumping test data.

The first steady state hydraulic tomography in unconfined aquifers was performed by *Cardiff et al.* [2009] using 9 pumping tests at the Boise Hydrogeophysical Research Site (BHRS) to estimate the distribution of depth-averaged *K*. They found that the *K* tomogram contained expected geological features, additionally, the uncertainty bounds on the estimation indicate that *K* was well constrained within the central portion of the research site where the pumping and observation well network is located.

A large-scale application of THT in fractured rock was demonstrated by *Illman et al.* [2009] at the Mizunami Underground Research site in Japan. Using 2 cross-hole pumping tests, they estimated the 3D distribution of *K* and S_s as well as their uncertainties. This was the first application of three-dimensional hydraulic tomography in the field which used transient drawdown data. Several continuous high *K* and low S_s zones were identified and interpreted as possible fault zones. This field investigation highlighted the potential use of hydraulic tomography in fractured rock environments to identify hydraulic connections between boreholes. However, the evaluation of the *K* and S_s tomograms was limited to available fault data, several drawdown data from one of the cross-hole pumping tests, and coseismic responses in wells. That is, the validity of *K* and S_s tomograms was not confirmed rigorously through the prediction of independent drawdown inducing events such as shown previously by *Illman et al*. [2007, 2008, 2010], *Liu et al*. [2007], and *Berg and Illman* [in review] through sandbox studies. Therefore, our review shows that there have been several field studies that illustrate the utility of hydraulic and pneumatic tomography, but a comprehensive study designed to validate hydraulic tomography results in three-dimensions using transient drawdown data is still lacking.

In order to further test the effectiveness of THT, we conducted 9 pumping tests in a three-dimensional network of pumping and observation wells completed in a highly heterogeneous glaciofluvial aquifer-aquitard system at a field site on the University of Waterloo campus. We

analyzed the transient drawdown data using the THT code of [*Zhu and Yeh*, 2005] to estimate K and S_s tomograms as well as their uncertainties in three-dimensions. We first performed stochastic inverse modeling of each pumping test individually to examine the effectiveness of such inversions. This is then followed by a sequential inversion of 4 pumping tests, which amounts to THT. For the THT, we examined the influence of different K and S_s values input to the models, as well as the inclusion of local-scale K data from a selected number of boreholes. The K and S_s tomograms were first evaluated qualitatively through comparison to site geology. The tomograms generated from these various cases were also quantitatively validated through direct comparison to local-scale K data from selected boreholes. Finally, we used the tomograms for forward simulations of 9 pumping tests of which 5 were not used in the construction of the tomograms. We assessed the results by comparing the simulated and observed drawdown responses for the drawdown data used for calibration and validation.

F2. EXPERIMENTAL DESIGN

F2.1 Site description

The study was conducted at the North Campus Research Site (NCRS) located at the University of Waterloo (UW) in Waterloo, Ontario, Canada. The surficial geology of the local area is glaciofluvial in origin and is highly heterogeneous. The field site is located within the Waterloo Moraine, which is an interlobate feature composed of kettle and kame deposits containing alternating layers of till and glaciofluvial material [*Karrow*, 1993]. Site specific geology has been described by *Sebol* [2000] and *Alexander et al.* [2011]. The main feature of the site is an 'aquifer zone' located approximately 8 to 13 m below ground surface (mbgs). This zone consists of 2 high K units that are separated by a discontinuous low K layer. The upper aquifer is composed of sand to sandy silt, and the lower aquifer is composed of sandy gravel. The low K unit separating the 2 aquifers is discontinuous and is known

to provide hydraulic connection [*Alexander et al.*, 2011]. In addition, despite being interpreted as continuous layers, none of the units extend across the entire study site. Situated above and below the aquifer zone are low *K* silts and clays. At approximately 18 mbgs is the dense Catfish Creek Till which acts as a hydraulic barrier [*Alexander et al.*, 2011] and is taken to represent the bottom boundary for this study.

The aquifer system in the Waterloo Moraine is generally confined by a laterally extensive upper aquitard layer, however, this aquitard is known to contain stratigraphic windows in some areas [*Martin and Frind*, 1998]. Based on previous pumping tests performed at the site [*Alexander et al.*, 2011], the aquifer at the NCRS behaves as a confined to semi-confined system. None of the drawdown responses observed during previous pumping tests suggest that the main aquifer zone behaves in an unconfined manner, which might indicate the presence of a stratigraphic window. Water levels collected in the vicinity of the site indicate that groundwater flow is towards the south-east.

Alexander et al. [2011] performed a detailed description of soil cores collected during the installation of 5 wells at the NCRS. In addition, they performed 471 permeameter tests and 270 grain size analyses to produce a detailed *K* profile along these boreholes. Furthermore, slug and pumping tests were performed to estimate *K* along these boreholes. Figure F1 shows *K* estimates along 5 boreholes from grain size analyses, permeameter analyses of core samples, slug tests at CMT ports, and from pumping tests in which the medium was treated to be homogeneous. This figure highlights the considerable heterogeneity at the site, with large changes in *K* over very short vertical distances. The estimated vertical correlation length was 15 cm, and *K* ranged over 5 orders of magnitude [*Alexander et al.*, 2011]. We also note that *K* estimates from slug and pumping tests are in general higher than those estimated through permeameter tests conducted in the lab. Likewise, *K* estimates from the permeameter analyses are in general higher than the grain size estimates.

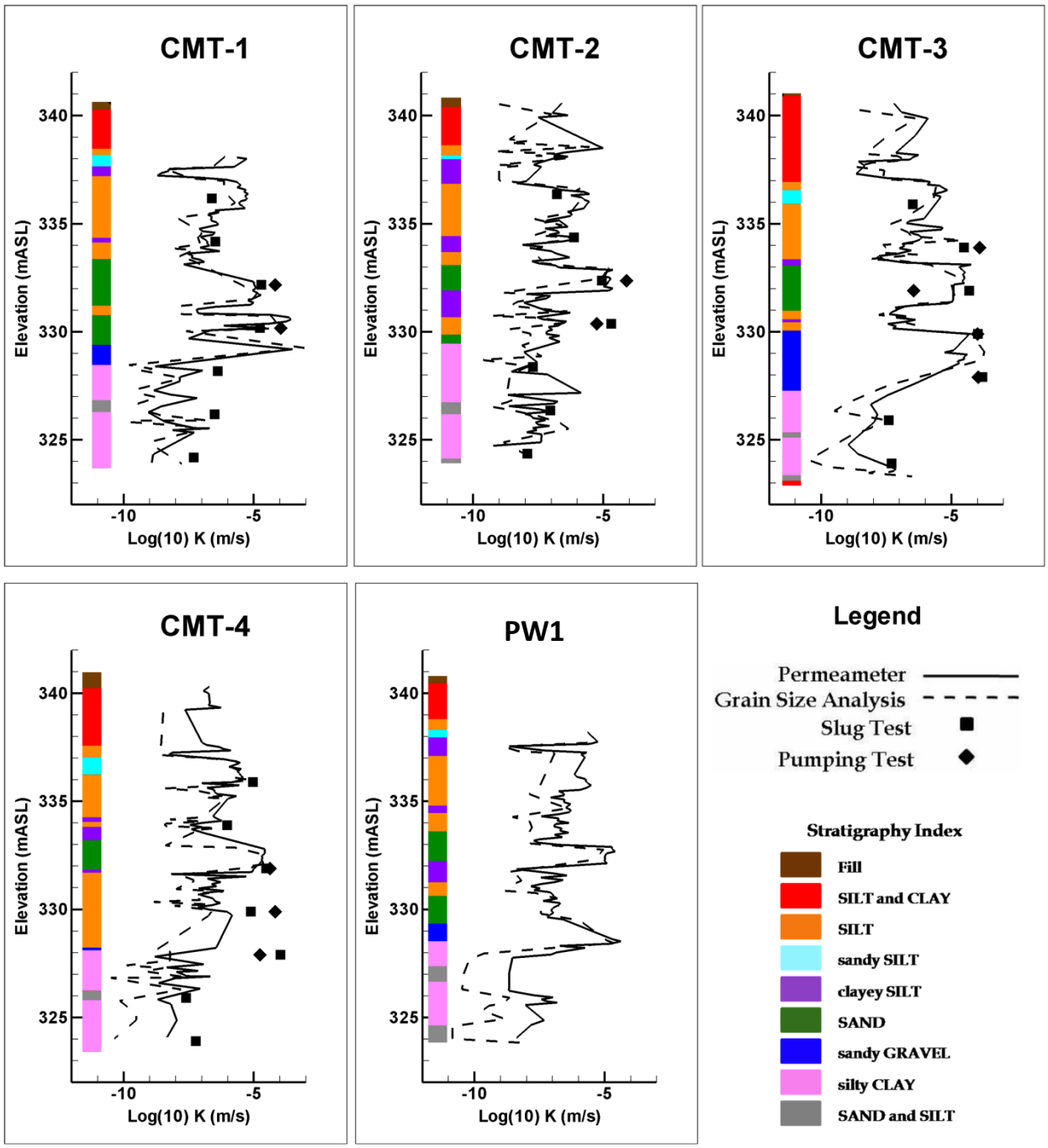


Figure F1: *K* estimates along 5 boreholes at the NCRS [after *Alexander et al.*, 2011].

F2.2 Well installation and site design

The NCRS was instrumented with a total 9 wells in a square pattern (1 at each corner, 1 at the center of each face, and 1 in the center) measuring 15 m × 15 m. Figure F2 is a schematic layout of the wells. Continuous multichannel tubing (CMT) wells (*Solinst Canada Ltd.*), containing 7 channels each (7 screened intervals), were used strictly as observation wells and were installed in between the 4 corners of square pattern (Figure F2). The screened intervals of the CMT wells were spaced 2-m apart with the upper screens located between 4.5 to 5.5 mbgs, and the deepest ports were 16.5 to 17.5 mbgs. The remaining 5 wells were pumping wells (PW), 3 of which were multi-screen wells (PW1 contained 8 screens, PW3 and PW5 contained 5 screens). These multi-screen wells were 10 cm in diameter, and contained screens 1 m in length. Each screen was separated from adjacent screens by 1 m of solid pipe. PW1 extended to approximately 18 mbgs; and, PW3 and PW5 to 12 mbgs. PW2 and PW4 were well nests consisting of 3 separate wells each (5 cm diameter). Each well in the nest had one screen that was 1-m long. Screen elevations for PW2 were 4, 7, and 8 mbgs; and, screen elevations for PW4 were 5, 8.5, and 11.5 mbgs.

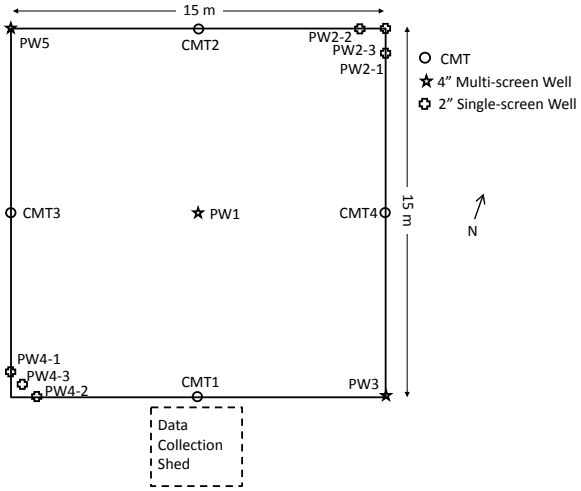


Figure F2: Two-dimensional plan view showing well locations.

Since these wells were installed for the purposes of hydraulic tomography, it was important that adjacent screens be hydraulically isolated. As such, the sand packs around the well screens were separated from each other by a layer of bentonite. The exceptions to this were the well nests (PW2 and PW4) which were installed in direct contact with the native formation. Figure F3 shows the three-dimensional perspective view of the wells, the various screens as well as bentonite seal elevations at the NCRS.

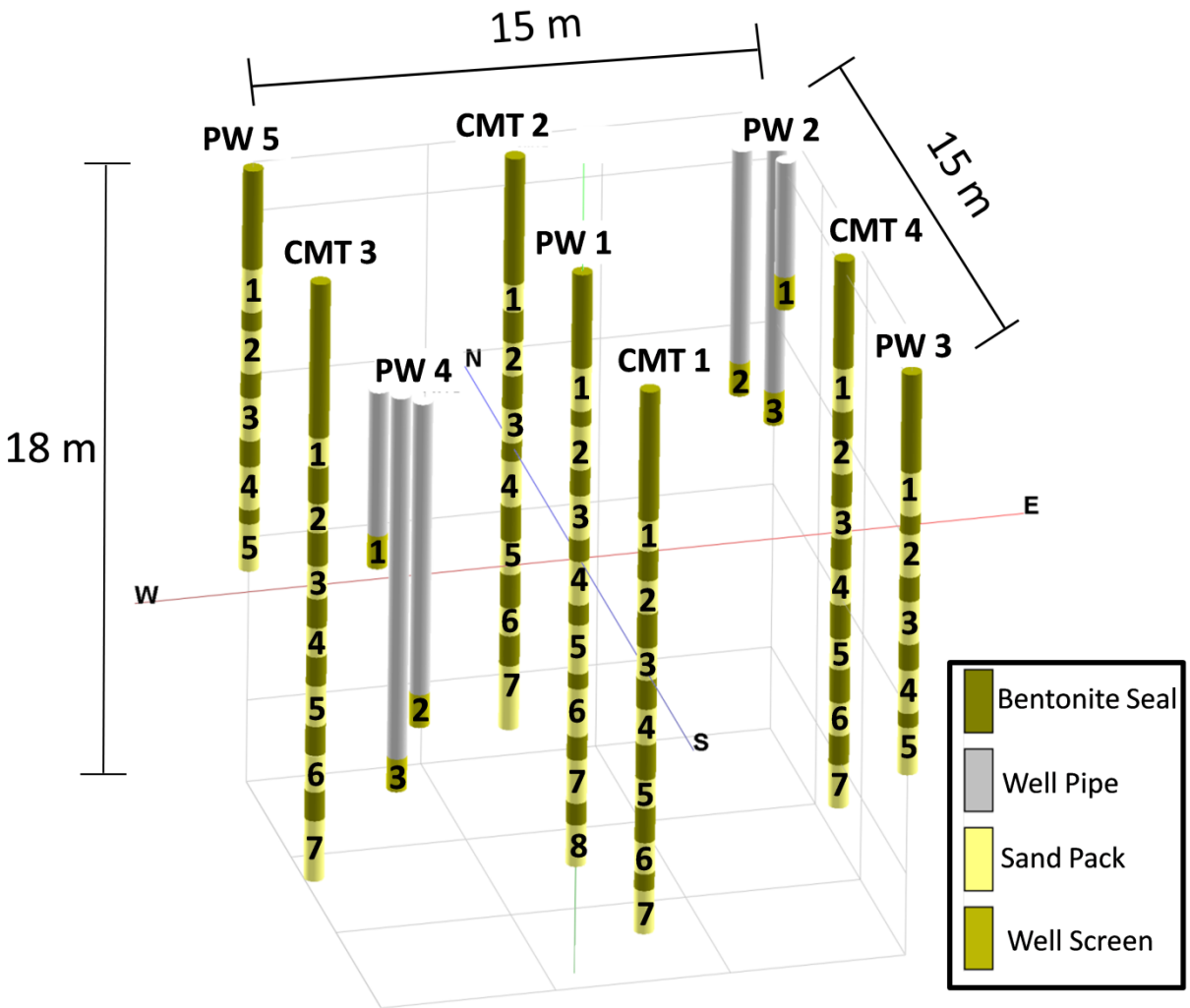


Figure F3: Three-dimensional perspective view of various wells and corresponding intervals at the NCRS.

F2.3. Sensors and Data Collection

A network of pressure transducers was installed at the NCRS to allow for data collection during pumping tests. Depending on the particular pumping test, up to 44 observation ports were monitored. The CMT ports were monitored with 0 - 15 psig (model MP100: Micron Systems) pressure transducers. Since the center channel of the CMT wells was too small to allow for both a pressure transducer and a water level tape (for verification purposes), these central ports were monitored manually with a water level tape. These ports were found to respond very slowly, thus, it was still possible to obtain a complete record for these ports. The well nests (PW2 and PW4) were monitored with 0 - 5 or 0 - 10 psig pressure transducers (model 3001 LT Leveloggers Junior, *Solinst Canada Ltd.*).

If one of the multi-screen wells (PW1, PW3, or PW5) was not being used for pumping, a FLUTe (*FLUTe Ltd.*) liner was installed in the well to prevent hydraulic short circuiting between adjacent screens within the well. Two of these liners contained 5 vented pressure transducers (Level Troll, *In Situ Inc.*), at elevations that corresponded with the screens of the PW wells. PW1, PW3, and PW5 were similar enough in construction to allow for the liners to be moved between wells depending on the pumping configuration. When PW1, PW3, or PW5 were used for pumping, the FLUTe systems were installed in the 2 unpumped wells. When PW2 or PW4 was pumped, the FLUTe systems were installed in PW3 and PW5, and a blank FLUTe liner was installed in PW1.

F2.4. Pumping test description

A total of 9 pumping tests were performed in the well network at the NCRS. These tests ranged in duration from 6.5 hrs to 22 hrs. Table 1 provides a summary of the pumping tests. For the multi-screen wells (PW1, PW3, and PW5) pumping was performed with a submersible pump (Model SQE05, *Grundfos Canada Inc.*) located between two packers that were inflated to isolate the target screen. Pumping in PW4 was performed using a surface lift pump.

The locations for the pumping tests were selected for their ability to stress the aquifer system (i.e. flow rate), as well as their spatial coverage of the site. In particular, emphasis was placed on pumping at different intervals from different boreholes so as to maximize our opportunity for stressing the aquifer-aquitard system from various locations. The pumped locations were; PW1-3, PW1-4, PW1-5, PW3-3, PW3-4, PW4-3, PW5-3, PW5-4 and PW5-5. PW4-3 had the highest flow rate at 30.2 L/min, followed by PW1-3 at 10.5 L/min. PW3-3 and PW3-4 had the lowest flow rates at 2.1 and 1.5 L/min respectively. The flow rates for the 5 remaining pumping tests were between 4.4 and 8.8 L/min.

Table F1: Summary of pumping tests performed at the NCRS.

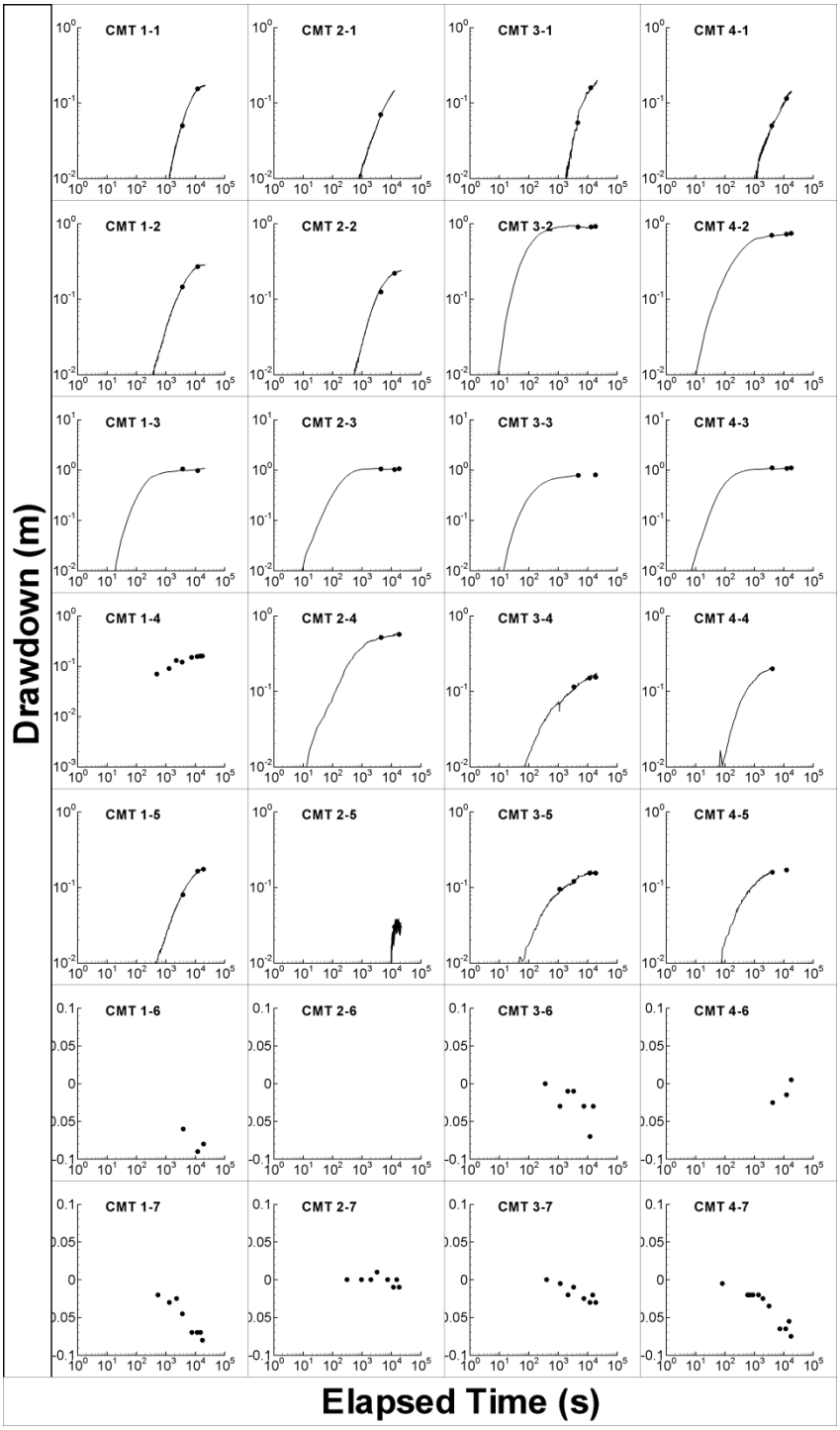
Pumped Location	Flow Rate (L/min)	Duration (hrs)	Maximum Observed Drawdown (m)	Port of Maximum Drawdown
PW1-3	10.5	6	1.10	CMT 4-3
PW1-4	6.3	8.5	0.98	CMT 2-4
PW1-5	4.4	22.5	0.18	CMT 2-4
PW3-3	2.1	22	0.57	CMT 1-3
PW3-4	1.5	22	0.08	CMT 1-5
PW4-3	30.2	22.5	0.88	CMT 3-4
PW5-3	7.8	22	0.92	CMT 2-3
PW5-4	7.8	8.5	0.45	CMT2-4
PW5-5	8.1	22	0.32	CMT2-3

For most of the pumping tests, the pumped interval was hydraulically isolated, as indicated by pressure transducer readings from above and below the packed interval. Two exceptions to this were, PW1-5 and PW5-5, both of which experienced pronounced drawdown above the packed-off interval. For these tests, the packers were operating properly, thus the short circuiting occurred either through the bentonite seal between the sand packs, or directly through the aquifer.

Figure F4a shows the drawdown response recorded at the CMT ports while Figure F4b shows the same but for the PW ports during the pumping test at PW1-3. Drawdown curves for the other 8 tests are included in the Supplementary Information section as Figures SF1-SF8 (For this appendix the supplementary figures are provided at the end of the appendix). Solid lines represent transducer recorded data, and the dots are water levels manually recorded with a water level tape. During this test, the drawdown response was most pronounced at ports CMT2-3, CMT3-3, and CMT4-3. The two deepest ports in the CMT wells saw very little to no drawdown. For these ports, sensor noise was evident in the recorded pressure signal. To remove the noise, we applied a 10-point centrally weight moving average. Since the data was collected at a high frequency (~ 4 hz at early time, 1 hz at late time) the application of this filter did not change the shape of the drawdown curve.

Examination of Figures F4a and F4b reveals that some ports recorded negative drawdown during this pumping test at early time. We observed this negative drawdown in deep ports completed in the silty clay unit just above the Catfish Creek Till during this and other pumping tests at the NCRS and attribute it to deformation-induced head change. The findings of *Berg et al.* [2011] indicate that when this behaviour manifests in pumping test data, it is still possible to estimate hydraulic parameters if late time data (after the water level draws down below static) is analyzed. As such, we include late time drawdown data for these ports if available, if not, the negative drawdown is treated as zero drawdown.

a)


Drawdown (m)
CMT 1-1
CMT 2-1
CMT 3-1
CMT 4-1
CMT 1-2
CMT 2-2
CMT 3-2
CMT 4-2
CMT 1-3
CMT 2-3
CMT 3-3
CMT 4-3
CMT 1-4
CMT 2-4
CMT 3-4
CMT 4-4
CMT 1-5
CMT 2-5
CMT 3-5
CMT 4-5
CMT 1-6
CMT 2-6
CMT 3-6
CMT 4-6
CMT 1-7
CMT 2-7
CMT 3-7
CMT 4-7
Elapsed Time (s)

b)

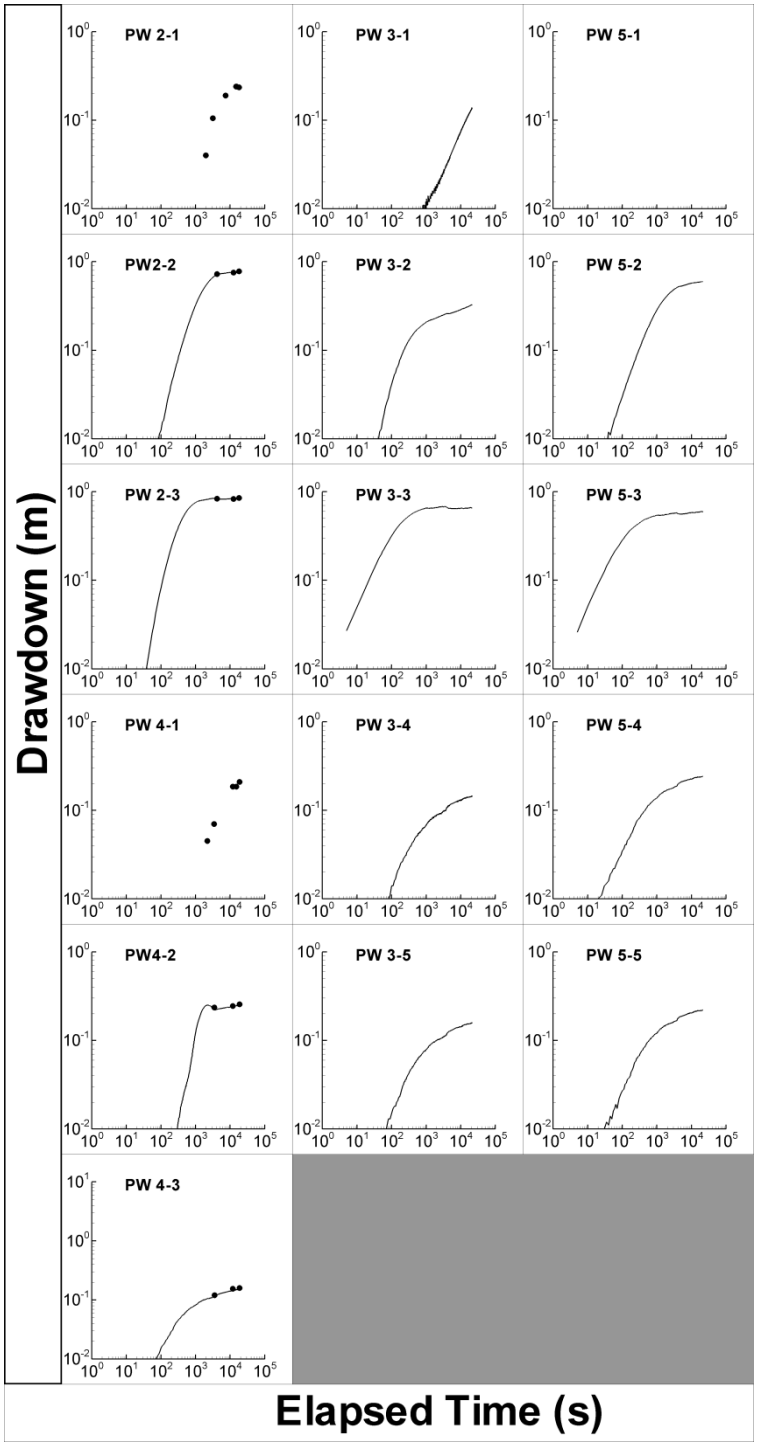


Figure F4: Observed drawdown in a) the CMT ports and b) the PW ports during the cross-hole pumping test at PW1-3. The solid line represents transducer recorded data, and the dots are manual water level measurements collected with a water level tape.

F3. INVERSE MODELING

We performed stochastic inverse modeling of individual pumping tests and THT using a code based on the Sequential Successive Linear Estimator (SSLE) developed by *Zhu and Yeh* [2005]. SSLE uses the successive linear estimator (SLE) concept and successively includes data sets, thus reducing the computational burden that would be encountered if all of the data sets were included simultaneously. The SLE approach is similar to cokriging [*Yeh et al.*, 1995a] which seeks mean parameter fields using point data and geologic/hydrogeologic structures (i.e. the spatial covariance and cross-variance functions of parameters and hydraulic heads) to condition the estimates. SLE is different from cokriging in that it successively updates both the conditional means and covariances of the estimates using a linear estimator based on differences between observed and simulated heads. In this way the nonlinear relation between the information and the parameters is considered [*Zhu and Yeh,* 2005].

Since HT experiments generate large data sets, classical inverse parameter estimate approaches may suffer from significant computational burden, numerical instabilities [*Yeh*, 1986; *Hughson and Yeh,* 2000] and even non-uniqueness. To efficiently handle these large data sets, *Zhu and Yeh*[2005] developed the SSLE which incorporates observation data from individual pumping tests sequentially, as opposed to analyzing all of the data simultaneously, thus reducing the computational load. Additionally, the issue of non-uniqueness is handled by treating the hydraulic parameter fields as spatial stochastic processes. The SSLE then seeks the mean parameter distributions which are conditioned on the observations from the pumping tests and any available 'hard' data (direct hydraulic parameter estimates from slug tests, permeameter tests, grain size analysis, etc.) [*Yeh and Liu*, 2000; *Zhu and Yeh*, 2005].

After the SSLE reaches the convergence criteria (number of iterations or the change in the estimated parameters is below a set threshold) for a specific pumping test, the next data set is included. This is continued until all of the data sets have been included. While the simultaneous

inversion of pumping tests is also possible with *Zhu and Yeh's* [2005] THT code extended by [*Xiang et al.*, 2009], the computational requirements are still significant, as such, we choose to sequentially analyze the pumping tests using the SSLE of *Zhu and Yeh* [2005].

Including tests in the SSLE sequentially, as oppose to analyzing the tests simultaneously means that the final inversion results many not honour the data included from the first test. To address this issue, *Zhu and Yeh* [2005] introduced a looped iteration scheme that cycles through the tests, thus revisiting the data from earlier tests periodically throughout the inversion. For example, for a 4-test inversion, the *Zhu and Yeh* [2005] looping scheme would analyze the tests in the following order: 1, 1, 2, 1, 2, 3, 1, 2, 3, 4. For this case we looped through the tests sequentially: 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4. In this way each test is included in the inversion the same number of times.

F3.1 Inverse Model Setup

The model domain used for all inversions and forward simulations was composed of variably-sized rectangular elements and measures 45 m by 45 m by 15 m. The smallest elements, measuring 0.5 m by 0.5 m by 0.5 m were located in the central portion of the model domain and correspond to the location of the 15 m x 15 m test plot. Outside of the test plot, the element size increased with increasing distance from the center of the model domain. In total, the model domain consists of 29,791 elements (31 x 31 x 31) and 32,768 nodes (32 x 32 x 32). The boundary conditions were no flow for top and bottom boundaries and constant head for the remaining outer boundaries. Figure F5 shows the model domain.

All stochastic inversions of individual pumping tests, and all three THT cases were executed using between 8 and 40 processors on a PC-cluster consisting (of 1 master and 12 slaves each with Intel Q6600 Quad Core CPUs running at 2.4 GHz with 16 GB of RAM per slave) at the University of Waterloo. The operating system managing the cluster was CentOS 5.3 based on a 64-bit system. The

total computational time for inverting data ranged from approximately 24 - 48 hrs for the stochastic inversion of one test to more than a week for THT using 4 tests.

Estimates of the hydrogeologic structure of the aquifer (correlation lengths and variance) were provided as inputs to the model. During some of the preliminary stochastic inversions of single tests, and the THT inversions, the variance was found to rapidly approach a value of 4 to 5 (these inversions are discussed further in sections 4 and 5 respectively). Since estimates of variance have been shown to have a negligible effect on the resulting tomograms [*Yeh and Liu*, 2000] a value of 5 was assigned as an initial value.

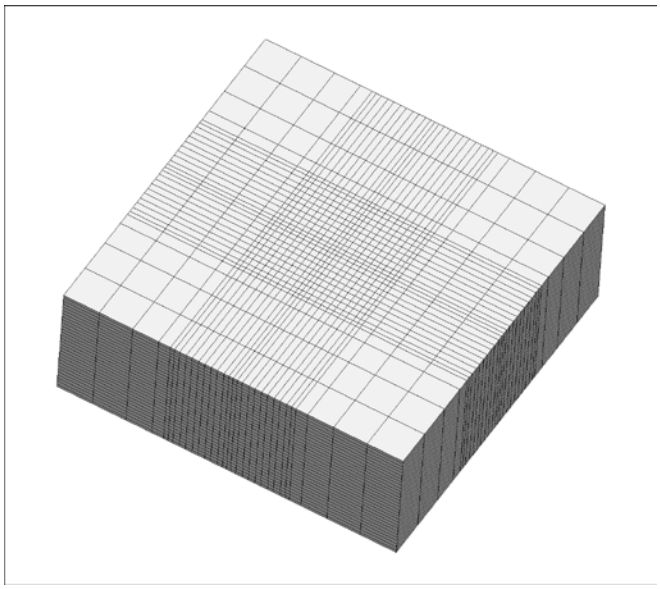


Figure F5: Finite element mesh used for simulations.

Based on the 1-D geostatistical analysis performed by *Alexander* [2009] along the 5 boreholes, the average vertical correlation length (λ_z) of K at the NCRS was estimated to be 0.15 m, while an omnidirectional variogram fit to all 471 data yielded a λ of 4.0 m. Due to the nature of the data set (discrete vertical profiles) and the availability of only 5 boreholes during the initial phase of

the field work at the NCRS, *Alexander et al.* [2011] was unable to obtain reliable correlation lengths in the horizontal plane (λ_x and λ_y) for K.

Additionally, based on the examination of stratigraphy from the 5 boreholes, *Alexander et al.* [2011] indicated that many of the units do not appear continuous across the site. Based on these factors, we selected an initial horizontal correlation length of 4 m (approximately half the distance between the central well and the edge of the plot). Because of the lack of information on geostatistical parameters for S_s, similar values were assigned for S_s. Several values were tried initially, and a value of $\lambda_x = \lambda_y = 4$ m produced tomograms that appeared to be most consistent with our knowledge of site geology. The average vertical correlation length determined by *Alexander et al.* [2011] is less than the thickness of an element block, as such, the vertical correlation length in the model is estimated as the thickness of an element block (0.5 m). This is tantamount to ignoring the fine scale heterogeneity, however, our decision was based on the limitations of computational resources available to us.

F4. STOCHASTIC INVERSE MODELING OF INDIVIDUAL PUMPING TESTS Each pumping test was analyzed individually to produce heterogeneous K and S_s fields. Approximately 5 data points from each drawdown curve at each observation port were included for each test. The total number of data points used ranged from 151 for PW3-4 to 233 for PW4-3. These data points were selected to define the shape of the curve (early, middle, and late time). During some of the pumping tests, some ports did not respond until very late into the test, or not at all. For these cases, zero drawdown was input into the inverse model.

Estimates of K and S_s to populate the initial homogeneous field were obtained by coupling PEST [*Doherty*, 2005] with HydroGeoSphere [*Therrien* et al., 2005] and individually matching the drawdown responses at each observation port for pumping tests performed at PW1-3. The model domain used for the parameter estimation was the same as the one described in section 3. The

geometric means of individual K and S_s estimates were rounded and then incorporated into the starting model ($K = 8.00 \times 10^{-6}$ m/s and $S_s = 1.00 \times 10^{-4}$ m^{-1}).

The inverse model ran for each pumping test listed in Table 1 until the convergence criteria were met (either maximum number of iterations, or no significant improvement in the difference between simulated and observed drawdown was attained). Figures F6a and F6c show the estimated K- and S_s-fields generated by inverting the pumping test performed at PW1-3 respectively. Figures F6b and F6d show the corresponding variance estimates of the ln K and ln S_s fields. Similar figures for the remaining 8 pumping tests are available online in the supplementary information section as Figures SF9-SF16.

Examination of Figures F6a and F6c reveal that, there is a significant amount of heterogeneity evident in the K distribution, relative to the estimated S_s field. All single test inversions displayed a similar tendency in which more heterogeneity being evident in the K distribution. The heterogeneity in K is generally thought to vary more than S_s, a tendency that has also been documented elsewhere [e.g, *Vesselinov et al.*, 2001a-b; *Liu et al.*, 2007]. Also, examination of Figures F6b and F6d reveal that the variance in the estimates of ln K are consistently lower than those for ln S_s, indicating that there is greater confidence in the estimates of ln K. The variance for both ln K and ln S_s is lower in the upper half of the model domain and higher in the lower half. This lower variance zone corresponds roughly with the aquifer zone where most of the drawdown occurs. The bottom portion of the aquifer experiences very little drawdown, even at late time, thus there is little information in this region to constrain the model.

The stochastic inversions for all 9 pumping tests are compared by analyzing their statistical properties, including geometric mean, variance, and the variogram parameters. Geometric mean and variance are calculated from the raw data set without considering spatial location. SGeMS [*Remy et al.*, 2008] was used to generate directional (x, y, and z) experimental variograms for ln K and ln S_s to

which either exponential or Gaussian variograms were fit manually. Many of the vertical variograms exhibited the 'hole effect' which is a result of the layered nature of the estimated fields. The results of these fits for ln K and ln S_s, for each pumping test are presented in Tables 2a and 2b. Examination of Tables 2a and 2b reveals that the geometric mean and variance estimates from the raw data as well as the variogram fitting parameters all reveal that the parameters are variable from one stochastic inverse model result to the next, suggesting that the inverse modeling of one test may not be sufficient to estimate these parameters accurately. However, we find that for both K and S_s, the lengths of the horizontal correlations are approximately double the vertical correlation lengths. The average horizontal correlation lengths of K and S_s from the 9 individual inversions were 7.5 m and 14.6 m. On the other hand, the vertical correlation lengths of K and S_s were 3.8 m and 7 m., respectively.

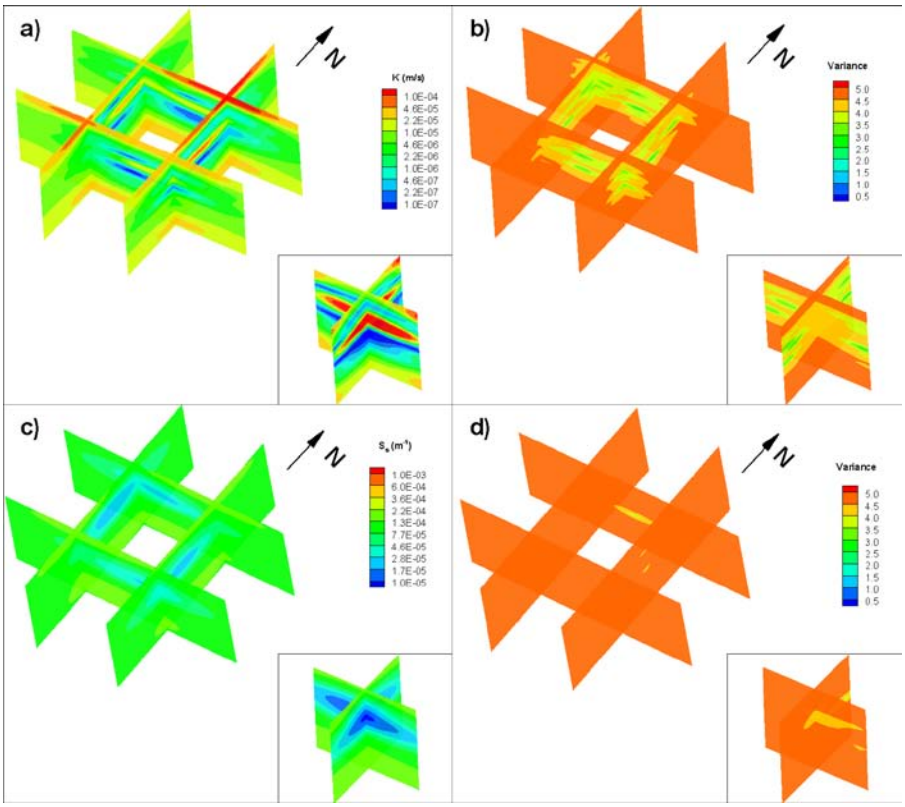


Figure F6: K and S_s fields from the stochastic inversion of a pumping test performed at PW1-3: a) K-field; b) variance associated with the estimated K field; c) S_s field; and d) variance associated with the

estimated S_s field. Note that the square formed by the slices in the main figure corresponds to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W). Similar figures for the remaining 8 pumping tests are available online in the supplementary information section as Figures SF9-SF16.

Table F2a: Geometric mean, variance, and correlation lengths of ln K for the stochastic inversion of individual pumping tests.

	ln K (m/s)						
	Geometric Mean*	**Variance***	**Sill**	λ_x	λ_y	λ_z	**Model**
PW1-3	5.1×10^{-6}	2.9	3.7^{xy}, 5.2^{z}	6.3	6.3	3.2	Exponential
PW1-4	6.4×10^{-6}	1.4	1.6^{xy}, 2.3^{z}	9.5	9.5	5.4	Exponential
PW1-5	8.1×10^{-6}	3.3	5.0	8.6	8.1	3.6	Exponential
PW3-3	4.0×10^{-6}	2.5	3.2^{xy}, 5^{z}	8.1	8.1	5	Exponential
PW3-4	9.5×10^{-6}	3.0	3.5^{xy}, 4.5^{z}	5.0	5.0	3.2	Exponential
PW4-3	9.8×10^{-6}	2.9	3.9^{xy}, 4.5^{z}	6.8	6.8	2.2	Exponential
PW5-3	6.0×10^{-6}	1.1	1.5^{xy}, 2.0^{z}	9.0	9.0	3.6	Exponential
PW5-4	9.4×10^{-6}	2.1	3^{xy}, 5.0^{z}	7.2	7.2	5.0	Exponential
PW5-5	8.2×10^{-6}	4.2	6.0	7.2	7.2	3.6	Exponential

*Calculated from raw data

^{xyz} direction variance was measured in

Table F2b. Geometric mean, variance, and correlation lengths of ln S_s for the stochastic inversion of individual pumping tests.

	l n S_s (m^{-1})						
	Geometric Mean*	**Variance***	**Sill**	λ_x	λ_y	λ_z	**Model**
PW1-3	8.6×10^{-5}	0.4	0.4^{xy}, 0.9^{z}	8.6	8.6	6.3	Gaussian
PW1-4	8.7×10^{-5}	0.3	0.4^{xy}, 1.1^{z}	9.9	9.5	5.9	Gaussian
PW1-5	2.0×10^{-5}	0.3	0.4	26.1	21.2	9	Exponential
PW3-3	9.4×10^{-5}	0.1	0.1^{xy}, 0.2^{z}	8.6	8.6	5.0	Gaussian
PW3-4	1.8×10^{-4}	0.3	0.4^{xy}, 0.2^{z}	23.0	15.8	11.7	Gaussian
PW4-3	1.8×10^{-4}	0.2	0.2^{xy}, 0.1^{z}	25.2	25.2	5.4	Exponential
PW5-3	9.8×10^{-5}	0.04	0.5^{xy}, 0.1^{z}	7.2	7.2	4.1	Exponential
PW5-4	9.4×10^{-5}	0.2	0.3^{xy}, 0.7^{z}	10.4	10.4	6.8	Gaussian
PW5-5	1.4×10^{-4}	0.2	0.2^{xy}, 0.5^{z}	12.2	12.2	9.5	Gaussian

*Calculated from raw data
[xyz] direction variance was measured in

F5. TRANSIENT HYDRAULIC TOMOGRAPHY (THT) ANALYSIS OF MULTIPLE PUMPING TESTS

We next performed THT in 3 different ways, with each case (1 – 3) incorporated different amounts of information regarding the site for the initial K and S_s fields. For case 1, the inversion started with homogeneous K and S_s fields, and did not include any information on site heterogeneity. For case 2, the model started as a simple 3-layer system using values obtained from the type curve analysis of three pumping tests (PW1-3, PW1-4, and PW1-5) using the solution by *Hantush* [1960]. The initial K field in case 3 is populated with K estimates from permeameter data collected from the 5 boreholes presented earlier (see Figure F1). Due to the significant computational effort of running the THT inversions for their full duration of the pumping tests (6.5 to 24 hrs), only the first hour of each pumping test was included in the inversion. During the first hour of each test, many ports show a strong response, and some even reach steady state. For each drawdown curve, 3 to 5 points were selected to define the curve (between 112 and 130 data points were included for each test). Unresponsive ports were also included in the inversion as they provide information on connectivity between the pumped and observation ports.

For each case, the tests with pumping taking place at intervals PW4-3, PW1-3, PW5-3, and PW3-3 were included in the inversion (in the order listed). These tests and the order they were included in the inverse model were based on the degree to which they stressed the aquifer (PW4-3 and PW1-3 had the highest flow rates of the 9 pumping tests) and their spatial coverage of the site (PW5-3 and PW3-3). This is consistent with the recommendations of *Illman et al.* [2008]. After all 4 pumping tests were included in the inversion, we implemented the looping scheme that we described

earlier and the simulations were run until the convergence criteria were met Convergence criteria are met when either the change in the estimated fields is below a minimum threshold (e.g. variance changes less than 0.01) or the maximum number of iterations has been achieved. The model domain for these THT inversions is identical to that described for the stochastic inversion of individual tests in section 4.

F5.1 Case 1 – THT with homogeneous K and S_s fields as starting model

For this case, the K and S_s fields were initially homogeneous (same starting values as those used for the stochastic inversion of individual tests) and no local-scale values of K and S_s were included at the start of the inversion. Figures F7a and F7c show the final K and S_s tomograms, respectively, after including the drawdown data from the 4 pumping tests (PW4-3, PW1-3, PW5-3, and PW3-3), while Figures F7b and F7d show the associated variance maps.

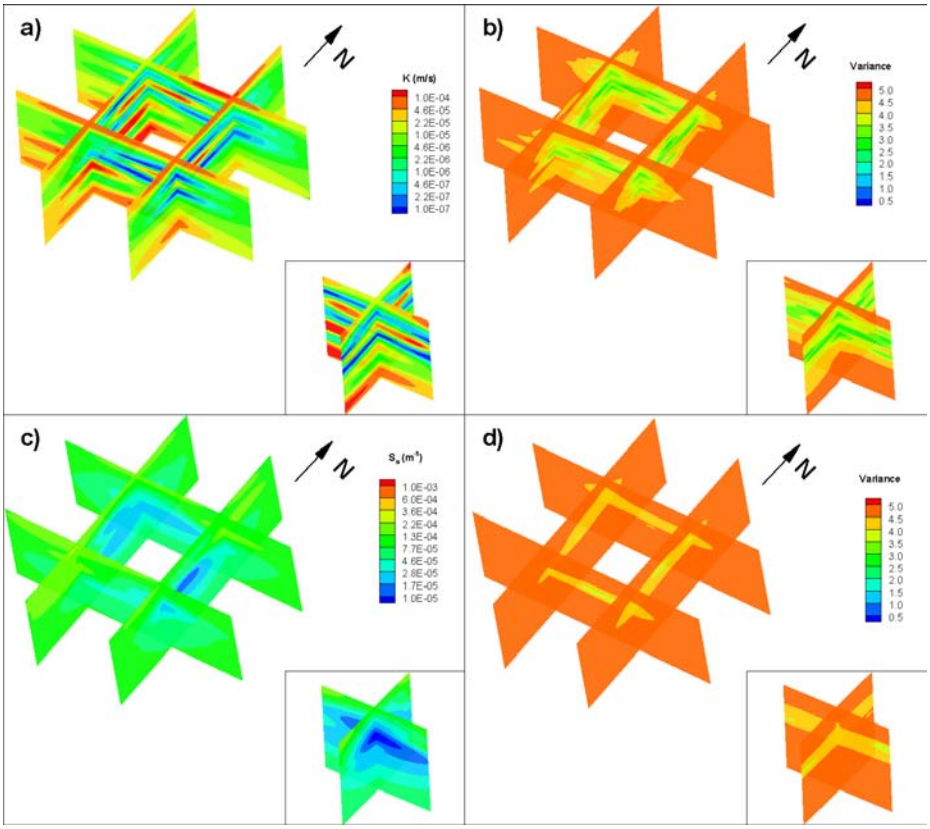


Figure F7: Case 1 THT *K*- and S_s-tomograms after the inclusion of 4 pumping tests (PW4-3, PW1-3, PW5-3, and PW3-3): a) *K*-tomogram*;* b) variance associated with the estimated *K* tomogram; c) S_s-tomogram; and d) variance associated with the estimated S_s-tomogram. Note that the square formed by the slices in the main figure corresponds to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

In general, this case was able to capture the heterogeneity evident in the upper portion of the aquifer. However, the inversion did assign high *K* material to the lower portion of the domain, which, based on core logs and permeameter data (see Figure F1) is known to be silt/clay which has low values of *K*. Examination of the variance maps (Figures F7b and F7c) indicates that the estimates in this region have a high degree of uncertainty, a reflection of the very small, to absent, drawdown responses observed in the lower ports, even at very late time (>20 hrs) (see for example, Figures F4a and F4b).

F5.2 Case 2 – THT with layered, homogeneous K and S_s fields as starting model

For this case, the aquifer system is conceptualized as a main aquifer system bounded above and below by aquitards (each 5 m thick) based on the information on stratigraphy from Figure F1. Initial values of K and S_s for these layers were estimated by performing a *Hantush* [1960] type-curve analysis on the observation points located within the central aquifer unit during pumping tests performed at PW1-3, PW1-4, and PW 1-5. Each drawdown response was individually matched using *Aqtesolv Pro 4.5* [*Duffield*, 2007] to estimate K and S_s for the 'aquifer zone' as well as leakage parameters for the adjacent aquitards. The leakage parameters can then be used to calculate estimates of K and S_s for the aquitards. However, since this solution does not account for observation ports situated within the aquitard units, there is much more uncertainty in the estimates of these parameters. As such, the geometric mean of the aquitard estimates was taken and assigned to both aquitards. Additionally the geometric mean of all K and S_s estimates made for the aquifer zone was used to represent the aquifer.

Due the heterogeneity at the site, the drawdown arrived very early at some ports, resulting in unrealistically small estimates of S_s ($< 1 \times 10^{-6}$ m^{-1}) when using the *Hantush* [1960] type-curve method. As such, any estimates smaller than 1×10^{-6} m^{-1} were omitted from the average. Values assigned to the aquifer as starting conditions for THT were 8.53×10^{-6} m/s and 2.23×10^{-5} m^{-1} for K and S_s respectively; and 4.47×10^{-7} m/s and 1.02×10^{-4} m^{-1} for the aquitard K and S_s respectively.

Figure F8 shows the results of the THT analysis starting with a simple layered model. While the upper and lower portions of the model domain appear significantly different than the Case 1, the pattern in the central aquifer zone is quite similar (see inset in Figures F8a and F8b).

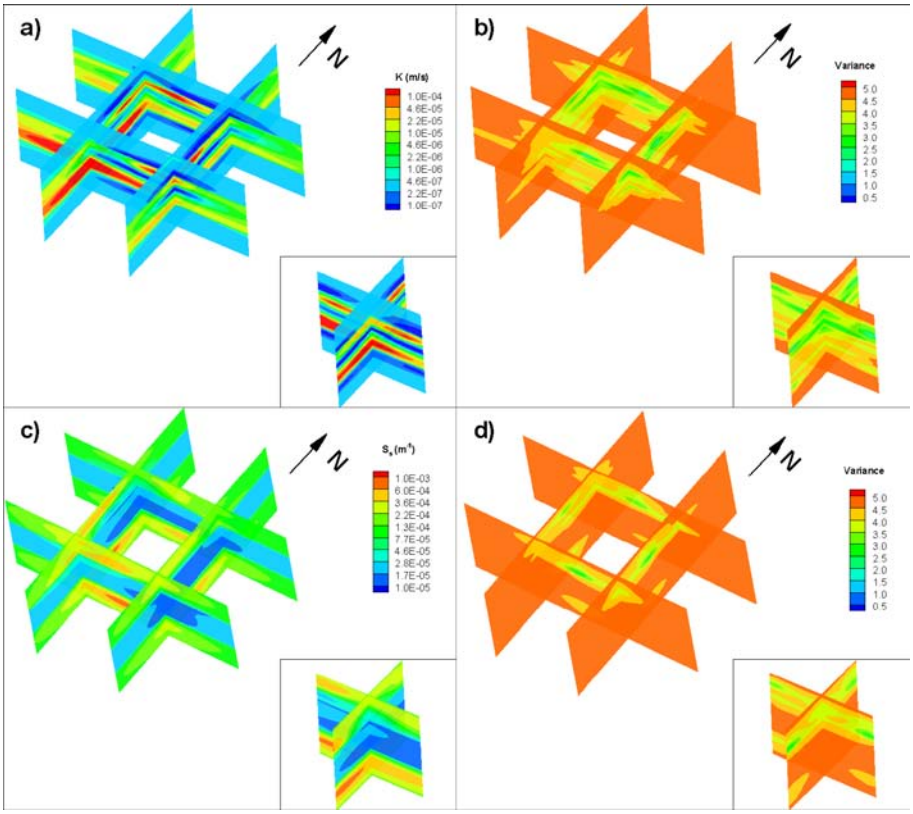


Figure F8: Case 2 THT *K*- and S_s-tomograms after the inclusion of four pumping tests (PW4-3, PW1-3, PW5-3, and PW3-3): a) *K*-tomogram; b) variance associated with the estimated *K*-tomogram; c) S_s-tomogram; and d) variance associated with the estimated S_s-tomogram. Note that the square formed by the slices in the main figure corresponds to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

F5.3 Case 3 – THT populated with permeameter *K* estimates for starting model

The THT code [*Zhu and Yeh*, 2005] allows for the inversion to begin with local-scale estimates of *K* and S_s when they are available. For this case, the THT analysis started with a homogeneous *K* and S_s field (same values as case 1), and had 127 values of *K* specified along boreholes PW1, CMT1, CMT2, CMT3, and CMT4 (See Figure F1). The *K* estimates were based on the permeameter measurements of *Alexander et al.* [2011]. Since the vertical element size was 0.5 m and *Alexander et al.* [2011]

sampled at intervals as small as 0.1 m, the geometric mean of multiple samples that fell within the same model element was used.

Illman et al. [2008] found that conditioning the inverse model with poor data can have a detrimental effect on the resulting tomogram. Since our estimates of S_s span a very wide range and have a high degree of uncertainty, we chose to omit S_s values from the conditioning.

The results of THT that include permeameter K data in the starting condition are presented in Figure F9. Initially, the tomograms appear quite different than those presented previously. However, the difference mainly occurs in the lower region of the aquifer where there is greater uncertainty in the estimated values. In the middle to upper portions of the model domain where the main aquifer zones exists, we see a very similar pattern of an aquifer zone separate by a low K zone.

Tables F3a and F3b summarize the statistical parameters of the K- and S_s- tomograms respectively. As seen in the inversion of the single tests discussed in section 4, many of the vertical variograms exhibited the 'hole effect' which is a result of the layered nature of the estimated fields. For K, the horizontal correlations lengths were approximately two to three times greater than the vertical correlation lengths. For S_s, the horizontal correlations lengths were approximately one and a half times greater than the vertical correlation lengths. The average horizontal correlation lengths of K and S_s from the THT analysis were 8.4 m and 9.2 m, respectively. On the other hand, the vertical correlation lengths of K and S_s were 2.4 m and 5.5 m, respectively.

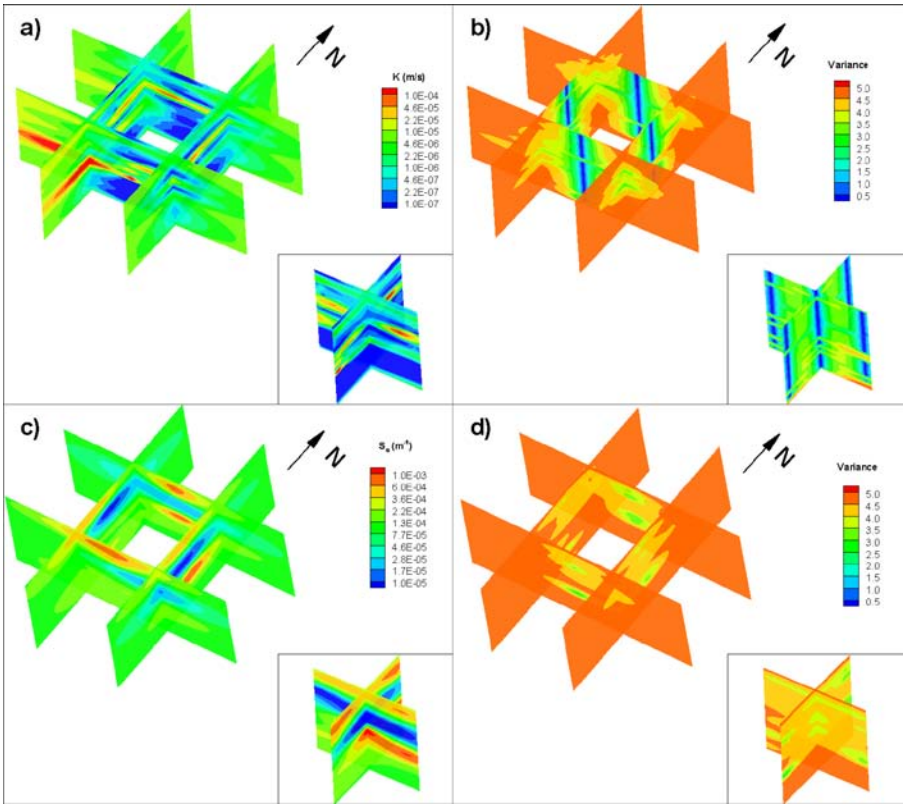


Figure F9: Case 3 THT *K*- and S_s-tomograms after the inclusion of four pumping tests (PW4-3, PW1-3, PW5-3, and PW3-3): a) *K*-tomogram; b) variance associated with the estimated *K*-tomogram; c) S_s-tomogram; and d) variance associated with the estimated S_s-tomogram. Note that the square formed by the slices in the main figure corresponds to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

Table F3a: Geometric mean, variance, and correlation lengths of ln *K* for the various THT cases.

	ln *K* (m/s)						
	Geometric Mean*	**Variance***	**Sill**	λ_x	λ_y	λ_z	**Model**
Case 1	7.4×10^{-6}	3.5	3.7^{xy}, 4.5^{z}	5.9	5.9	2.3	Exponential
Case 2	4.5×10^{-7}	4.1	5^{xy}, 6^{z}	6.3	6.3	1.8	Exponential
Case 3	9.4×10^{-7}	4.0	5.0^{xy}, 4.0^{z}	14.0	12.0	3.0	Exponential

*Calculated from raw data

^{xyz} direction variance was measured in

Table F3b: Geometric mean, variance, and correlation lengths of ln S_s for the various THT cases.

	ln S_s (m^{-1})						
	Geometric Mean*	**Variance***	**Sill**	λ_x	λ_y	λ_z	**Model**
Case 1	5.9×10^{-5}	0.4	0.5^{xy}, 0.6^{z}	10.8	10.8	6.3	Gaussian
Case 2	1.0×10^{-4}	1.5	1.6^{xy}, 3.5^{z}	9.9	9.9	6.3	Gaussian
Case 3	1.1×10^{-4}	0.9	1.3^{xy}, 2.5^{z}	7.0	7.0	4.0	Gaussian

*Calculated from raw data

^{xyz} direction variance was measured in

F5.4 THT Calibration

The K- and S_s-tomograms generated for Cases 1, 2 and 3 were used to simulate the 4 pumping tests used for their creation (PW4-3, 1-3, 5-3, and 3-3). Forward simulations were performed using HydroGeoSphere (HGS) [*Therrien et al.*, 2005] and the model domain described in section 3 for the inverse simulations.

Figure F10 is a scatterplot comparing simulated and observed drawdown at all ports for the 1st hour of each pumping test (i.e., the same data included in the THT code for the creation of these tomograms). These scatterplots, can be considered calibration scatterplots as it is an assessment of how well the tomograms were able to match the data. In general, the matches were quite good with the data clustering around the 1:1 line although some bias is evident for some of the matches. Some exceptions do occur, most notably, with results from PW1-3 for Case 3 revealing a slope of 0.25 indicating that the simulated drawdown during the first hour of the test is considerably lower than the observed, possibly indicating that estimates of K in the vicinity of PW1-3 are too high. In contrast, the slope for PW1-3 for cases 1 and 2 are quite close to 1 (0.97 and 0.96 respectively).

Overall, the scatterplots for cases 1 and 2 appear to look similar, whereas case 3 tends to perform slightly worse. This is likely a result of the limitations imposed by starting the inversion with

the permeameter K data, suggesting that some of the estimated values K from the permeameter may not be accurate. This may be a result of incomplete core recovery. In particular, *Alexander et al.* [2011] noted that approximately 80% core recovery was obtained for CMT1, CMT2, CMT4 and PW1, and only 69% recovery was obtained for CMT3. This reinforces the suggestion of *Illman et al.* [2008] that conditioning of the model with field data should be done carefully.

As discussed earlier, the THT analysis for case 1 began with homogeneous K and S_s fields while case 3 incorporated some information on stratigraphy based on Figure F1. Given that the K and S_s tomograms from case 2 were able to correctly estimate a low K zone near the bottom of the model, while the tomograms from case 1 could not, we consider the tomograms from case 2 to be the most accurate among the 3 cases to this point.

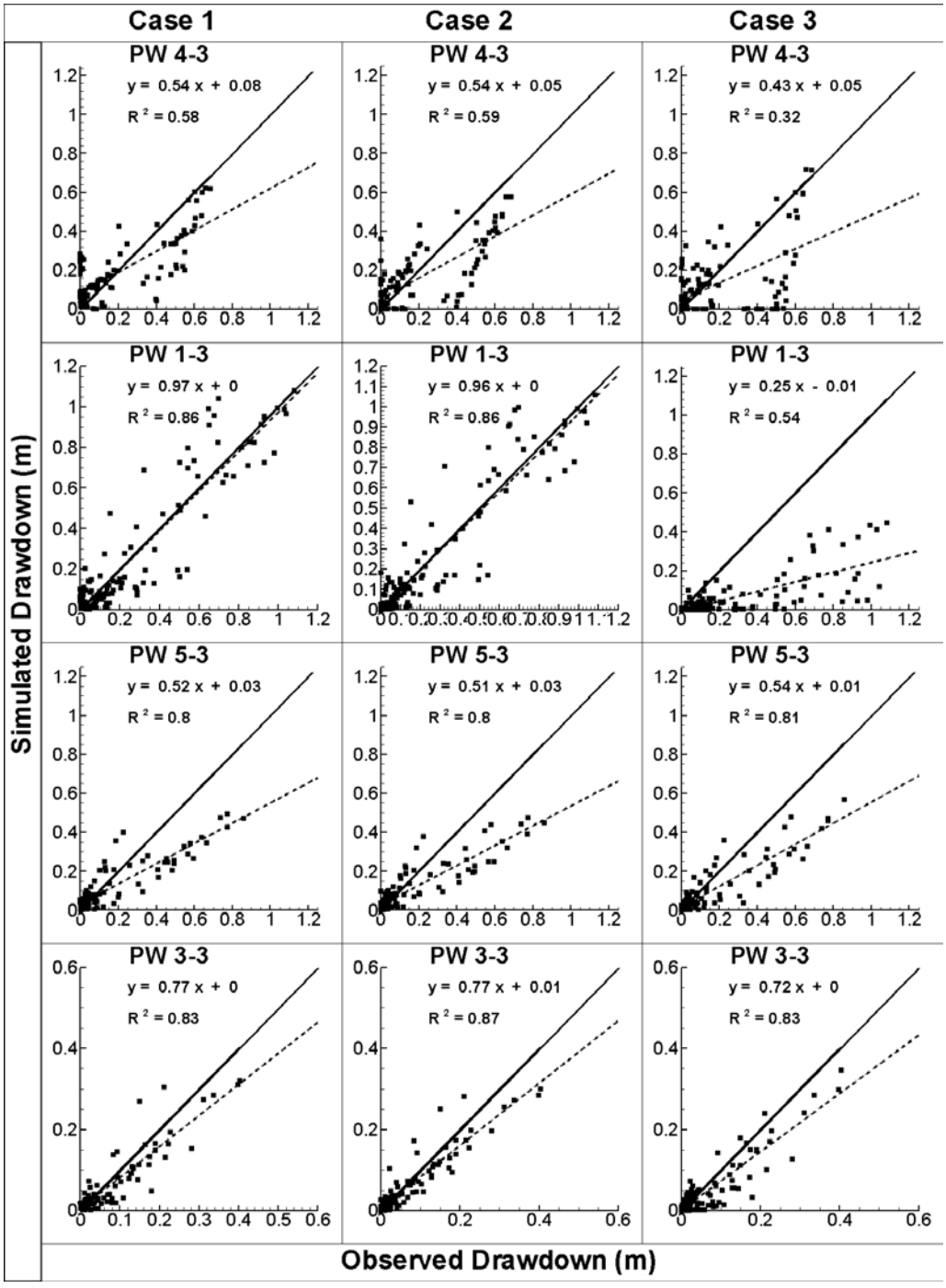


Figure F10: Calibration scatterplots for THT Cases 1 - 3. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

F6. DISCUSSION

To assess the accuracy of the K and S_s tomograms generated by the stochastic inversion of individual pumping tests and the 3 THT cases, we evaluate them: 1) visually (i.e., do the tomograms reflect our knowledge of the site based on core data?); 2) by comparing K profiles along boreholes to permeameter values; and, 3) simulating all 9 pumping tests using the K and S_s tomograms from each case and comparing the simulated drawdown to the observed drawdown.

F6.1 Visual assessment

Here, we compare our inverse modeling results to the stratigraphy inferred from the core logs (Figure F1) as well as to Figure F11. In particular, many of the estimated K and S_s fields generated by the stochastic inversion of individual pumping tests capture the presence of an aquifer zone, however, the details and extent are lacking (see Figure F6 and Figures SF9 – SF16). Additionally, as a result of starting with homogeneous K and S_s fields, the inversion produces a high K zone at the lower part of the model domain where very little drawdown was observed. Based on the permeameter data, the lower portion of the model domain should be represented as a low K zone.

All 3 THT cases (Figures F7 – F9) share some very similar features, most notably, the double aquifer system located in the upper portion of the model domain. This is consistent with the 3D model of stratigraphy that was developed for this site (Figure F11). The stratigraphic model was created based on the soil classification performed by *Alexander et al.* [2011] by analyzing the soil cores collected along PW1, CMT1, 2, 3, and 4. In total 8 soil classifications were assigned, ranging from silty clay to sandy gravel. The interpolation between boreholes was performed using an inverse distance approach built into RockWorks 2006 [RockWare, 2007]. The overall domain dimensions are identical to those for the THT analysis, however, for this case the elements are uniform in size (0.5 x 0.5 x 0.5 m) and total 243,000 elements.

While case 1 appears to display the double aquifer in the upper portion of the model domain, it does not accurately reproduce the bottom half of the model domain. In this portion, case 1 contains high *K* materials which are clearly absent from the stratigraphic map in Figure F11. Starting case 2 as a simple layered system honours the low *K* zone in the bottom half of the aquifer. This also has the effect of improving S_s estimates in the upper and lower portions of the model domain as reflected by the S_s-variance field shown in Figure F8d. Since case 3 was populated with the permeameter data that in general follows the stratigraphy (Figure F1 and F11), it is reasonable to expect that it most closely matches the stratigraphy. Additionally, estimates of S_s were also improved for this case as reflected by the S_s-variance field shown in Figure F8d.

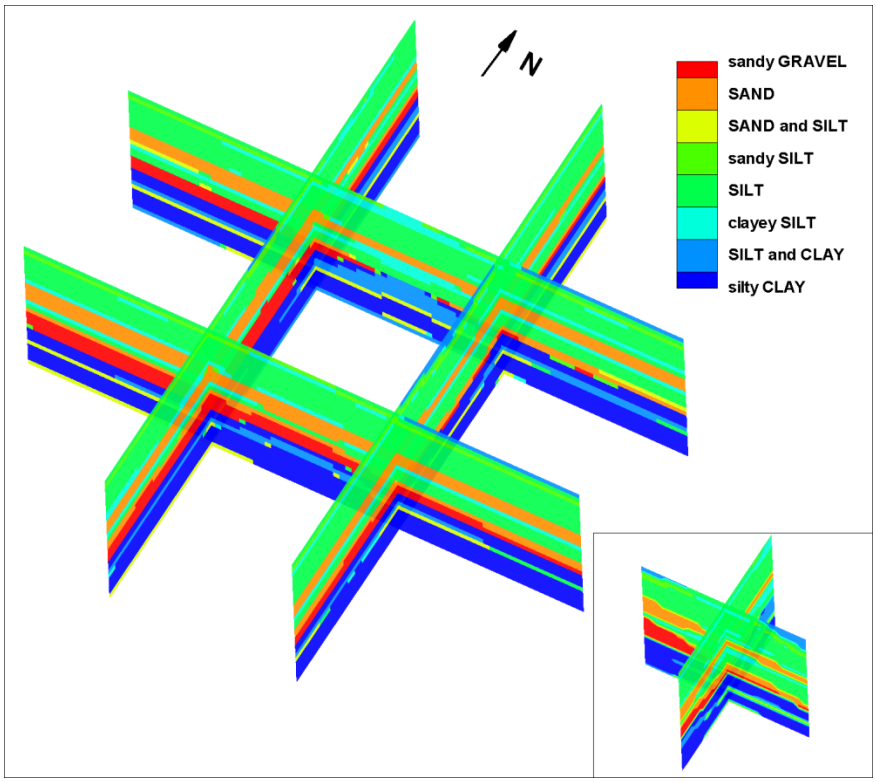
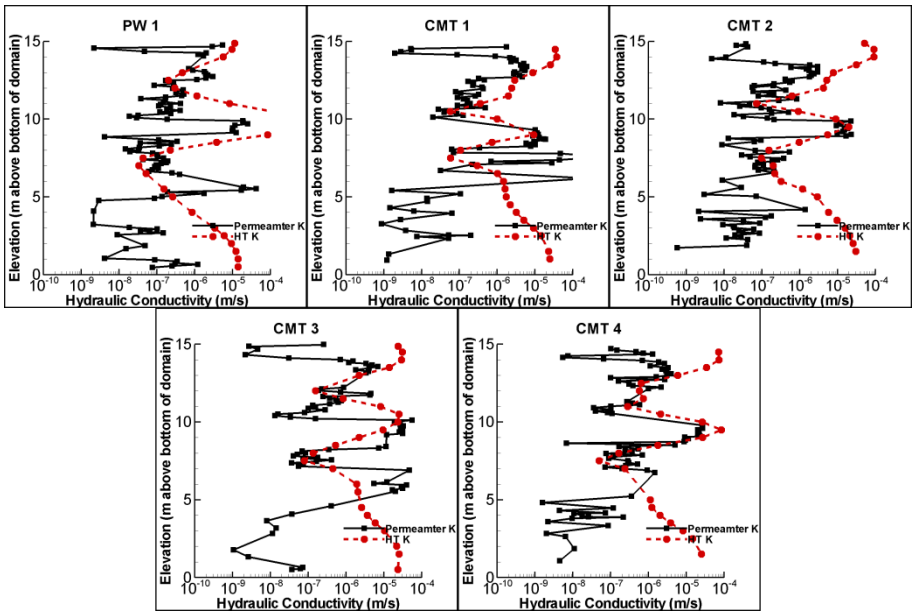


Figure F11: Three-dimensional stratigraphy model created based on the descriptions of core samples. Note that the square formed by the slices in the main figure correspond to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

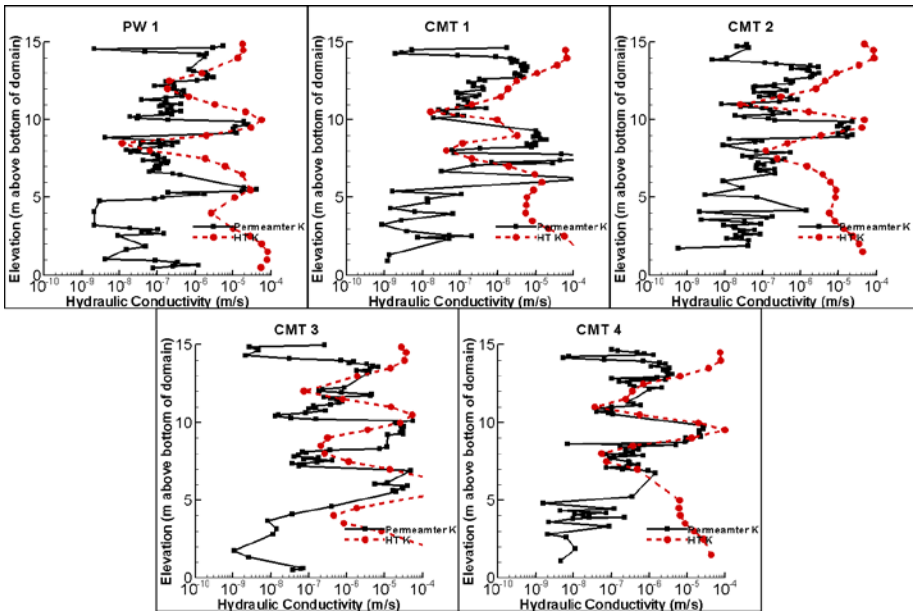
F6.2 Comparison to permeameter *K* data

The permeameter *K* data obtained by *Alexander et al.* [2011] is compared to the *K* distribution from the stochastic inverse modeling of the pumping test at PW1-3 and the *K* tomograms from three THT cases. The comparison is made along 5 boreholes (PW1, CMT1, CMT2, CMT3, and CMT4). Figures F12a, F12b, F12c, and F12d show the comparison for the stochastic inversion using the cross-hole test at PW1-3, and THT cases 1, 2, and 3 respectively.

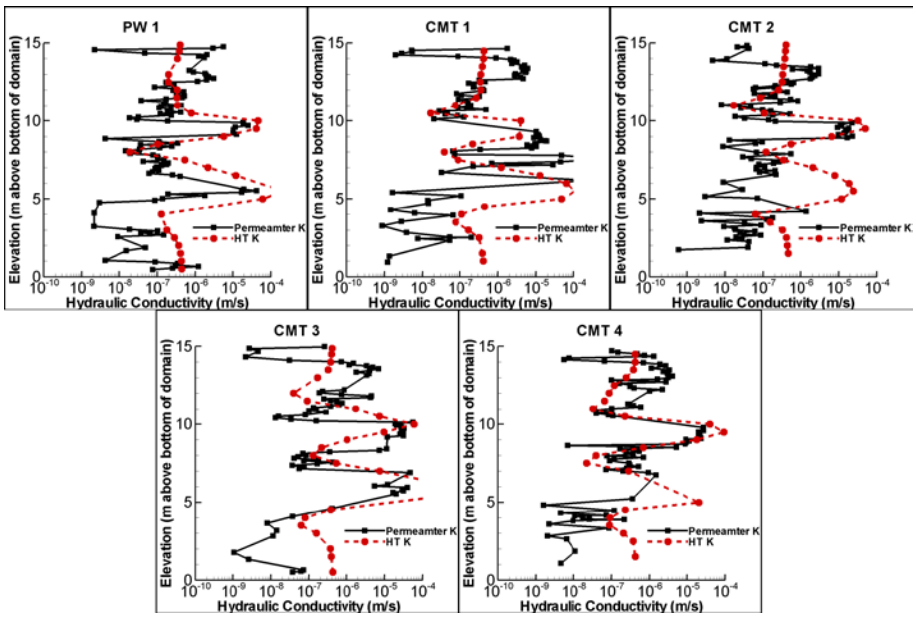
a)



b)


PW 1
CMT 1
CMT 2
CMT 3
CMT 4
Elevation (m above bottom of domain)
Hydraulic Conductivity (m/s)
Permeamter K
HT K

c)


PW 1
CMT 1
CMT 2
CMT 3
CMT 4
Elevation (m above bottom of domain)
Hydraulic Conductivity (m/s)
Permeamter K
HT K

d)

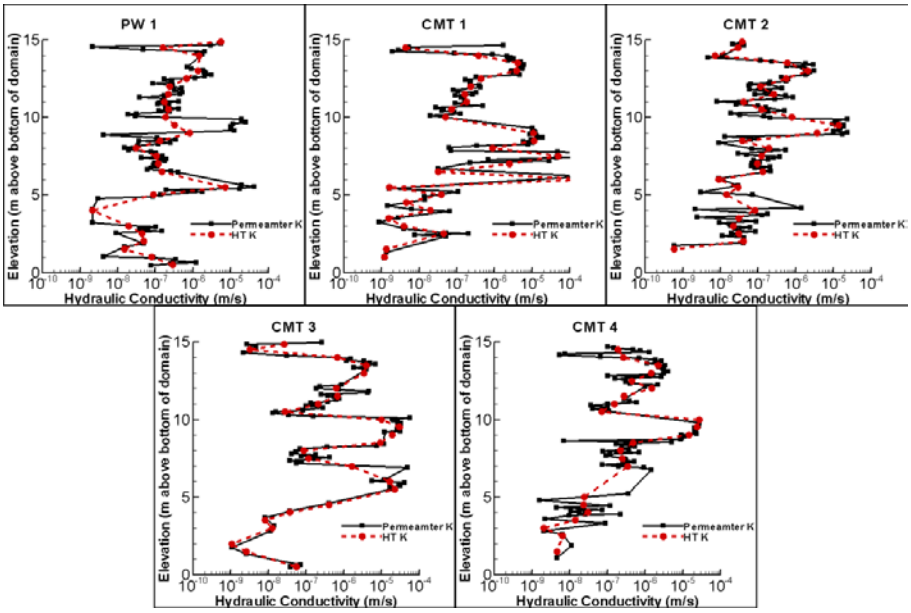


Figure F12: Vertical *K* profile along boreholes for the stochastic inversion of the pumping test at PW1-3 (a), and all THT cases, 1 (b), 2 (c), and 3 (d).

Examination of Figure F12a shows that the stochastic inversion of a single pumping test (PW1-3) indicates the presence of a single high *K* zone located approximately 10 m above the base of the model domain. However, significant deviation from the permeameter data is evident in the upper and lower portions of the model domain where high *K* values are estimated, in contrast to the low *K* values measured by the permeameter tests. This is consistent with the visual interpretation of the estimated *K* field (Figure F6).

Figure F12b shows that the *K* tomogram from Case 1 captures the main high *K* unit (located approximately 10 m above the base of the model domain) and the adjacent low *K* zones. The capturing of the main high *K* unit is particularly evident for CMT4. However, the *K* profile as estimated by THT for this case deviates significantly from the permeameter data near the upper and lower boundaries of the model domain.

Figure F12c shows that Case 2 is able to quite accurately capture both aquifers located at 10 m and 5 m above the base of the model domain. However, we notice that the values of K of the aquifer units are noticeably higher in the K tomogram in comparison to the permeameter data. Additionally, the upper and lower portions of the model domain are better represented for this case in comparison to case 1. This suggests that starting the THT analysis with K and S_s fields that approximate the known stratigraphy can lead to more realistic tomograms.

Figure F12d shows the K profile from the K tomogram from Case 3 compared to the permeameter K data. Despite being allowed to vary after the estimation of the first pumping test is complete, estimates of K along the 5 boreholes for Case 3 are still quite close to the original data incorporated in the THT inversion. However, as discussed in section 5.4, in terms of model calibration, case 3 performed the worst among the 3 THT cases. This could suggest that permeameter K data, even with the availability of a large number of estimates, may not entirely be suitable for representing large scale groundwater flow within this aquifer-aquitard sequence as heterogeneous as this site. This once again reinforces the argument of *Illman et al.* [2008] that the data included as 'true' must be selected carefully.

F6.3 Prediction of independent pumping tests

The prediction of independent pumping tests that are not used in the inversion is another way to validate the tomograms. For each of the stochastic inversions, as well as the 3 THT cases, all 9 pumping tests were simulated using the resulting K and S_s tomograms. This comparison is made using scatterplots, where observed and simulated drawdowns are compared at 5 times throughout the duration of the pumping tests (these times are different for each test, and range from early time to late time). Forward simulations were performed using HGS [*Therrien et al.*, 2005] and the model domain described in section 3 for the inverse simulations.

F6.3.1 Validation of K and S_s distributions from the stochastic inversion of individual tests

All 9 pumping tests were simulated for each of the K and S_s fields generated from the stochastic inversion of individual pumping tests (i.e., a total of 81 forward runs consisting of 9 K and S_s fields for each of the 9 pumping tests). Figure F13 is a sample scatterplot using the K and S_s fields estimated for stochastic inversion of the pumping test performed at PW5-4. As expected, the best match is for PW5-4, the test which was used for model calibration. Some of the other tests are also matched quite well (PW1-5, PW4-3), however, the matches of other tests are poor. Results from the stochastic inverse modeling of other single pumping tests yielded similar results and are presented in Table F4. This table summarizes the parameters for the best-fit lines for all 9 pumping tests by presenting the minimum, maximum, and mean values for the slope, intercept, and the coefficient of determination (R^2).

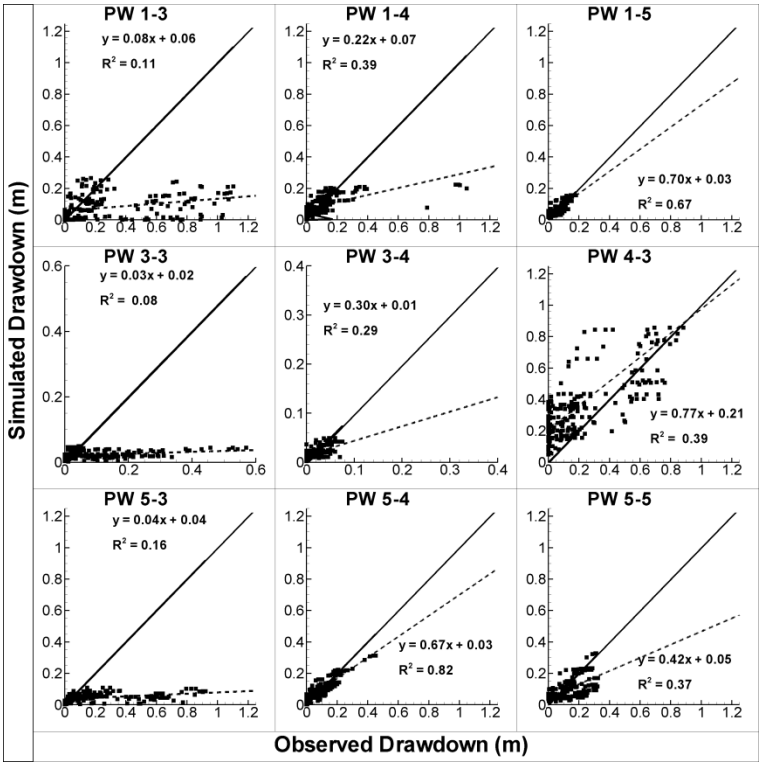


Figure F13: Scatterplot for all 9 pumping tests using the estimated K and S_s fields from the stochastic inversion of the pumping test at PW5-4. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

Table F4: Summary of single test inversion validations (slope etc.)

	Slope			***Intercept***			***R^2***		
	Min	**Max**	**Average**	**Min**	**Max**	**Average**	**Min**	**Max**	**Average**
PW1-3	0.15	0.90	0.36	0.01	0.18	0.07	0.03	0.92	0.31
PW1-4	0.03	1.32	0.45	0.01	0.17	0.04	0.10	0.84	0.27
PW1-5	0.05	0.89	0.41	0.01	0.14	0.04	0.12	0.90	0.40
PW3-3	0.05	0.89	0.38	0.01	0.38	0.13	0.003	0.69	0.20
PW3-4	0.05	1.27	0.47	0.01	0.17	0.06	0.13	0.73	0.37
PW4-3	0.04	0.78	0.33	0.01	0.07	0.03	0.17	0.87	0.40
PW5-3	-0.02	0.51	0.20	0.02	0.33	0.11	0.0002	0.58	0.18
PW5-4	0.03	0.77	0.36	0.01	0.21	0.06	0.08	0.82	0.36
PW5-5	0.09	0.80	0.42	0.02	0.20	0.06	0.15	0.88	0.40

F6.3.2 Validation of K and S_s tomograms from THT

For each of the THT cases, all 9 pumping tests were simulated and the observed and simulated results were compared using scatterplots (see Figures F14, F15, and F16 for cases 1, 2, and 3 respectively). We note that THT for all 3 cases was performed using the 1st hour of data recorded for each pumping test, however, the entire test duration was simulated for each pumping test (6.5 - 22.5 hrs). Thus, the difference between Figure F10 (calibration plots for each THT case) and the validation plots (Figures F14, F15, F16) is the simulation of the entire duration of the test.

All 3 cases appear to perform quite similarly, with quite good matches for the pumping tests used for calibration purposes (PW1-3, PW4-3, PW5-3, and PW3-3). Poorer matches are seen for pumping tests not included in the inversion, in particular, matches for PW1-4 and PW5-4 are quite poor for all 3 cases. Interestingly, these ports are adjacent to pumping tests included in the THT inversion. Intermediate quality matches are observed for PW1-5, PW3-4, and PW5-5. Overall Case 2 appears to perform slightly better than cases 1 and 3, particularly for independent pumping tests performed at PW1-5, PW3-4, and PW5-5.

F6.4 Overall assessment of transient hydraulic tomography

As part of the THT analysis, we also examined the use of 2 different starting models (cases 1 and 2) as well as the inclusion of permeameter K data along boreholes (case 3). All 3 cases produced fairly similar representations of the central portion of the aquifer where a strong response was recorded, however, the most noticeable difference between these cases was for the lower portion of the domain, where very little to no drawdown was recorded. The lack of drawdown in this lower portion of the system provides a challenge to the inversion resulting in higher uncertainties in this region. Cases 2 and 3 preserved the low K material in the bottom portion of the aquifer, indicating, that for unresponsive regions (or regions with very little drawdown) the inclusion of additional

information, either with a heterogeneous starting model, or the point estimates of hydraulic parameters may improve estimates for these regions. In terms of predicting drawdown, both cases 2 and 3 are very similar, however, case 2 better reproduces the observed drawdowns for the pumping tests at PW1-3 and PW4-3. As such, we identify case 2 as the best performing THT inversion.

Validation of tomograms generated by steady state hydraulic tomography [*Illman et al*., 2010] and transient hydraulic tomography [*Berg and Illman*, in review] have been performed in the laboratory. These validations were performed by simulating 16 independent cross-hole pumping tests and comparing the simulated drawdowns to the observed drawdowns. For both the steady state and transient hydraulic tomography cases, excellent matches were obtained in the laboratory under controlled conditions.

It is clear that the validations performed as part of this study (Figures F14, F15, and F16) are not as good as those from the laboratory studies, however, there are other sources of error and uncertainties in forcing functions (e.g., initial conditions, boundary conditions, source and sink terms, etc.,) in the field. Some possible sources of error that may have affected this study are: 1) the use of a coarse grid limiting our abilities to capture the fine scale details necessary and 2) the larger well spacing and larger spacing between monitoring ports which amounted to more than half the correlation length (for both the horizontal and vertical directions) recommended by *Yeh and Liu* [2000].

One should keep in mind that both limitations are due to technological constraints both of which should be overcome in the future. Despite these challenges, we feel that THT was still able to capture the key features of the aquifer-aquitard sequence at this site and holds promise as a tool for the characterization of heterogeneity in three dimensions at the field scale.

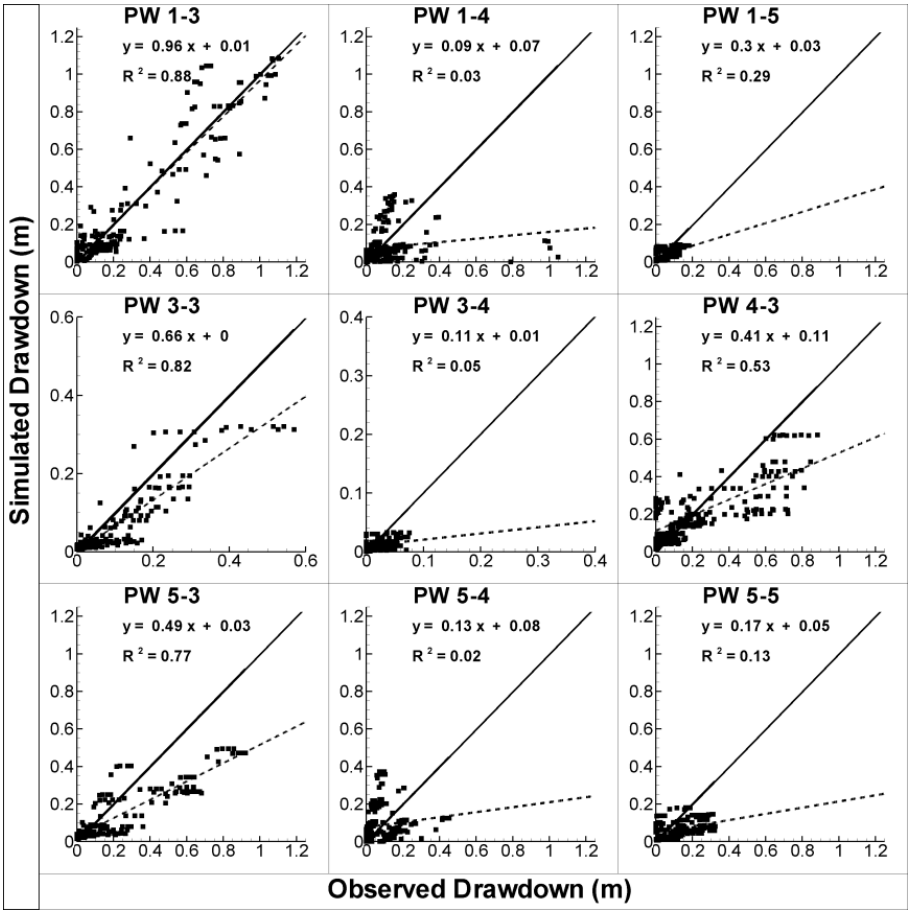


Figure F14: Scatterplot for all 9 pumping tests using the tomograms generated from THT Case 1. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

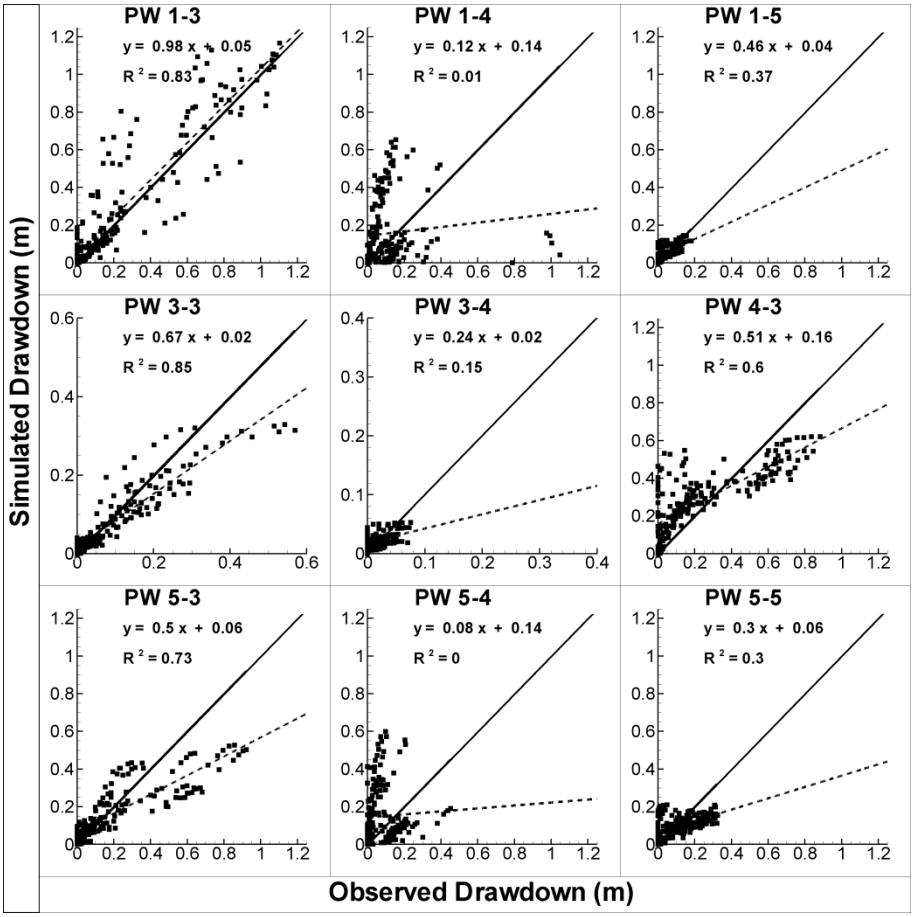


Figure F15: Scatterplot for all 9 pumping tests using the tomograms generated from THT Case 2. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

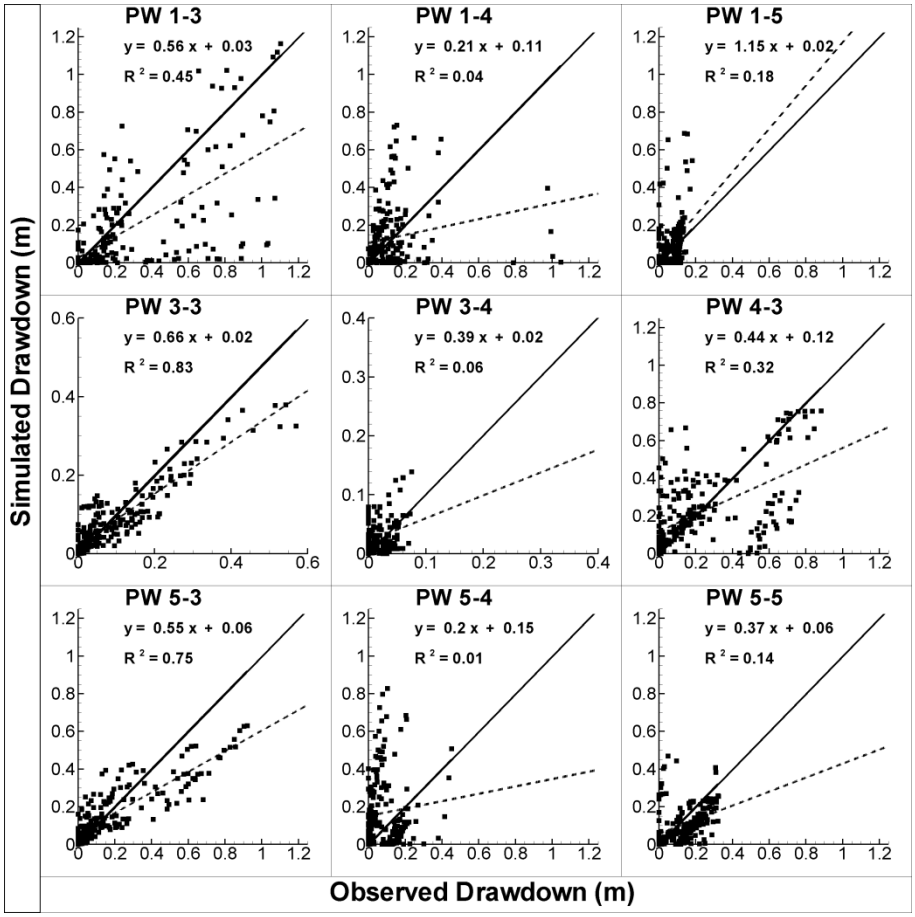


Figure F16: Scatterplot for all 9 pumping tests using the tomograms generated from THT Case 3. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

F7. CONCLUSIONS

In this study we performed 9 pumping tests in a three-dimensional network of pumping and observation wells to validate the transient hydraulic tomography [THT] code of *Zhu and Yeh* [2005] based on the stochastic inverse method. The data collected from these pumping tests were then analyzed to estimate heterogeneous distributions of K and S_s. This was accomplished by: 1) performing stochastic inversions of individual pumping tests; and 2) performing THT using 4 pumping tests. The THT analyses were performed using three different types of *a priori* information: 1) effective K and S_s fields estimated from pumping test data; 2) 3-layer K and S_s fields with each

layer being homogeneous based on parameter estimates from type-curve analysis based on the *Hantush* [1960] solution; and 3) homogeneous K and S_s fields with the inverse model populated with local-scale K obtained via permeameter analysis of core samples along 5 boreholes. The resulting K and S_s distributions were then validated visually by comparison to stratigraphy, by comparison to permeameter K estimates, and by simulating all 9 pumping tests and comparing the simulated drawdown to the observed records.

This study leads to the following major conclusions:

1) It is possible to interpret individual or multiple pumping tests in a highly heterogeneous geologic medium by means of a three-dimensional transient hydraulic tomography (THT) code of *Zhu and Yeh* [2005], which treats the geologic medium as a locally isotropic, heterogeneous continuum. The approach yields highly resolved maps of K and S_s heterogeneity as well as their uncertainties.

2) Many of the single test inversions captured the main aquifer zone, however, the details are lacking in comparison to the THT inversion. Not only are more details evident in the estimated fields with addition of multiple tests, but the ability of the K and S_s tomograms to predict independent pumping tests improves. For the estimated tomograms from single test inversions, the variance of ln K ranged from 1.06 to 4.20, while the variance of ln S_s ranged from 0.04 to 0.36. This suggests that the degree of heterogeneity in K is greater than S_s at this site. The same could be said for variance estimates obtained from THT inversions.

3) Our THT analysis of transient drawdown data from 4 pumping tests showed that the most salient features of the glaciofluvial aquifer-aquitard system can be captured with reasonable fidelity. In particular, the THT analysis detected presence of 2 aquifers as well as the aquitard. However, the THT analysis without the knowledge of site geology (Case 1) yielded unsatisfactory results for

regions with little drawdown, at the lower portion of the model domain, that is known to contain low K material.

4) The THT analyses which started with K and S_s fields that better reflected the stratigraphy (Case 2) or populating the model with K values (Case 3) did not result in significantly improved results in terms of the predictability of independent pumping tests over Case 1. However, the stratigraphy along boreholes was better represented. This suggests that including information about the geology for these regions may produce more realistic K and S_s tomograms.

5) We conclude that while still in the early stages of field implementation, THT appears to be a promising approach for characterizing aquifer heterogeneity for small scale plots. However, more work needs to be done to validate THT at a larger scale. This will be our focus for future studies.

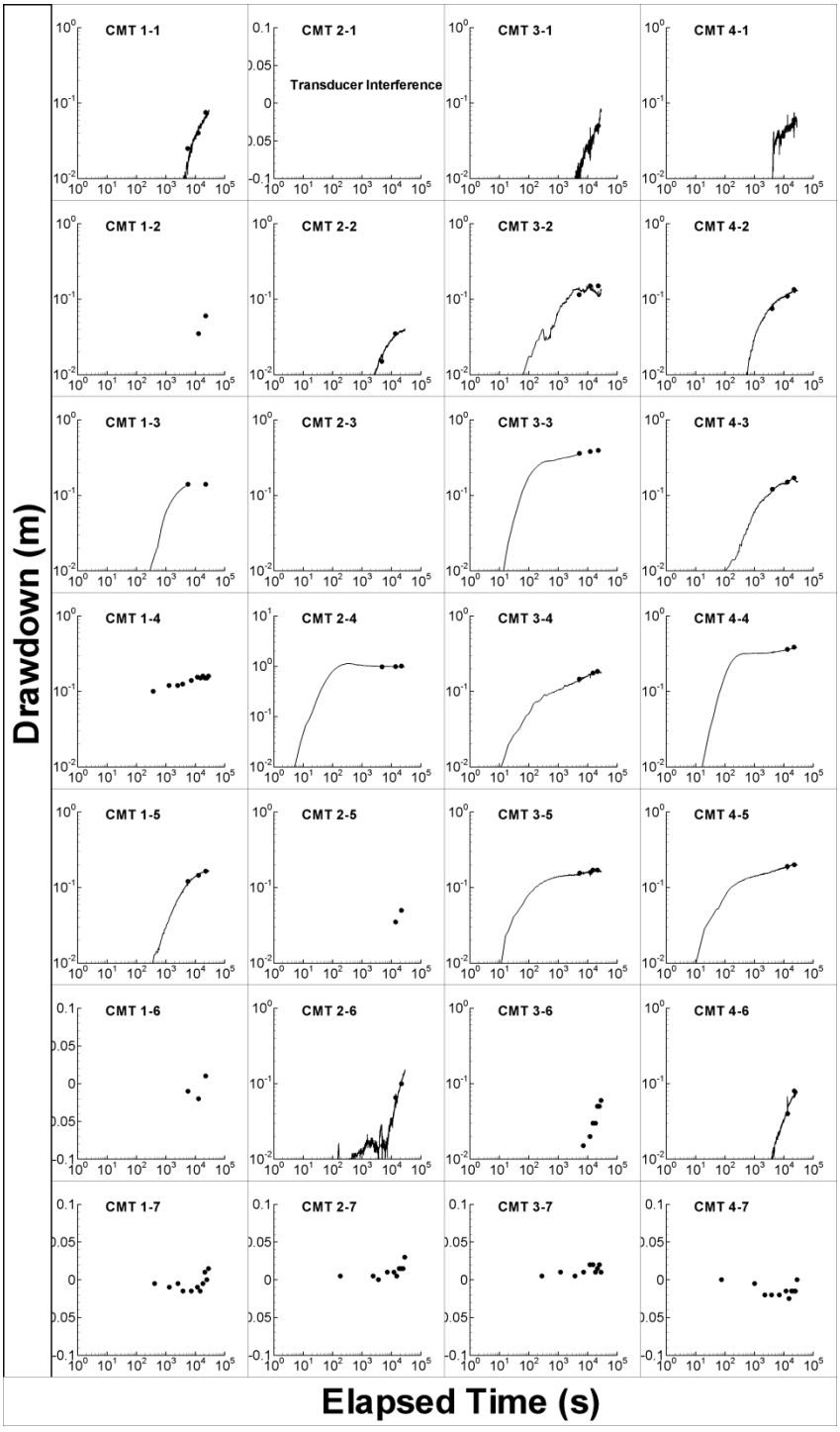


Figure SF1a. Observed drawdown (CMT wells) response for pumping at PW1-4. Note: at CMT2-1 there were problems with the transducer and the collected data was unreliable.

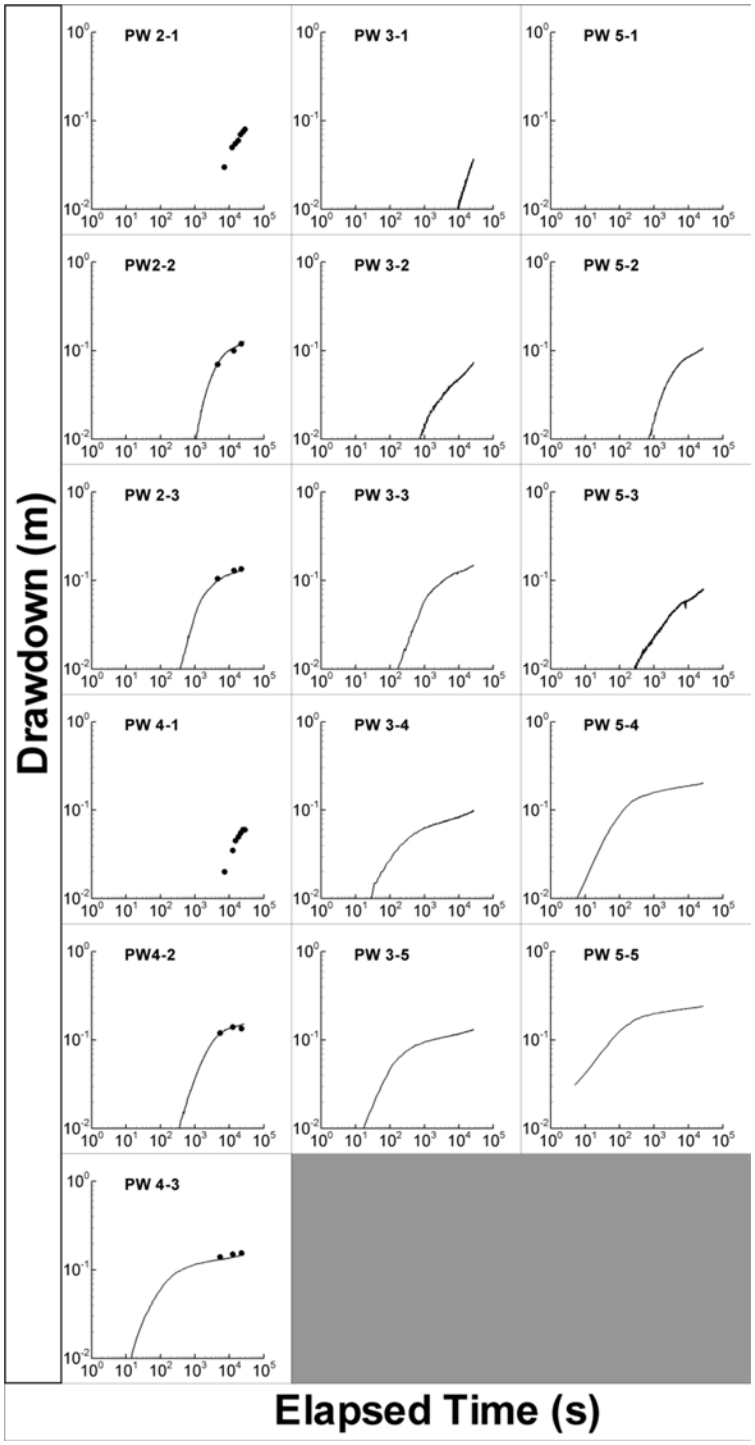


Figure SF1b. Observed drawdown (PW wells) response for pumping at PW1-4.

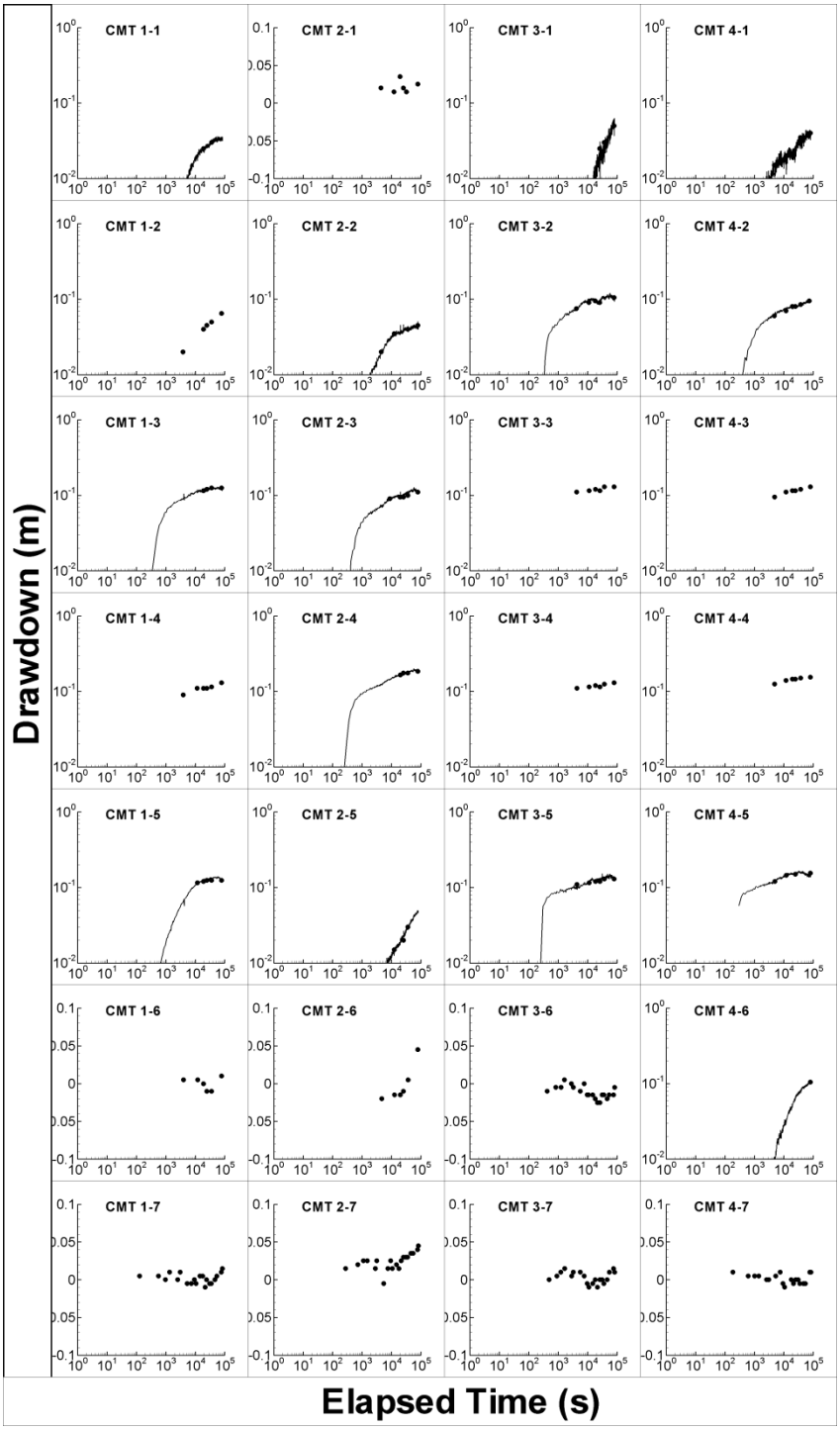


Figure SF2a. Observed drawdown (CMT wells) response for pumping at PW1-5.

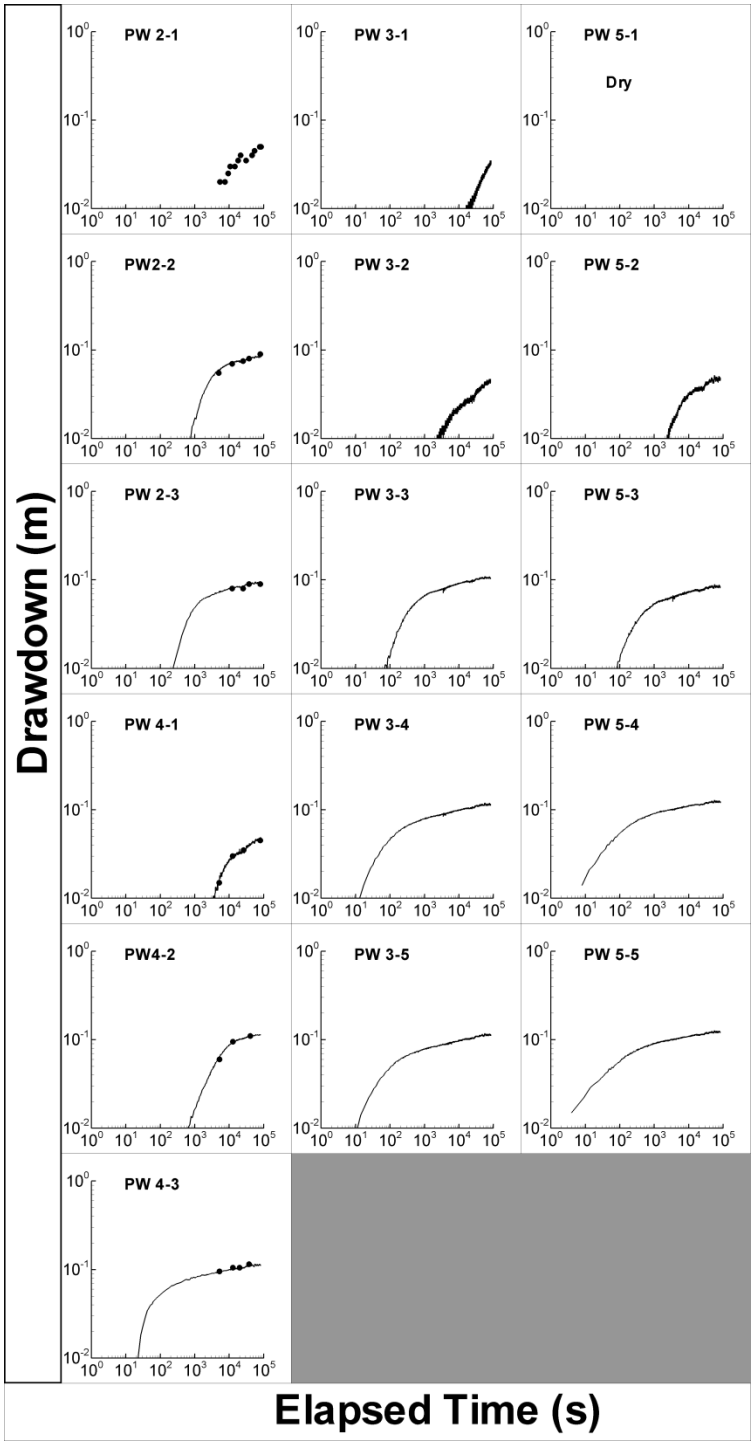


Figure SF2b. Observed drawdown (PW wells) response for pumping at PW1-5.

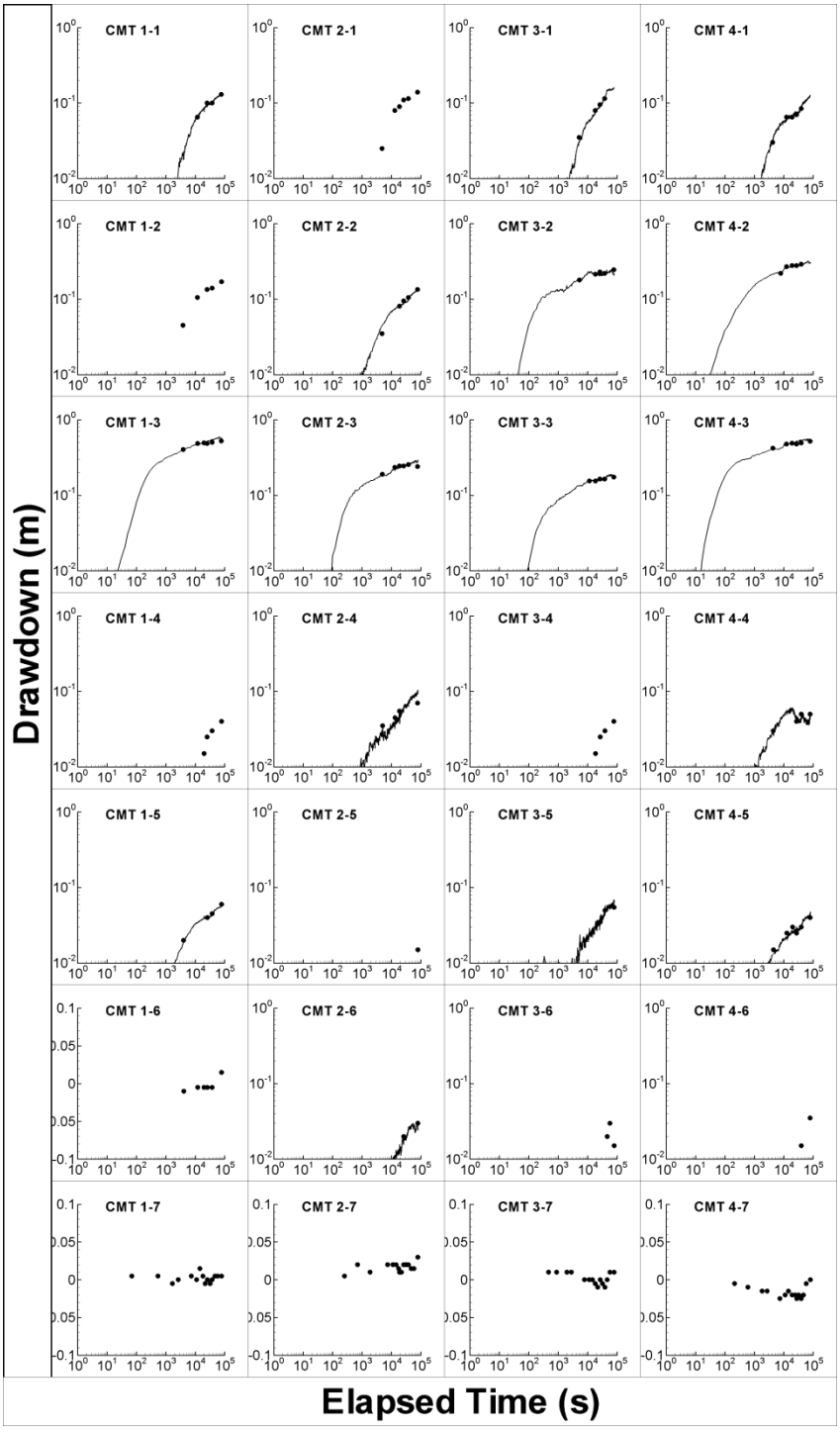


Figure SF3a. Observed drawdown (CMT wells) response for pumping at PW3-3.

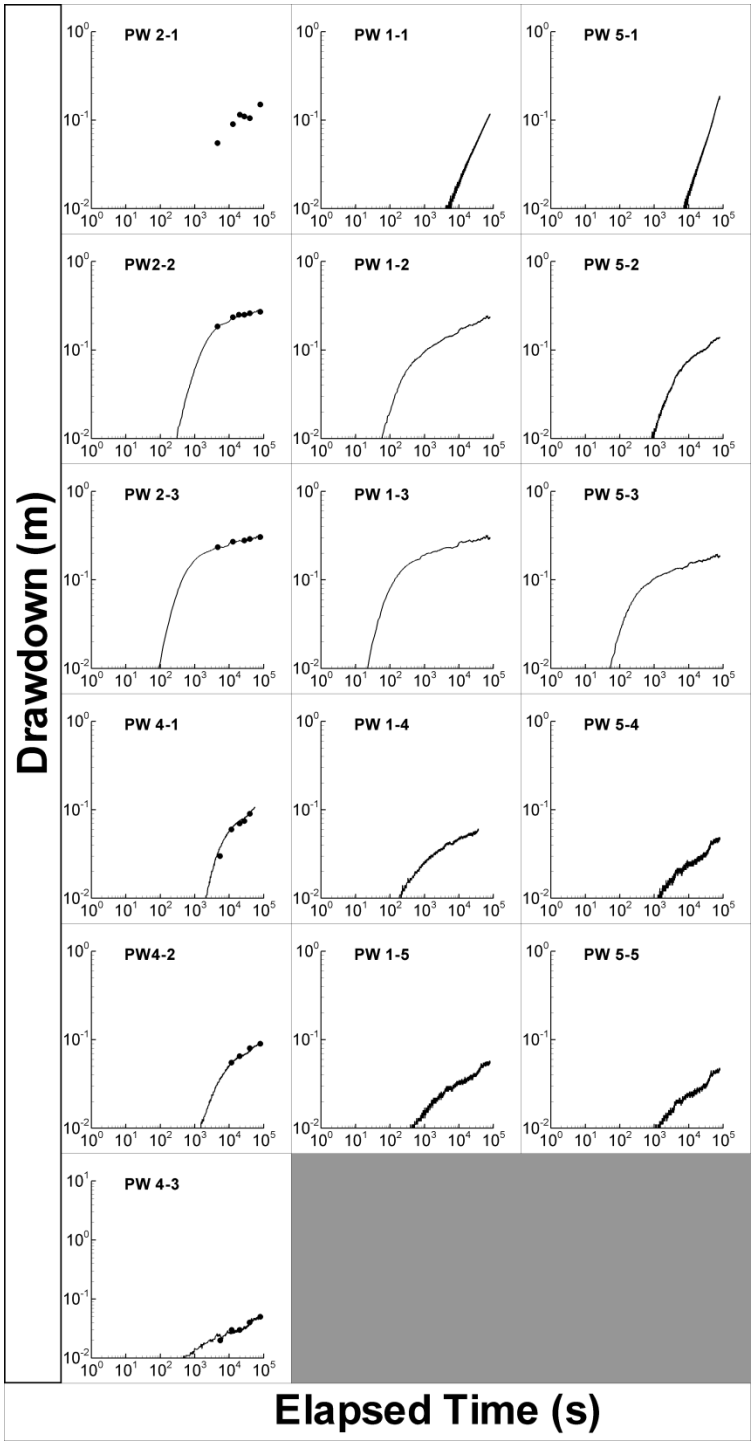


Figure SF3b. Observed drawdown (PW wells) response for pumping at PW3-3.

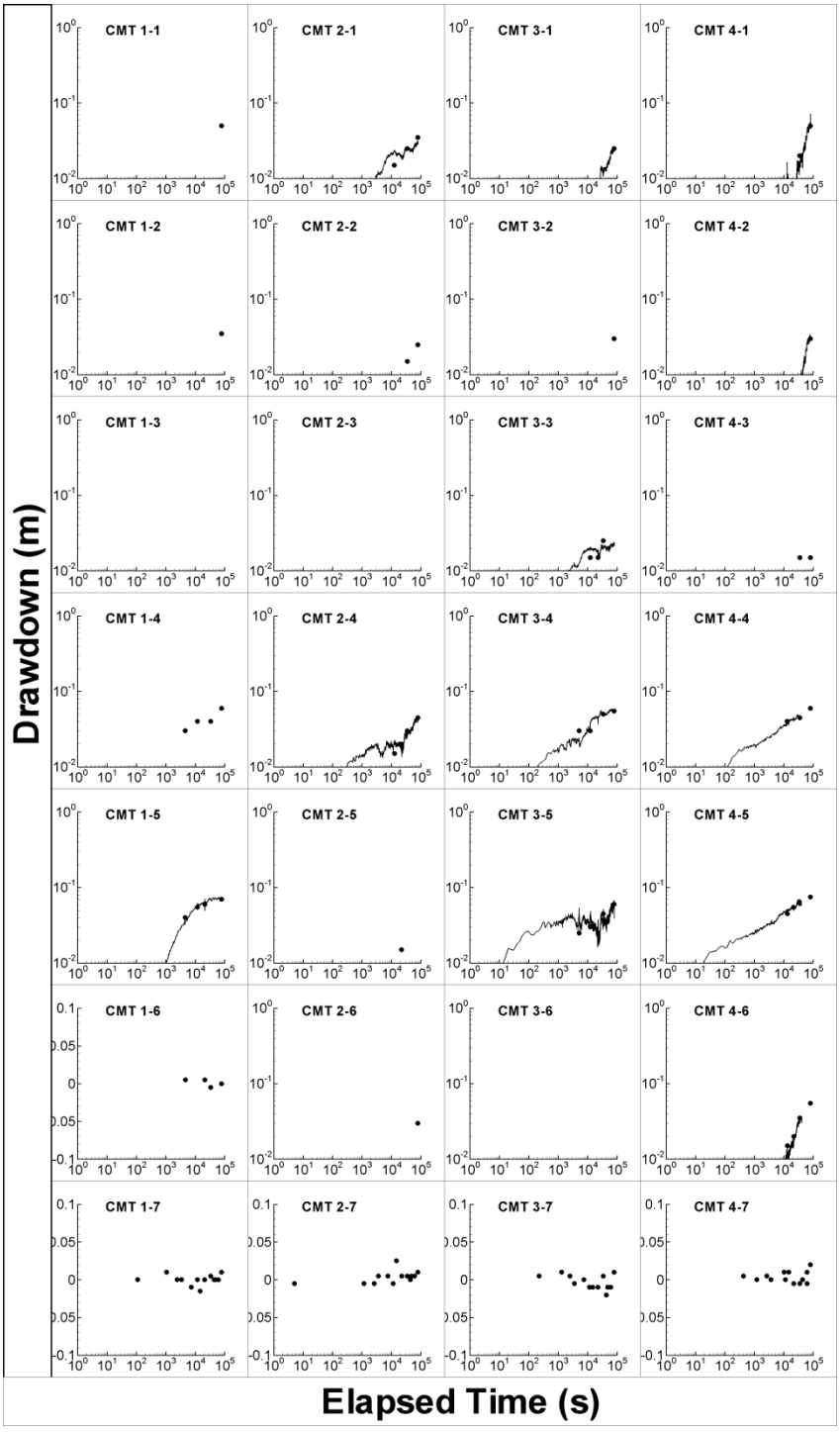


Figure SF4a. Observed drawdown (CMT wells) response for pumping at PW3-4.

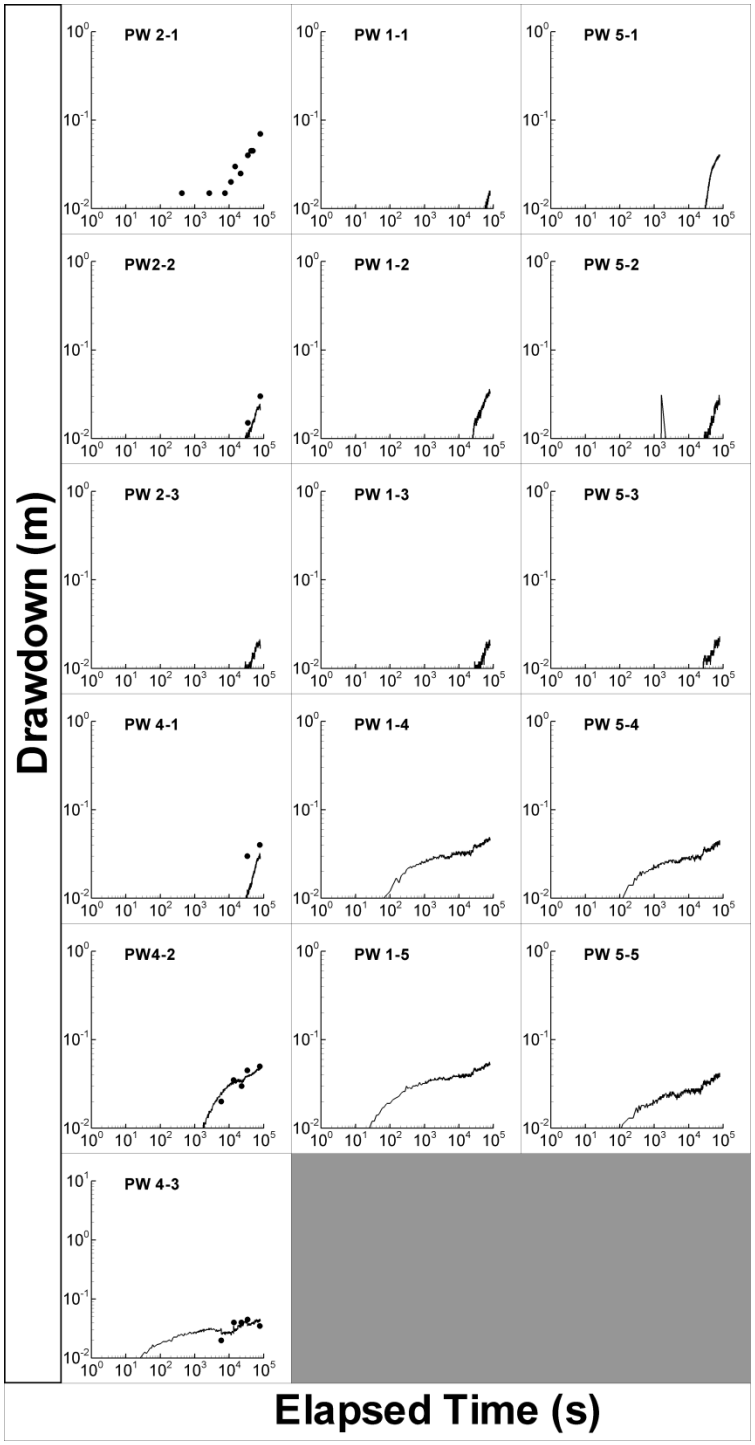


Figure SF4b. Observed drawdown (PW wells) response for pumping at PW3-4.

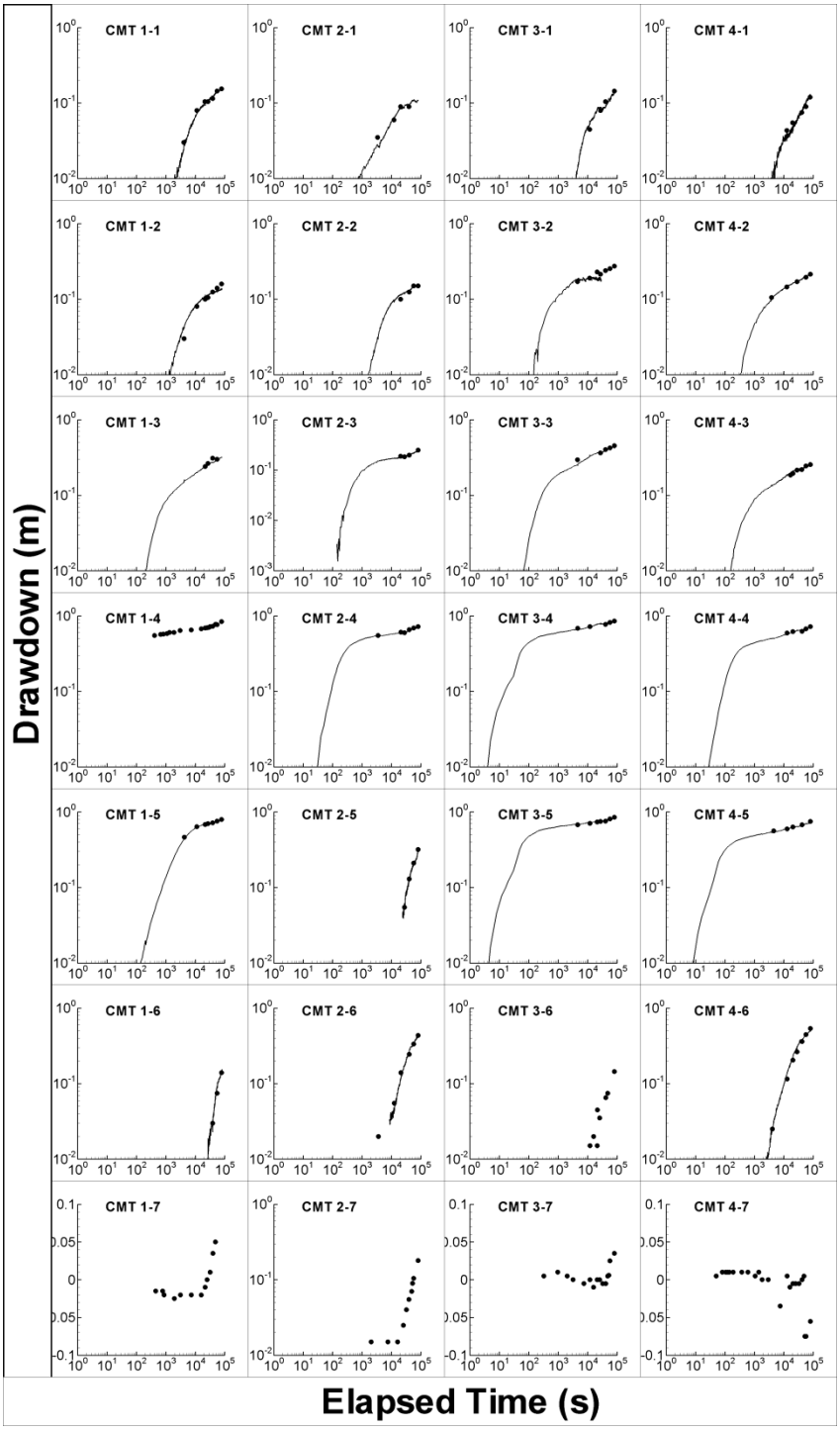


Figure SF5a. Observed drawdown (CMT wells) response for pumping at PW4-3.

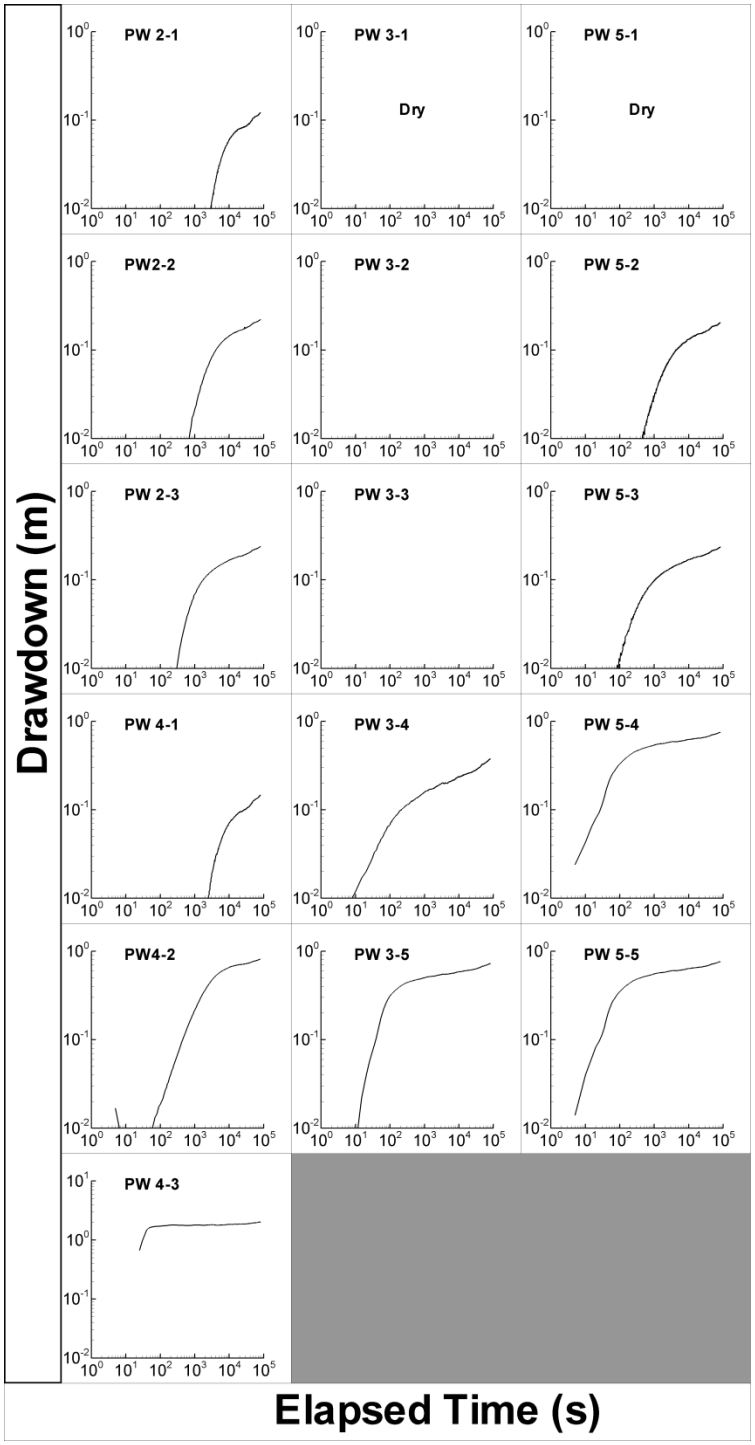


Figure SF5b. Observed drawdown (PW wells) response for pumping at PW4-3.

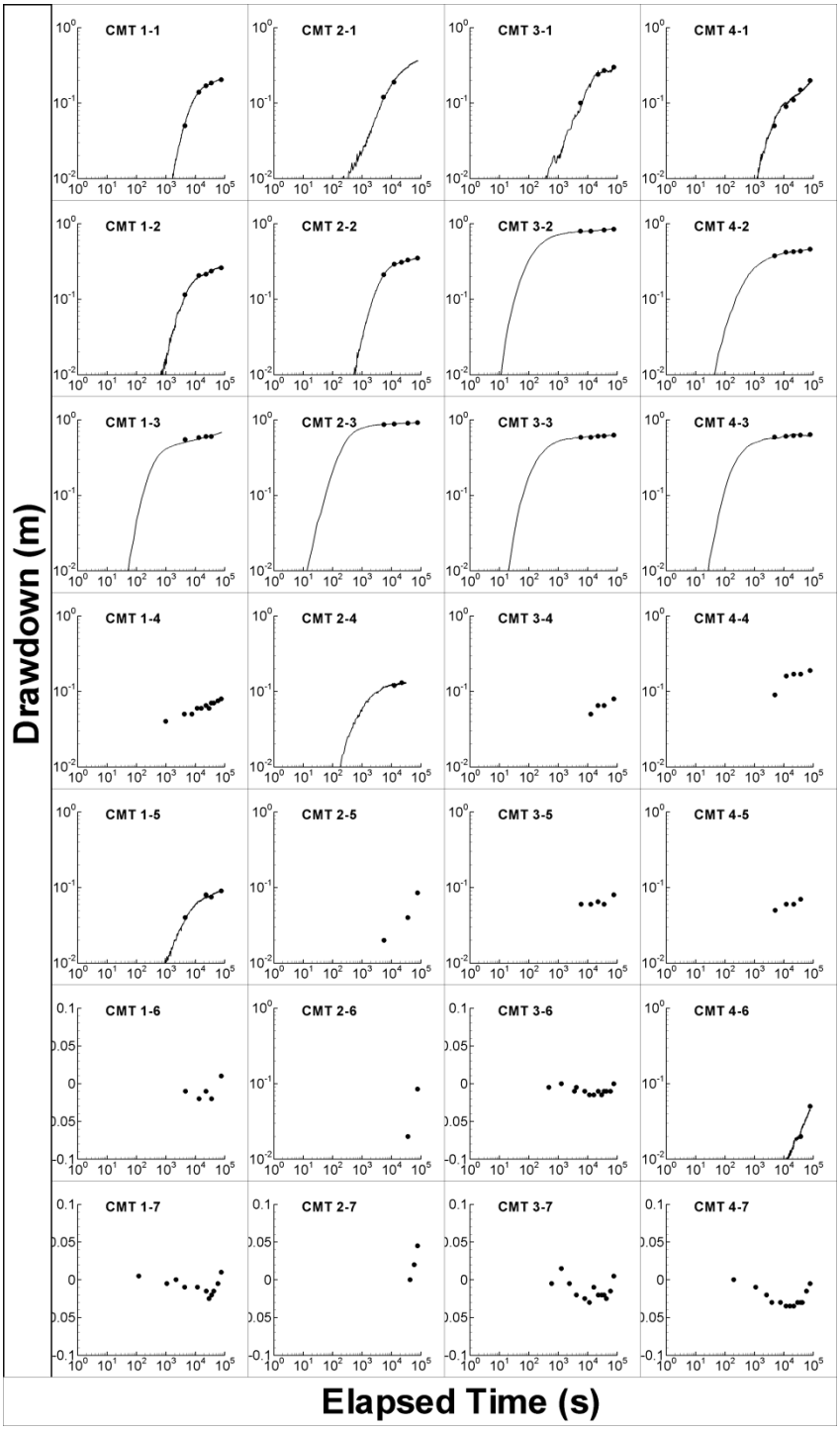


Figure SF6a. Observed drawdown (CMT wells) response for pumping at PW5-3.

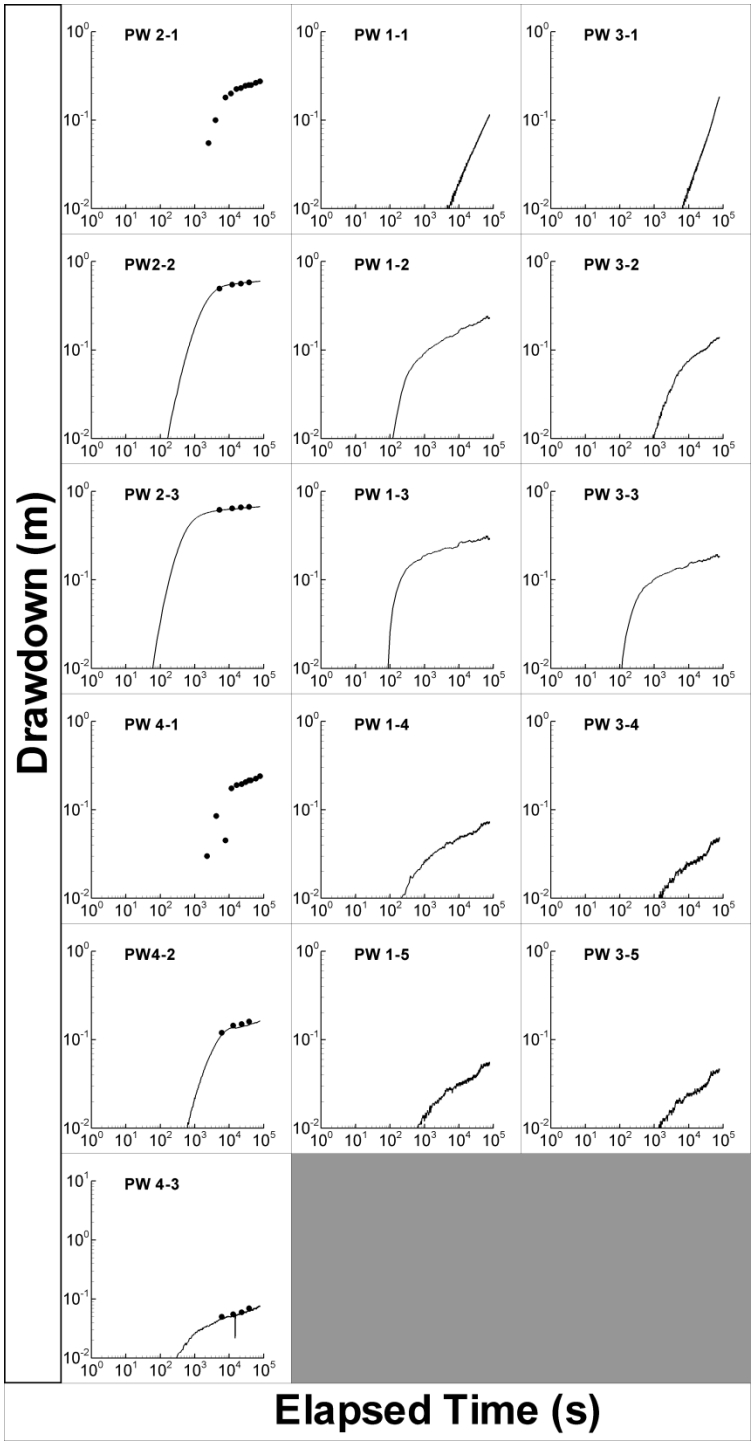


Figure SF6b. Observed drawdown (PW wells) response for pumping at PW5-3.

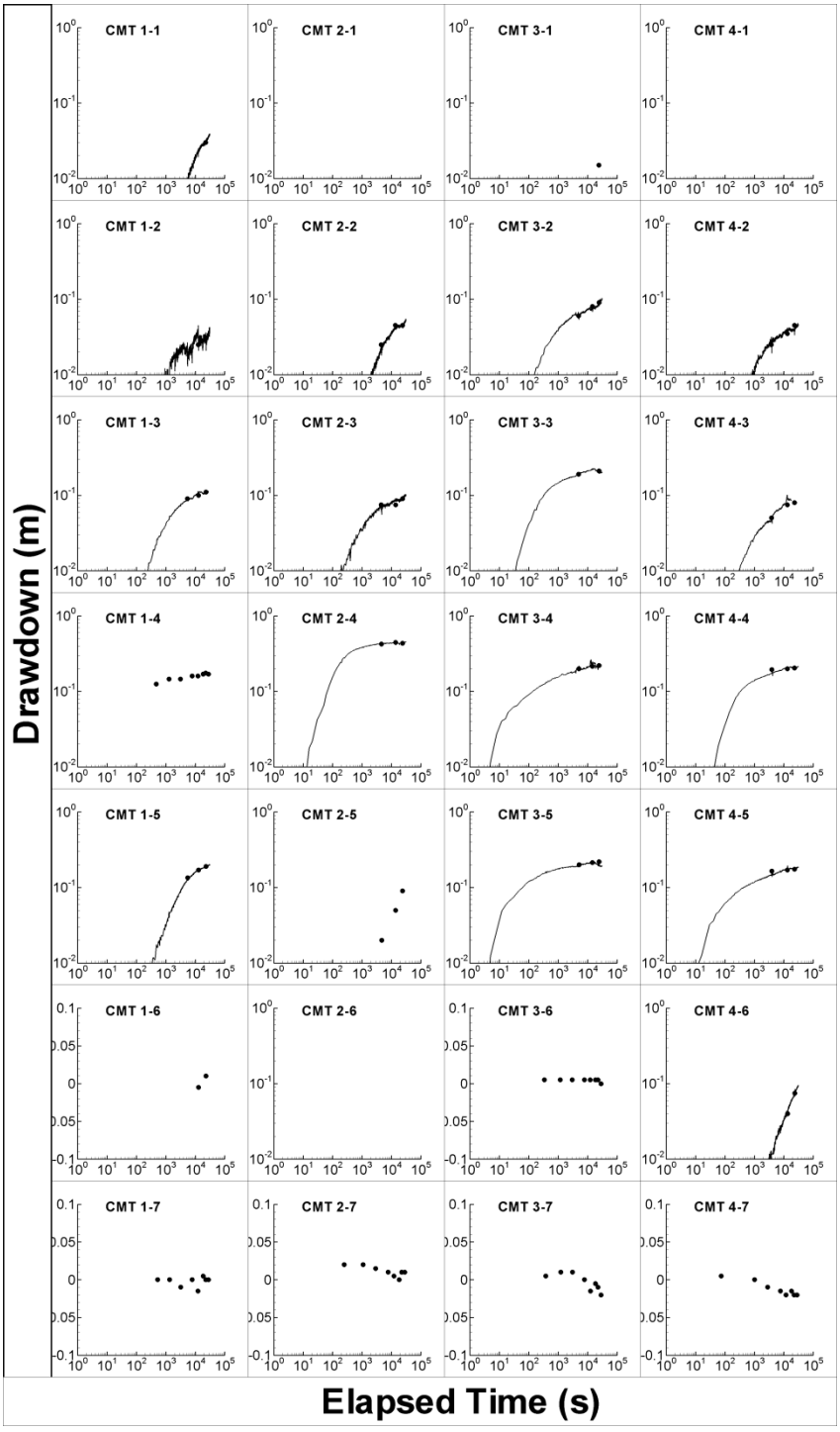


Figure SF7a. Observed drawdown (CMT wells) response for pumping at PW5-4.

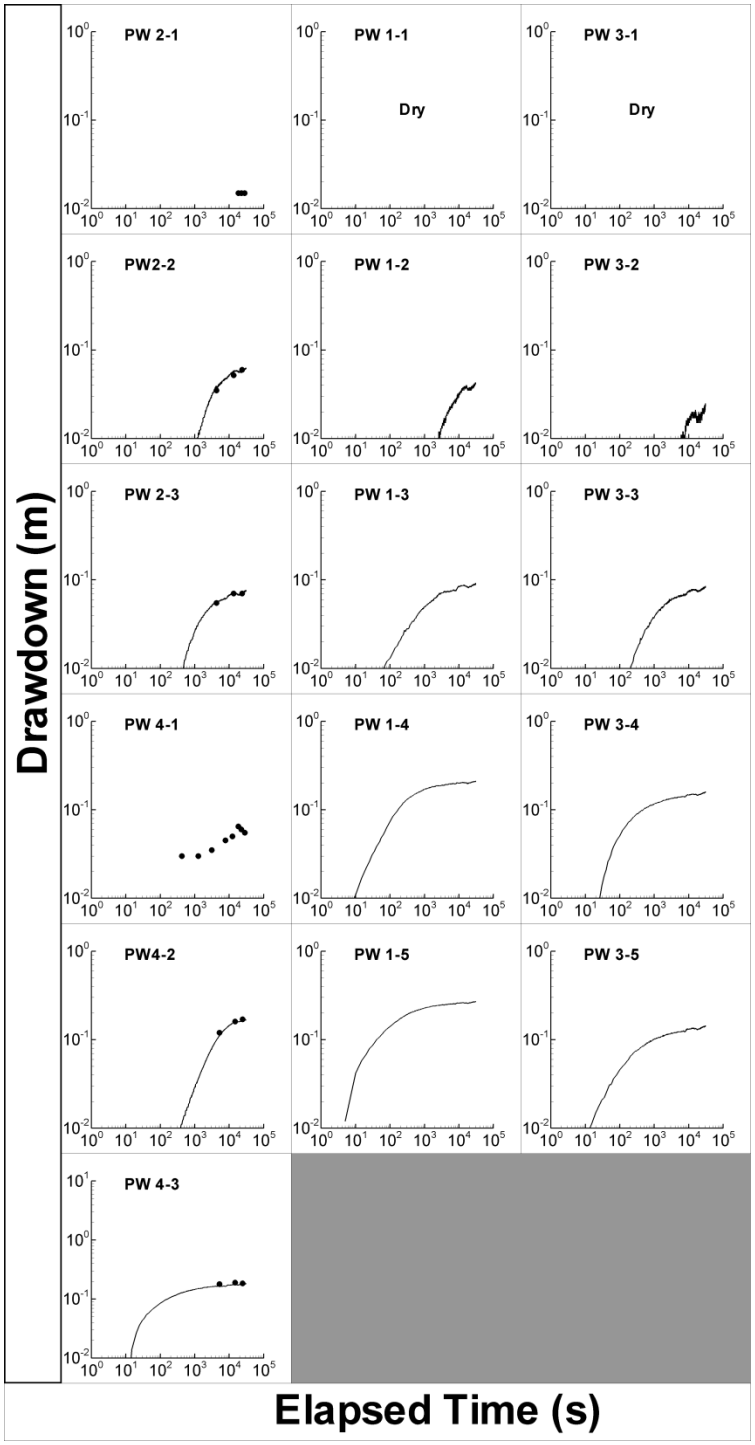


Figure SF7b. Observed drawdown (PW wells) response for pumping at PW5-4.

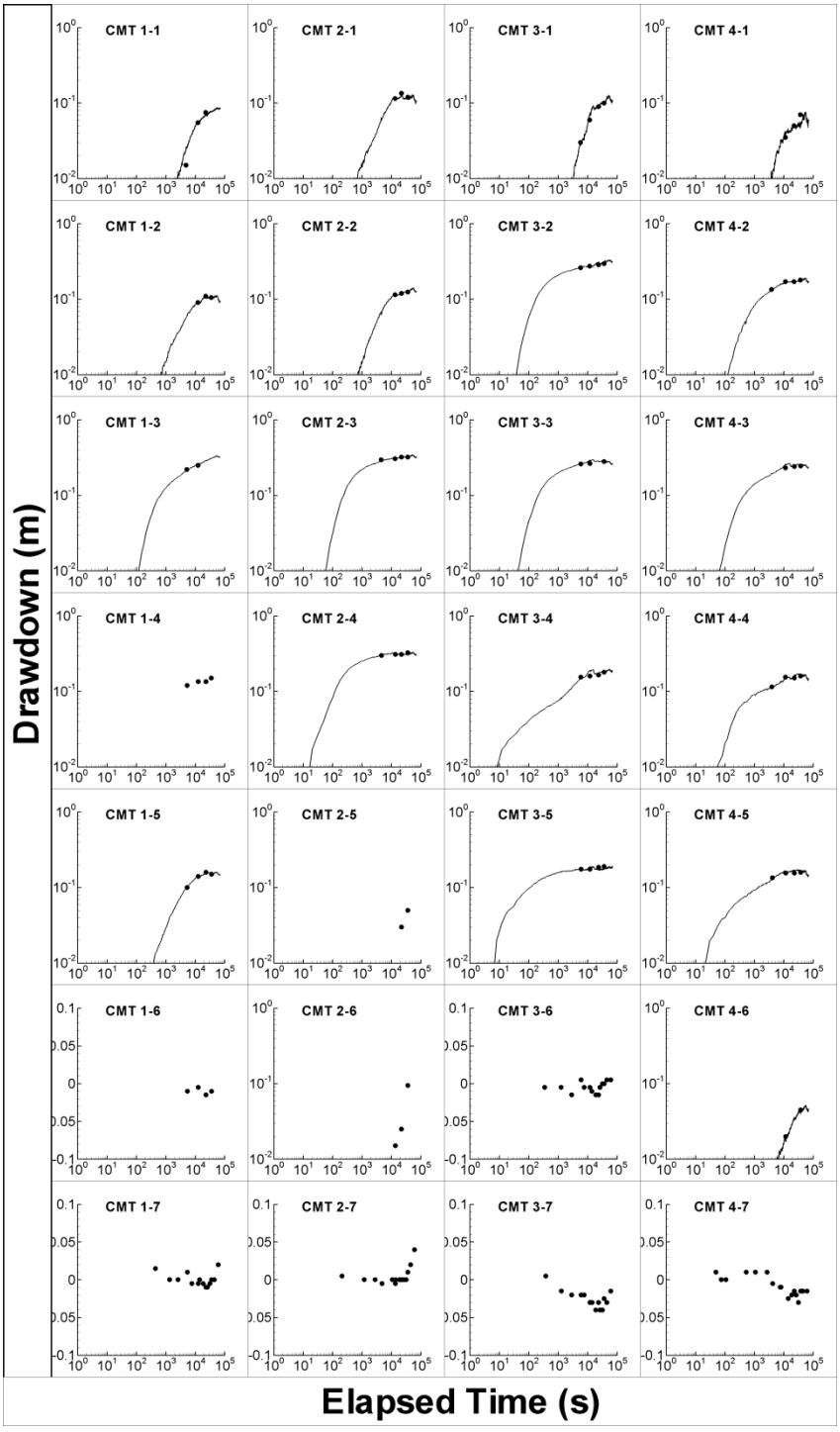


Figure SF8a. Observed drawdown (CMT wells) response for pumping at PW5-5.

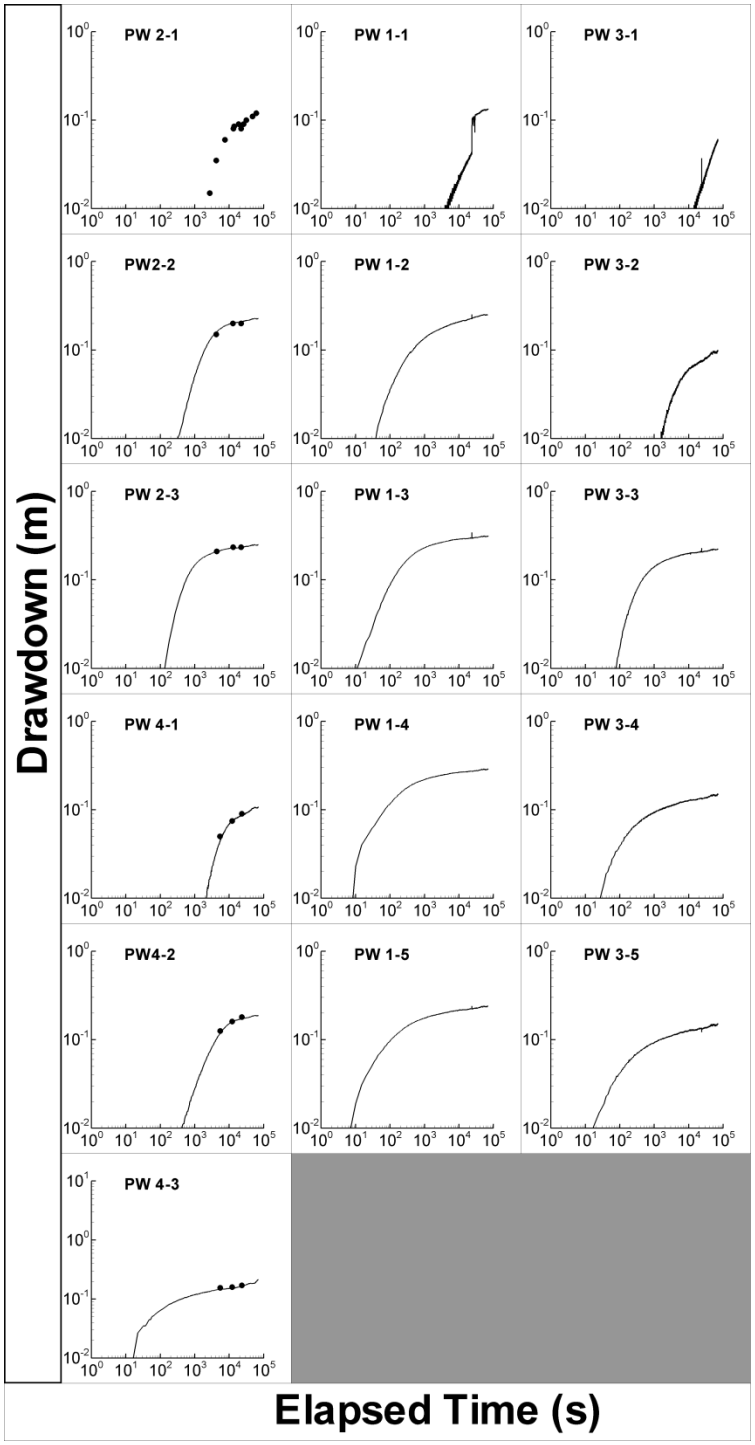


Figure SF8b. Observed drawdown (PW wells) response for pumping at PW5-5.

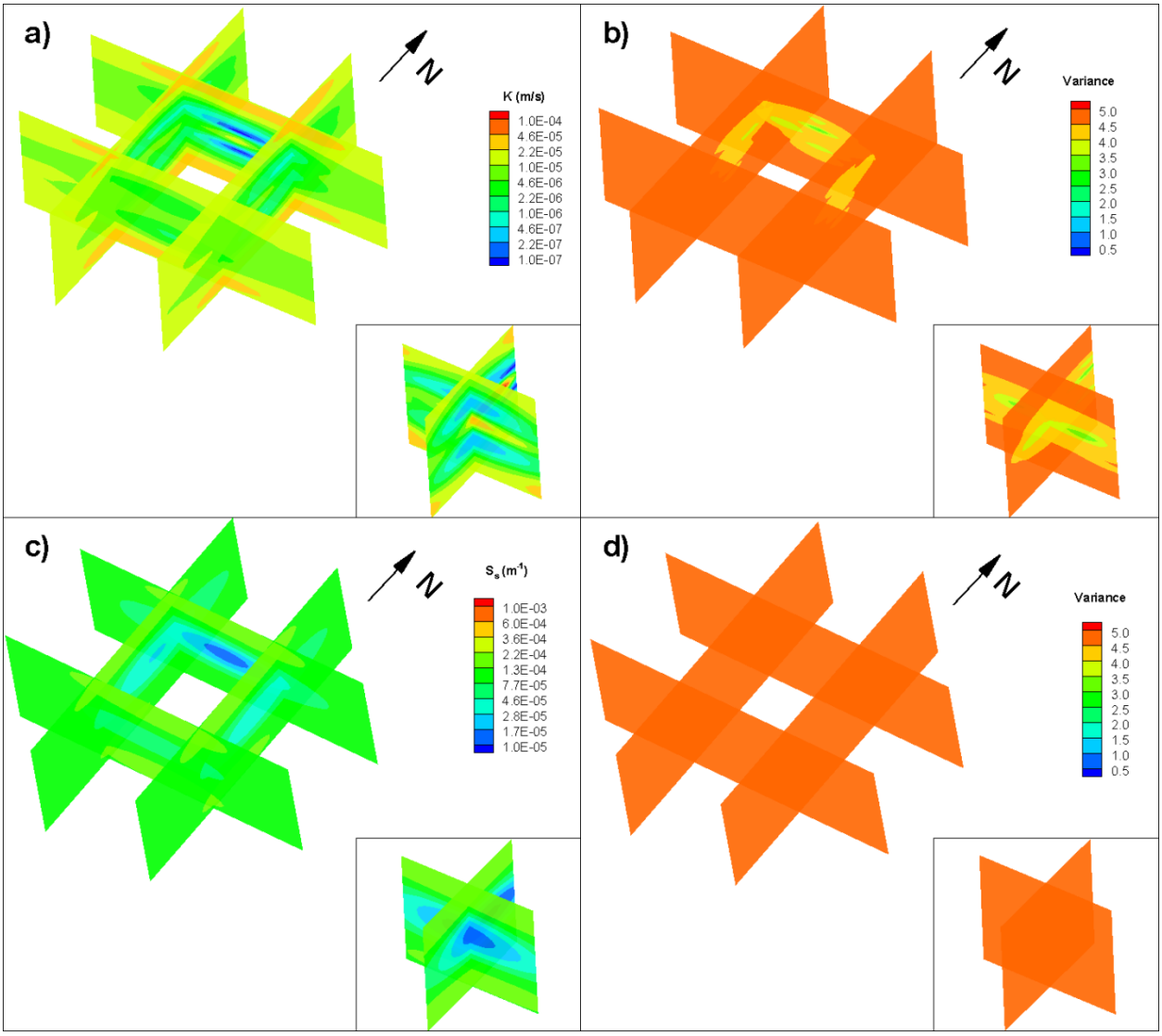


Figure SF9. Single-test inversion K- and S_s-tomograms for the cross-hole test at PW1-4.

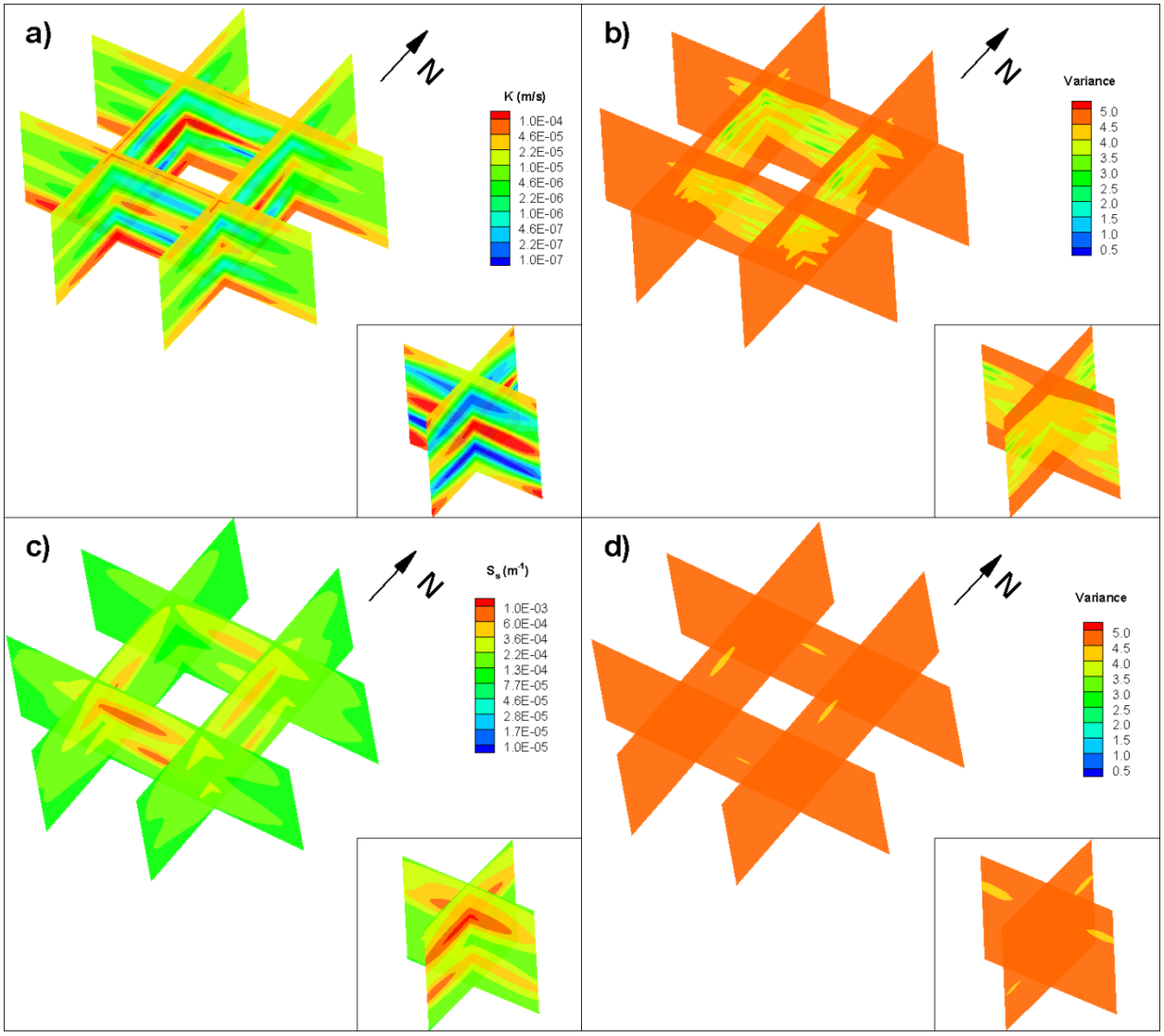


Figure SF10. Single-test inversion K- and S_s-tomograms for the cross-hole test at PW1-5.

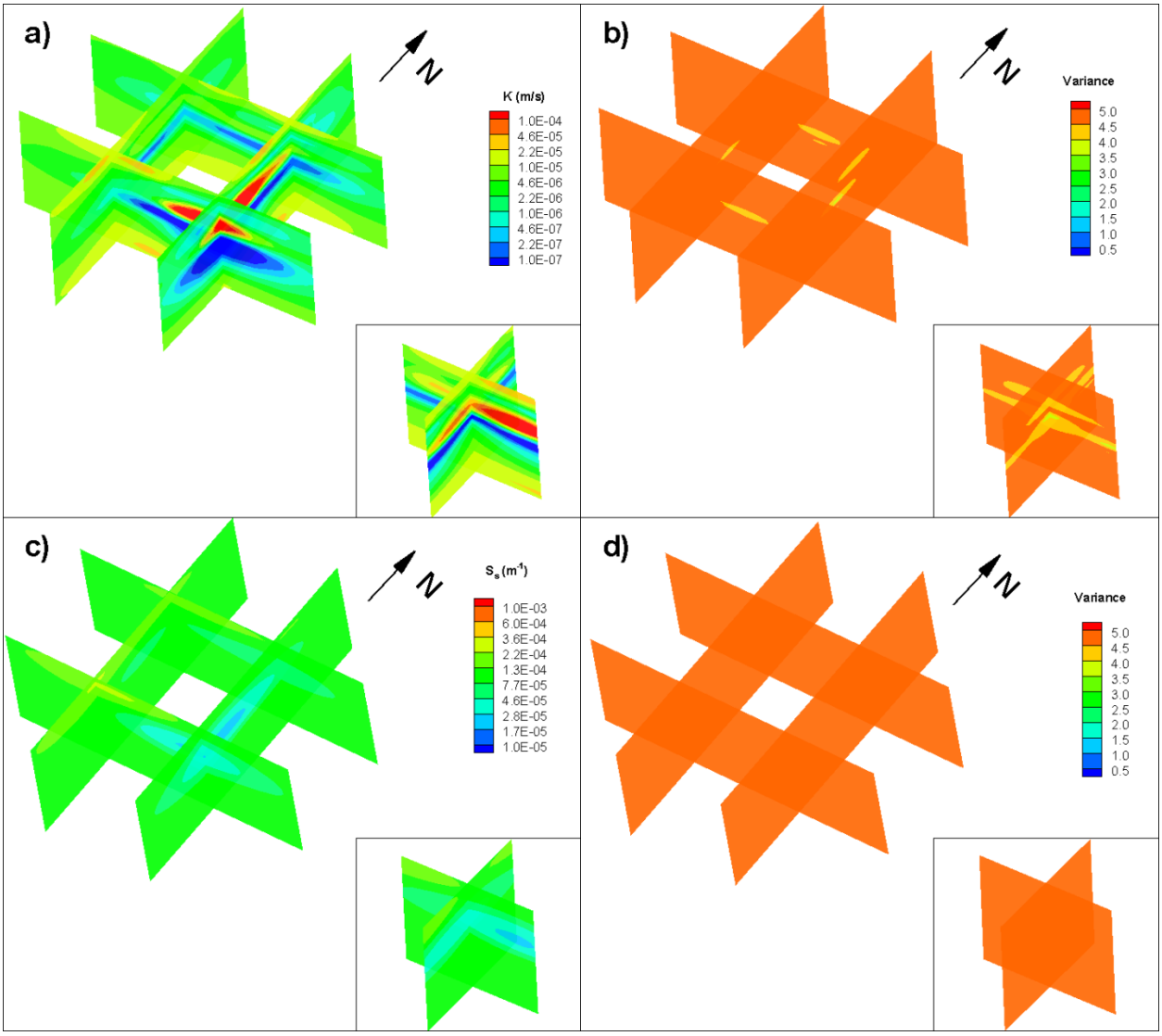


Figure SF11. Single-test inversion *K*- and S_s-tomograms for the cross-hole test at PW3-3.

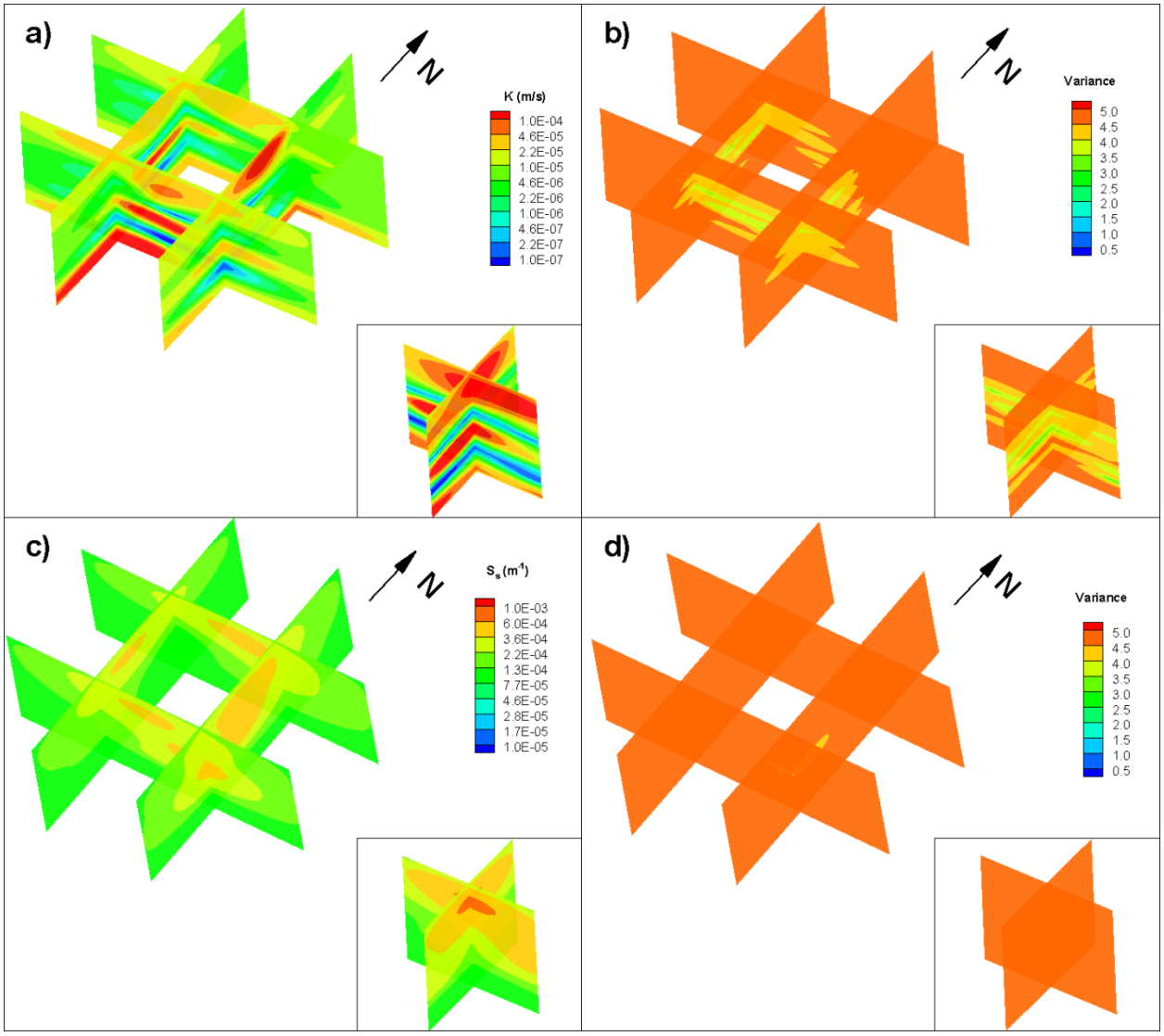


Figure SF12. Single-test inversion *K*- and S_s-tomograms for the cross-hole test at PW3-4.

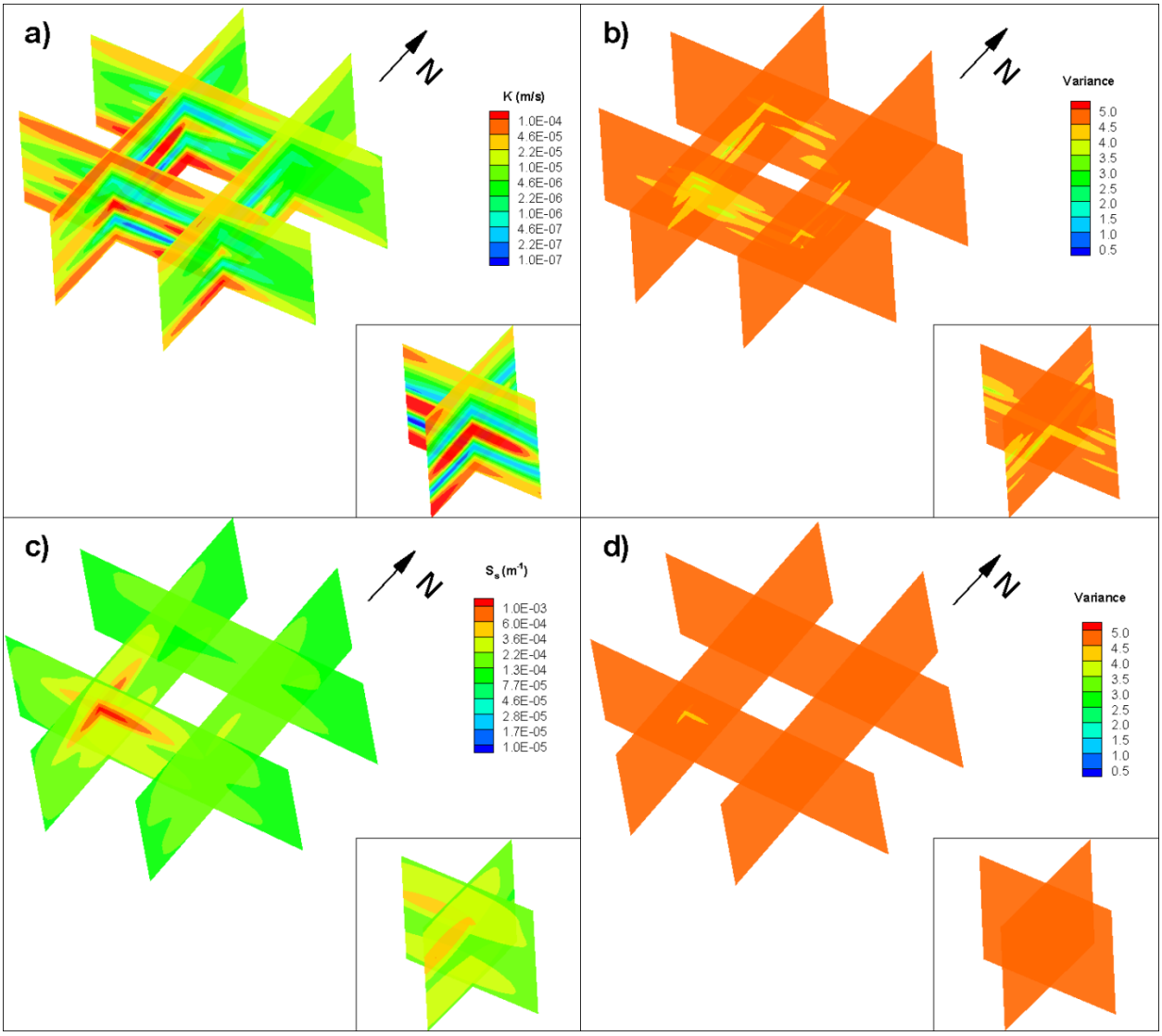


Figure SF13. Single-test inversion *K*- and S_s-tomograms for the cross-hole test at PW4-3.

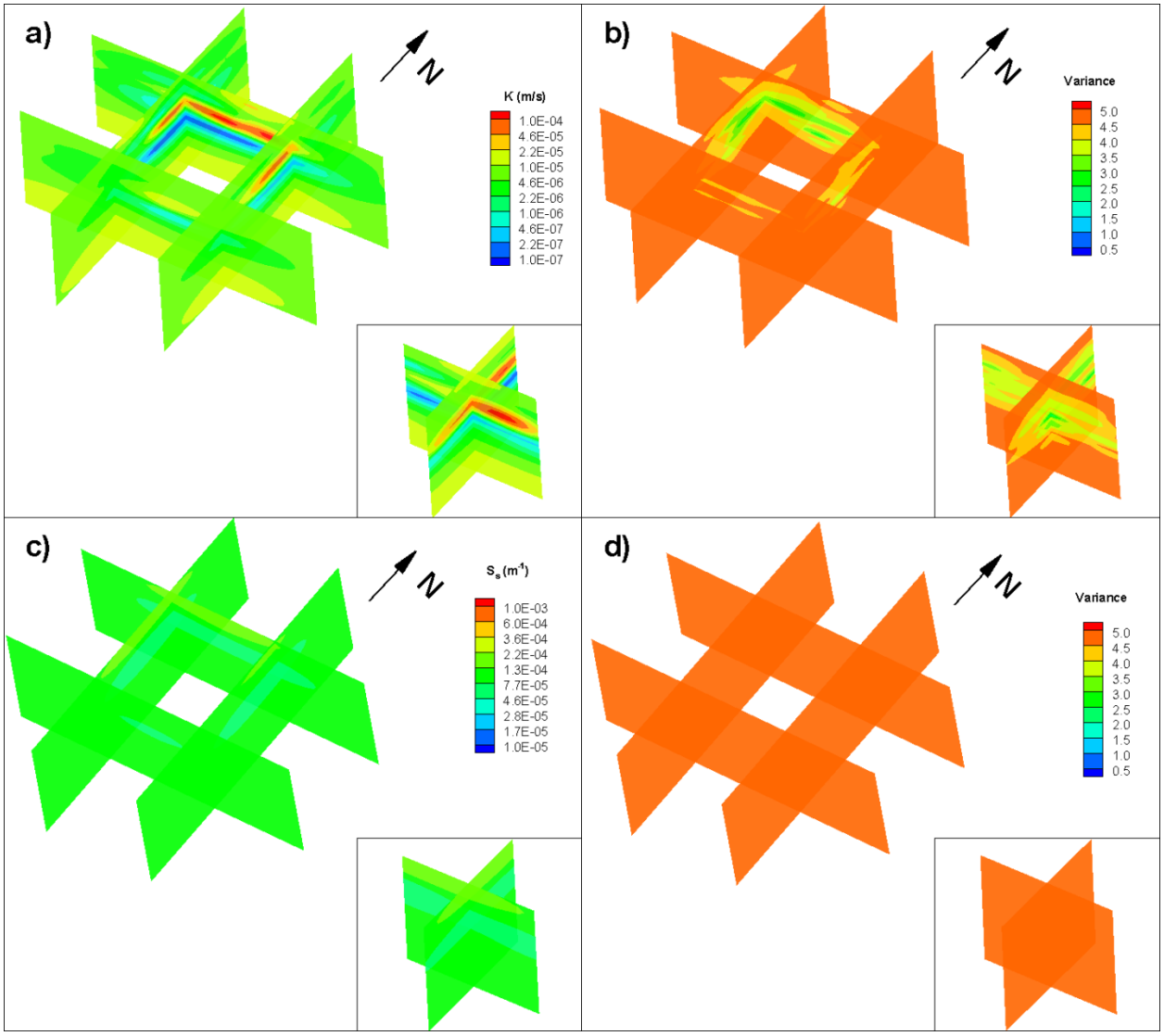


Figure SF14. Single-test inversion K- and S_s-tomograms for the cross-hole test at PW5-3.

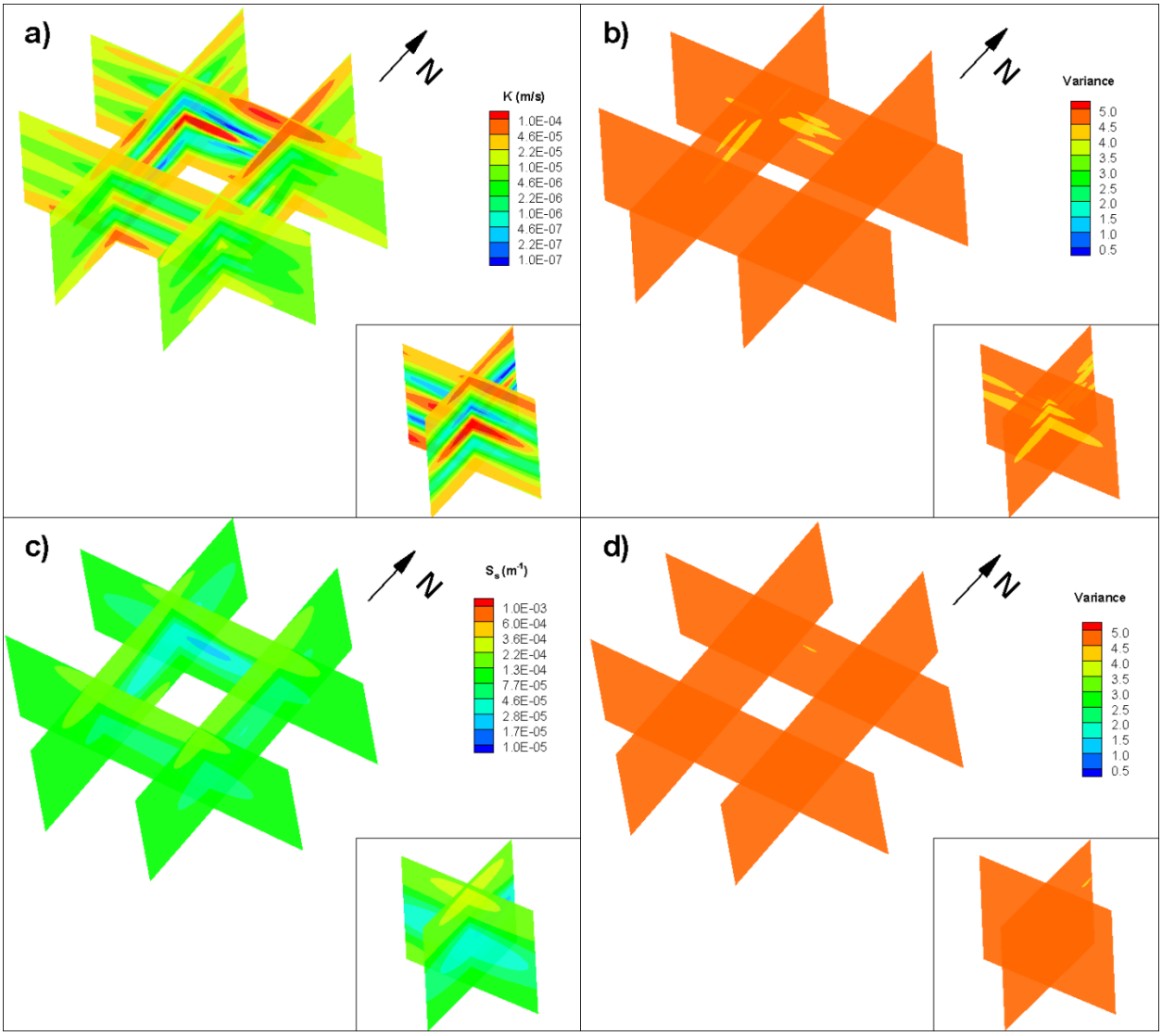


Figure SF15. Single-test inversion *K*- and S_s-tomograms for the cross-hole test at PW5-4.

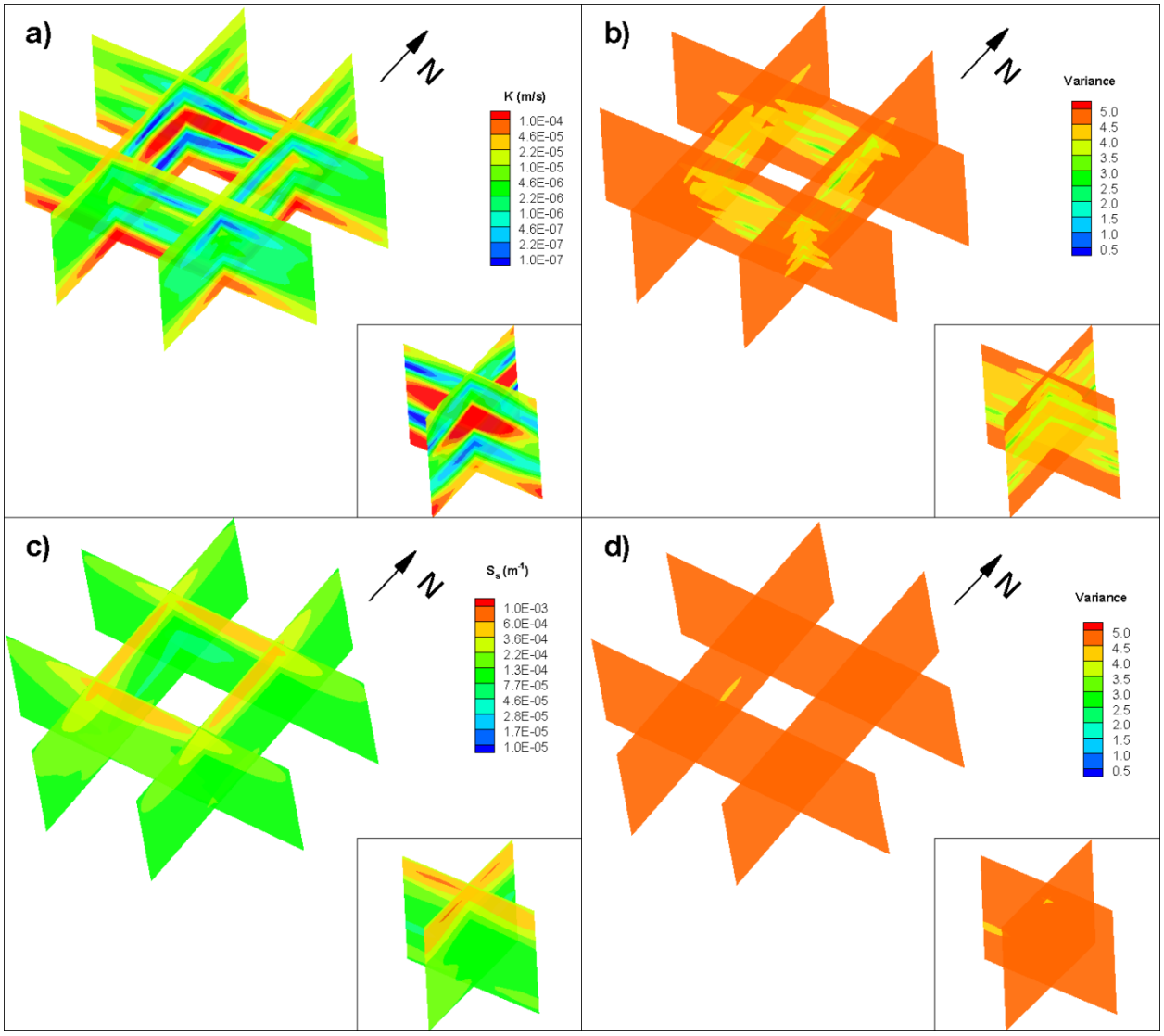


Figure SF16. Single-test inversion K- and S_s-tomograms for the cross-hole test at PW5-5.

Appendix G

Comparison of Subsurface Heterogeneity Modeling Approaches Through Transient Groundwater Modeling in a Highly Heterogeneous Glaciofluvial Aquifer-Aquitard Sequence

This manuscript by *Berg, S.J. and W.A. Illman* will be submitted in the near future.

EXECUTIVE SUMMARY

The accurate delineation of subsurface heterogeneity in hydraulic conductivity (K) and specific storage (S_s) continues to be a topic of great importance. Over the last several decades, numerous approaches have been developed to idealize this heterogeneity. A recently developed tool for imaging K and S_s heterogeneity is hydraulic tomography (HT) which, to date, has not been compared to other methods for modeling subsurface heterogeneity in the field. We attempted this comparison by characterizing the highly heterogeneous glaciofluvial aquifer-aquitard system at a field site located on the University of Waterloo campus, in Waterloo, Ontario, Canada. Site characterization included core sampling at 5 boreholes yielding 471 permeameter K measurements and detailed core logs. Additionally 9 pumping tests were performed at discrete intervals with a 3D network of monitoring points. Using these data, we constructed 5 separate K and S_s distributions that incorporate heterogeneity in various degrees of detail. Methods compared include: 1) effective parameters; 2) ordinary kriging; 3) conditional simulation based on the Transition Probability/Markov Chain Geostatistical method; 4) geological modeling, and 5) transient hydraulic tomography. The resulting K and S_s fields were then evaluated by assessing their ability to capture the salient features at the site, as well as their ability to reproduce 9 pumping tests. Our results show that transient hydraulic

tomography was best able to reproduce these tests, as well as capture the key aquifer features observed by detailed core analysis.

G1. INTRODUCTION

All groundwater investigations require some information regarding aquifer structure and heterogeneity. For relatively homogeneous systems it may be possible to characterize the aquifer using more traditional type curve approaches, such as *Theis* [1935] for confined aquifers, *Hantush* [1960] for semi-confined aquifers, and *Neuman and Witherpsoon* [1969] for multi-aquifer systems. However, rarely do aquifers conform to the assumptions required by type-curve solutions. As such, many approaches for characterizing aquifer heterogeneity have been developed. These include: 1) methods that upscale small-scale hydraulic parameter estimates; 2) those that rely on the interpolation of such small-scale measurements; 3) estimating hydraulic parameters based on the response of the system to stimuli (e.g. geostatistical or stochastic inverse modeling of hydraulic head to infer hydraulic conductivity (K)); and 4) creating stratigraphic or geological models, the layers of which can be calibrated to hydraulic stimuli (e.g., pumping tests). Despite the wide range of approaches available, the selection of the appropriate characterization method, modeling/interpolation method, and data with which to populate these models is an ever present challenge.

Considerable research has been performed to estimate effective (or homogeneous) values of K for heterogeneous aquifer systems (see *Sanchez-Vila* [2006] for a detailed review). These approaches range from averaging point data (e.g. calculating the geometric mean) to interpreting multiple drawdown curves from pumping tests through inverse modeling. While effective parameter estimates may be applicable for larger scale problems, or aquifer systems with a relatively low variance, using homogeneous estimates of K for highly heterogeneous systems, or to predict transport

over short distances can pose a challenge. As such, methods for mapping the heterogeneous distribution of hydraulic parameters are commonly employed.

One of the most common approaches for interpolating small-scale data is to use geostatistical estimators [*de Marsily et al.,* 2005]. These methods often assume that the hydraulic parameters are log-normally distributed and stationary. The user then fits a variogram model (exponential, spherical, etc.,) to the experimental variogram based on observed data (permeameter data, slug tests, etc.). Some geostatistical estimators, such as ordinary kriging, generate smoothed fields that are unable to accurately represent abrupt changes in parameters values (e.g. the contact between two material types). To overcome this, categorical interpolation methods such as indicator kriging [*Journel,* 1983; *Journel and Isaaks*, 1984; *Journel and Alabert*, 1990; *Journel and Gomez-Hernandez*, 1993] and Transition Probability/Markov Chain geostatistical methods [*Carle*, 1999; *Carle and Fogg*, 1997; *Weissmann et al.,* 1999] have been developed. These approaches interpolate categories, as opposed to discrete values, making it possible to reproduce abrupt material changes and juxtapositional tendencies of different layers. For cases where very high resolution conditioning data are available, very realistic hydrofacies models can be constructed [e.g. *Maji and Sudicky*, 2008]

Another approach is to construct geological models based on stratigraphic, or hydrostratigraphic units [e.g., *Koltermann and Gorelick*, 1996; *Martin and Frind*, 1998; *DeMarsily et al*., 2005; and *Alexander et al.*, 2011]. These models are often based on the interpretation of soil cores collected during the installation of wells. While soil cores provide excellent information on the 1D heterogeneity at the location of the well, their collection is expensive and often results in few cores being collected. This can pose a challenge for interpreting the lateral extent of layers at a site. Once constructed, individual layers within these models can either be assigned hydraulic parameters values, or these values can be estimated by coupling a groundwater model with a parameter estimation program and calibrating layer estimates to site data (e.g. pumping test data).

Recently, hydraulic tomography has been introduced as an alternate method for characterizing and modeling aquifer heterogeneity. The first mention of hydraulic tomography that we are aware of was by *Neuman* [1987] who proposed hydraulic tomography as an analogue to geophysical tomography. Since then many researchers have developed inverse methods to analyze multiple pumping tests for the estimation of hydraulic parameters [e.g., *Gottlieb and Dietrich*, 1995; *Snodgrass and Kitanidis*, 1998; *Yeh and Liu*, 2000; *Bohling et al.*, 2002; *Brauchler et al.*, 2003; *McDermott et al.*, 2003; *Li et al.*, 2005; *Zhu and Yeh*, 2005; 2006; *Fienen et al.*, 2008; *Castagna and Bellin*, 2009; *Xiang et al.*, 2009; *Liu and Kitanidis*, 2011].

Many of these methods have been assessed through numerical studies [*Gottlieb and Dietrich*, 1995; *Yeh and Liu*, 2000], laboratory studies [*Liu et al.*, 2002; *Illman et al.*, 2007, 2008, 2010; *Liu et al.*, 2007; *Yin and Illman*, 2009] and field studies [*Bohling et al.*, 2007; *Straface et al.*, 2007; *Li et al.*, 2008; *Cardiff et al.*, 2009; *Illman et al.*, 2009]

While HT has been compared to other heterogeneity characterization or modelling approaches in the laboratory for both steady state hydraulic tomography (SSHT) [*Illman et al.*, 2010] and transient hydraulic tomography (THT) [*Berg and Illman, in review (a)*], HT has yet to be compared to other heterogeneity characterization or modelling techniques in a field setting. Additionally, to our knowledge, there are no studies that compare the ability of different heterogeneity modeling approaches in the field. As such, the goal of this study is to apply several heterogeneity modeling approaches to the highly heterogeneous glaciofluvial aquifer-aquitard sequence at the North Campus Research Site (NCRS) situated on the University of Waterloo campus in Waterloo, Ontario, Canada.

This site was previously characterized by *Alexander et al.* [2011] who interpreted small-scale K data obtained via permeameter and grain size analyses using ordinary kriging. In this study, due to the large number of approaches available for modeling subsurface heterogeneity, we select what we

consider to be a broad sampling of representative and widely used approaches. These include: 1) calculation of effective parameters through model calibration of a pumping test; 2) ordinary kriging of *K* values obtained by performing permeameter tests on soil cores (based on the interpretation of *Alexander et al.* [2011]); 3) conditional simulations based on the Transition Probability/Markov Chain method using soil core logs; 4) construction of a stratigraphic model and its calibration to a single pumping test; and 5) transient hydraulic tomography analysis of 4 pumping tests.

While other studies have examined and compared various inverse modeling strategies [*Zimmerman et al*., 1998; *Franssen et al*., 2009], these studies relied on synthetic true fields, thus allowing the results of the various strategies to be compared to the true fields. However, for this study, the true parameter fields are unknown, thus, we evaluate the various approaches by: 1) comparing the *K* obtained by these various cases to permeameter *K* values measured along 5 boreholes; and 2) simulating 9 pumping tests performed at the field site and comparing the simulated drawdowns to the observed.

G2. SITE DESCRIPTION EXPERIMENTAL DESIGN

The near surface geology at the NCRS is highly heterogeneous and composed of alternating layers of tills and glaciofluvial deposits [*Karrow,* 1993]. Many of these layers are discontinuous and shorter than the width of the study site. Located at a depth of approximately 18 m is the Catfish Creek till, which is a stony, silty, sand till. This unit is assumed to act as a hydraulic barrier [9] and is treated as the base of the site for this study. Further details to the site geology can be found in *Karrow* [1979; 1993], *Sebol* [2000], and *Alexander et al.* [2011].

Between ground surface and the Catfish Creek till, there are two coarser units that act as aquifers. Despite being separated by an 'aquitard' layer, there are many stratigraphic windows that provide a hydraulic connection between the two units [*Alexander et al.*, 2011].

Alexander et al. [2011] analyzed 5 continuous soil cores collected during the installation of pumping and monitoring wells. This analysis consisted of detailed core logging, 471 permeameter tests and, 270 grain size tests to estimate small-scale *K*. The analysis showed that the site is highly heterogeneous, consisting of thin discontinuous beds that can abruptly change material types. The estimated vertical correlation length for the site was approximately 15 cm [*Alexander*, 2009], and *K* was found to vary over 5 orders of magnitude.

Currently there are several other sites in the literature that have been characterized in great detail through the analysis of a large number of small-scale samples [e.g. *Sudicky*, 1986; *Wolf*, 1998; *Rehfeldt et al.*, 1992; and *Sudicky* et al., 2010]. Table G1 summarizes the relevant statistics of *K* for these sites. For all of the sites, anisotropy is evident in the correlation lengths with the horizontal correlation lengths being greater than the vertical. The ln *K* variance ($\sigma^2 \ln K$) ranges from 0.14 for the Cape Cod site [*Wolf*, 1988] to 6.5 for the North Campus Research Site [*Alexander et al.*, 2011]. The collection of such detailed data sets requires considerable efforts for both collection and analysis. The results of this data are then used create heterogeneous models of hydraulic parameters which can then be used to make predictions of groundwater flow and contaminant transport at these sites.

Table G1: Summary of statistical properties for highly sampled heterogeneous sites.

	n	Geometric Mean, K_G (m/s)	Minimum	Maximum	$\sigma^2 \ln K$	λ_1	λ_2	λ_3	Source
Cape Cod	825	3.5×10^{-4}			0.14	2.4	1.4	0.26	*Wolf*, 1988
Borden	1279	7.2×10^{-4}	6×10^{-6}	2×10^{-4}	0.29	2.8	2.8	0.12	*Sudicky*, 1986
Columbus	2187	5.5×10^{-4}	4.1×10^{-7}	1.5×10^{-2}	3.5-5.6	7.5-22.6		1-2.3	*Rehfeldt et al.*, 1992
North Bay	1878	3.5×10^{-5}	1.4×10^{-7}	4.2×10^{-4}	2	17.19	7.39	1	*Sudicky et al.*, 2010
NCRS	471	3.0×10^{-7}	5.8×10^{-10}	2.8×10^{-4}	6.5	N/A*		0.15	*Alexander et al.*, 2011
NCRS (THT)	29791	4.5×10^{-7}	1.7×10^{-8}	1.6×10^{-3}	4.1	6.3	6.3	1.8	*Berg et al., in review (b)***

*Due to the limited number of horizontal locations, *Alexander et al*, [2011] were unable to determine horizontal correlation lengths. A correlation length 4 m was obtained when an omni-directional variogram was fit to the data.

**Case 2 of *Berg et al.* [*in review*] (b)

Four continuous multichannel tubing (CMT) wells containing 7 channels each were installed at the NCRS as observation wells. The screens were spaced 2 m apart with the upper most screens located between 4.5 and 5.5 m below ground surface (mbgs), and the deepest screens set 16.5 to 17.5 mbgs. Five pumping wells were installed, 3 of which were multi-screened wells (10 cm diameter); two screened at 5 different elevations (PW3 and PW5) and 1 screened at 8 different elevations (PW1). The screened intervals were 1 m long and were separated from adjacent screens by 1 m of solid pipe. PW1 extended to approximately 18 mbgs; and, PW3 and PW5 to 12 mbgs. The remaining two pumping wells, PW2 and PW4 (5 cm diameter), were well nests consisting of 3 separate wells, each screened over a 1 m interval. Screen elevations for PW2 were 4, 7, and 8 mbgs; and, screen elevations for PW4 were 5, 8.5, and 11.5 mbgs.

The well array was set up in a 9-spot square pattern, measuring 15 m × 15 m × 18 m deep. Figure G1 shows a plan view of the NCRS and Figure G2 shows a 3D schematic of the wells.

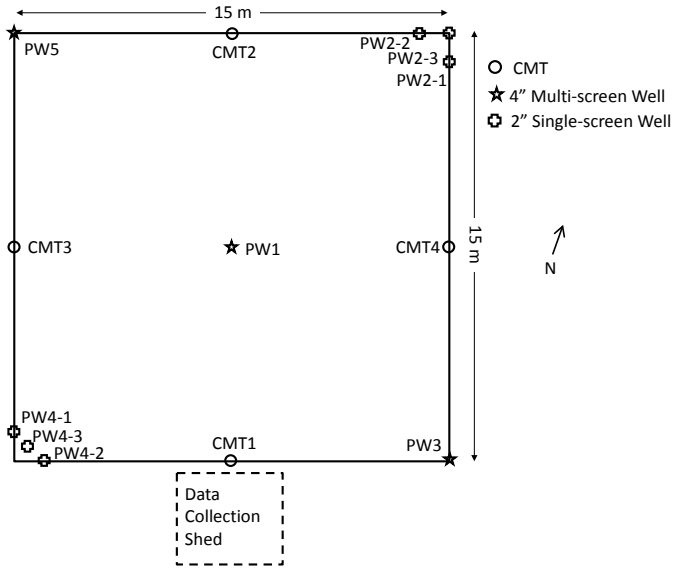


Figure G1: Plan view of the NCRS [*after Berg and Illman, in review(b)*].

In total, 9 pumping tests that ranged in duration from 6.5 hrs to 22 hrs were performed at the site. Table G2 provides a summary of the pumping tests including the volumetric flow rate, duration, maximum observed drawdown, and the port at which the maximum drawdown was observed. A more in-depth discussion of site geology and instrumentation can be found in *Alexander et al.* [2011] and *Berg and Illman* [*in review (b)*]

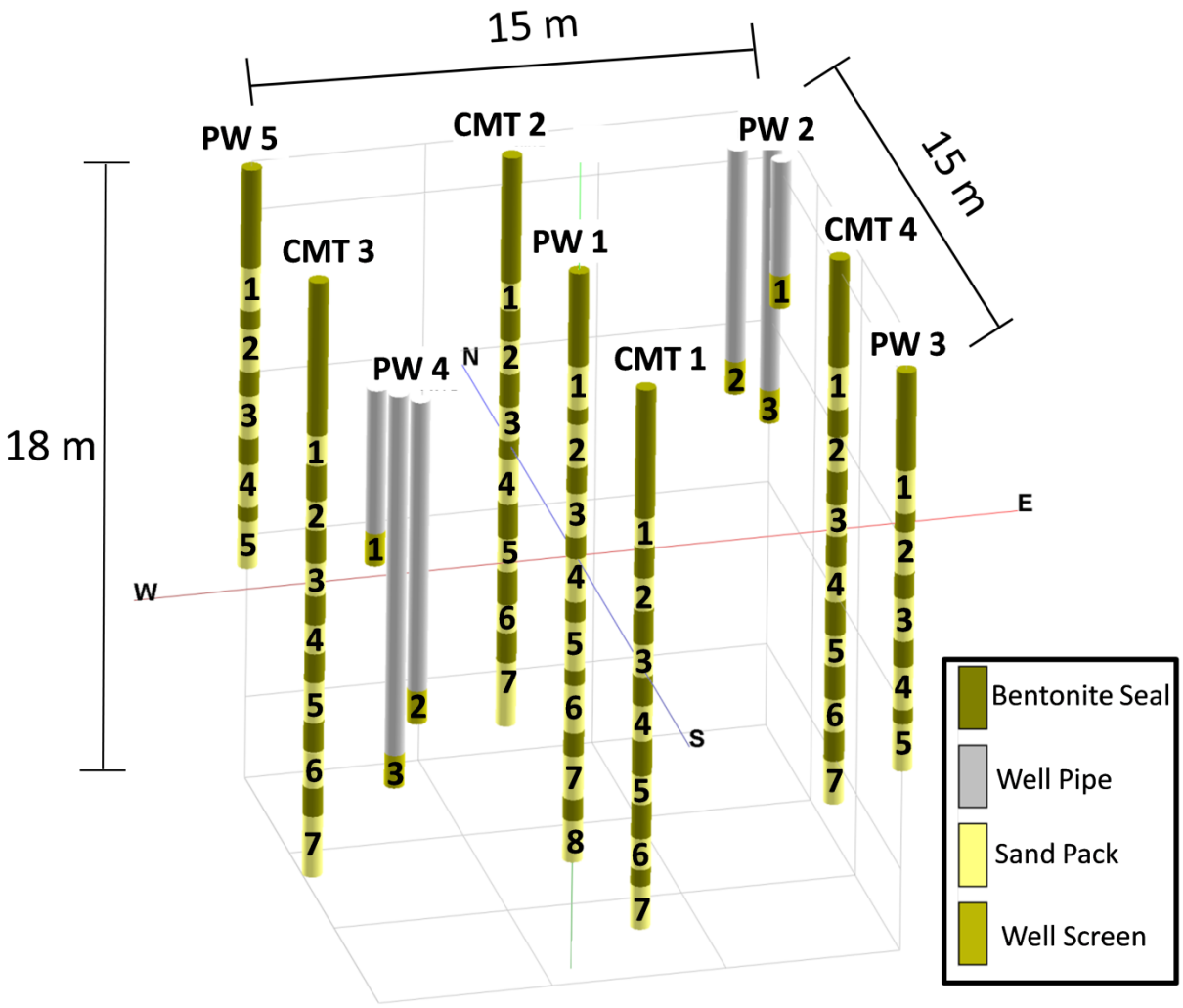


Figure G2: Three-dimensional perspective view of various wells and corresponding intervals at the NCRS [*after Berg and Illman, in review(b)*].

Table G2: Summary of pumping tests performed at the NCRS [*after Berg and Illman, in review (b)*].

Pumped Location	Flow Rate (L/min)	Duration (hrs)	Maximum Observed Drawdown (m)	Port of Maximum Drawdown
PW1-3	10.5	6	1.10	CMT 4-3
PW1-4	6.3	8.5	0.98	CMT 2-4
PW1-5	4.4	22.5	0.18	CMT 2-4
PW3-3	2.1	22	0.57	CMT 1-3
PW3-4	1.5	22	0.08	CMT 1-5
PW4-3	30.2	22.5	0.88	CMT 3-4
PW5-3	7.8	22	0.92	CMT 2-3
PW5-4	7.8	8.5	0.45	CMT2-4
PW5-5	8.1	22	0.32	CMT2-3

G3. SUBSURFACE HETEROGENEITY MODELING APPROACHES

G3.1 Estimation of effective parameters

The use of analytical solutions such as *Theis* [1935] or *Hantush* [1960] often requires assumptions that cannot be met at a given site because of complexities in the geology or forcing functions that cannot be controlled. To overcome these limitations, we estimated effective values of K and S_s by coupling the parameter estimation code, PEST [*Doherty*, 2005], with the groundwater flow model HydroGeoSphere (HGS) [*Therrien et al.*, 2005] while treating the parameter fields to be homogeneous.

The 3D model domain used for the estimation measured 45 m × 45 m × 15 m and was composed of variably-sized rectangular finite elements. The smallest elements, measuring 0.5 m × 0.5 m × 0.5 m were located in the central portion of the model domain and corresponded to the location of the 15 m × 15 m test plot in plan view. Outside of the test plot, the element size increases with increasing distance from the center of the model domain. In total, the model domain consists of

29,791 elements and 32,768 nodes. The boundary conditions are no-flow for top and bottom boundaries and constant head for the remaining outer boundaries.

The estimation of the equivalent K and S_s was performed by individually matching the simulated drawdown at the 28 CMT ports to that observed during the pumping test performed at PW1-3. Thus, unique equivalent K and S_s estimates were obtained for each observation port that experienced drawdown. The equivalent K values from 13 intervals ranged between 2.1×10^{-6} m/s and 1.1×10^{-5} m/s with a geometric mean of 6.2×10^{-6} m/s and a ln K variance of 0.49. Likewise, the equivalent S_s values from 8 intervals ranged between 4.3×10^{-6} m/s and 1.7×10^{-3} m/s with a geometric mean of 1.2×10^{-4} m^{-1} and a ln S_s variance ($\sigma^2 \ln S_s$) of 3.2 We treat the geometric mean of these individual estimates as the effective K and S_s values for this site.

G3.2 Geostatistical Analysis and Ordinary Kriging of Local Point Scale *K* Data

A commonly applied method for interpolating small-scale data is ordinary kriging. For this case, we used the kriged permeameter K field of *Alexander et al.*, [2011]. Based on the analysis of 471 permeameter tests performed on soil cores collected during the installation of PW1, and CMTs 1 through 4, *Alexander et al.* [2011] determined that the K data were both log-normally distributed and that the mean and variance within the investigation area were stationary. *Alexander et al.* [2011] kriged the permeameter K data by fitting an exponential omni-directional variogram to the experimental variogram for 5 boreholes. Figure G3 shows the location at which the samples for the permeameter analyses were collected. Figure G4 shows the corresponding kriged K field with the square created by the 4 slices corresponding to the outer edges of the 15 m x 15 m field plot. The inset image is a cross-section of the inner square of the main figure. This corresponds to cross-sections from CMT1 to CMT2, and CMT3 and CMT4. The kriged K field reveals considerable detail in the region that corresponds with the field site, however, due to the lack of information outside of

this area, there is considerable smoothing near the boundaries. The main features of this field are: 1) the presence of a two aquifers separated by low *K* material; and 2) the bottom portion of the model domain consists of very low *K* material ($< 1 \times 10^{-7}$ m/s).

For the purposes of creating forward simulations of the pumping tests (described later), the kriged parameter field was transferred to a model domain similar to the one used to for the effective parameter estimates (section 3.1) except the elements were uniform for this case, measuring 0.5 m × 0.5 m × 0.5 m (243,000 elements and 256,711).

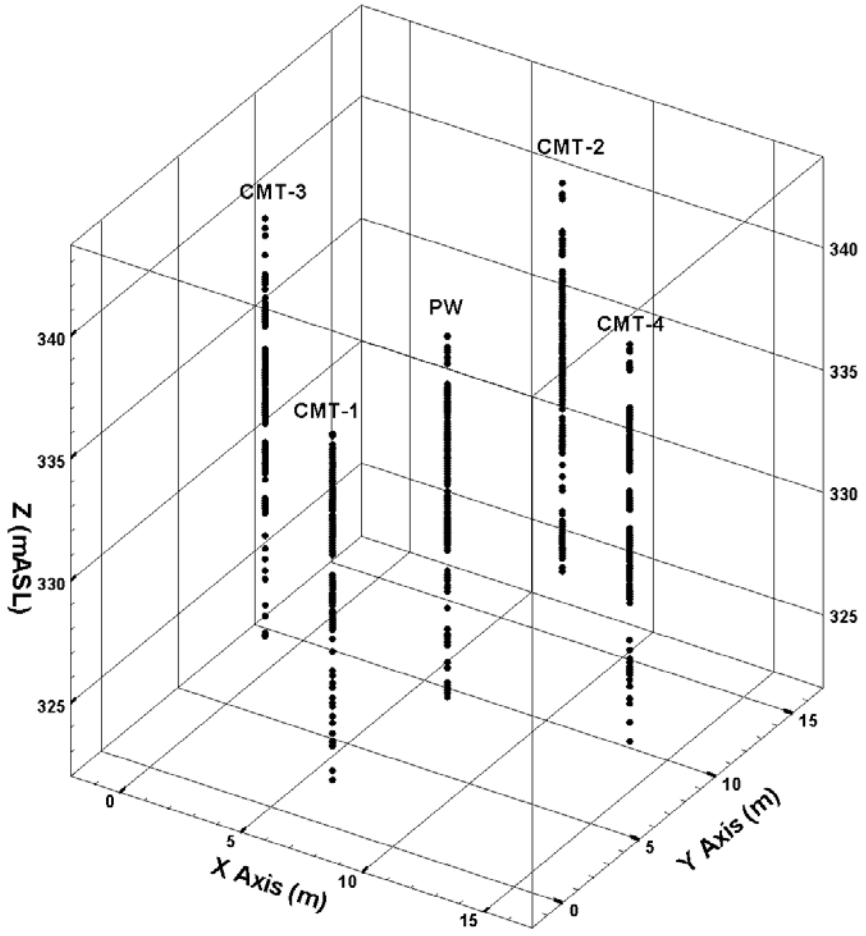


Figure G3: Location of core samples used for permeameter analysis to create the kriged *K* field in Figure 4 (after *Alexander*, 2009).

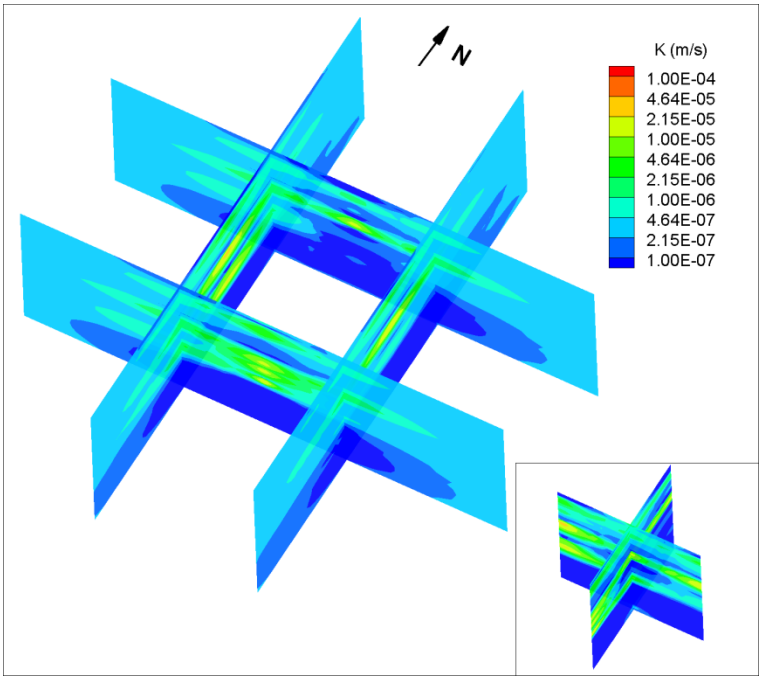


Figure G4: Three-dimensional perspective view of the kriged permeameter *K* field. The square created by the slices represents the outer edge of the 15 m x 15 m plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

G3.3. Transitional Probability/Markov Chain Probability Geostatistical Approach

The 8 material types identified by *Alexander et al.* [2011] (sandy Gravel, Sand, Sand and Silt, sandy Silt, Silt, clayey Silt, Silt and Clay, and silty Clay) were simplified to 4 (sandy gravel, sand, silt, and silty clay) for the purposes of using the transition probability/Markov Chain geostatistical approach to create conditional realizations of the site [*Carle*, 1999; *Carle and Fogg*, 1997; *Weissmann et al.*, 1999]. The reduction of categories was performed based on grouping categories with similar *K* (e.g. clayey Silt, Silt and Clay, and silty Clay). It was also found that reducing the number of categories improved the simulated to observed matches of the transition probabilities. Using the data along boreholes PW1, CMT1, CMT2, CMT3, and CMT4 the bi-directional (horizontal and vertical) experimental transition probabilities were computed and matched automatically with a

Markov chain model using TPROGs [*Carle*, 1999]. Figures G5 and G6 are the observed and simulated Markov Chain transition probabilities for the horizontal plane and z-direction, respectively. The data density in the horizontal direction (Figure G5) was quite low as a result of the low density of observation points in the horizontal direction, thus it was difficult to assess the quality of the match for many of the ports. However, the transition probabilities fit by the Markov Chain appear to capture the main trends in the data for each plot. For the vertical direction, (Figure G6) the data density was considerably higher and the Markov Chain model was able to accurately capture the observed transition probabilities for many of the material types.

The observed locations of the material categories along the boreholes were then used to condition the simulation which uses the modeled Markov chain transition probabilities and simulated annealing to generate multiple realizations of the model domain [*Carle*, 1999]. The same model domain as that described for the kriged case in section 3.2 was used and a total of 10 realizations were created. Figure G7 shows an example realization with (a) K and (b) S_s values assigned for each material type. K values were assigned by taking the geometric mean of permeameter samples that corresponded with each material type. The material types were assigned K values of 1.0×10^{-8} m/s, 8.7×10^{-7} m/s, 8.8×10^{-6} m/s, and 4.6×10^{-5} m/s for silty clay, silt, sand and sandy gravel respectively. Because of the difficulty in estimating S_s from small scale samples, estimates of S_s for the different material types were assigned based on the individual estimates obtained through the parameter estimation procedure described in section 3.1. Each observation port was categorized based on the predominant material surrounding it, and the geometric mean of multiple S_s values for each material type was taken. The assigned S_s values were 1.4×10^{-4} m/s, 2.2×10^{-4} m/s, 1.1×10^{-4} m/s, and 1.0×10^{-4} m/s for silty clay, silt, sand and sandy gravel respectively.

Since each of the 10 realizations was equally probable, the selection of a single realization posed a challenge. Thus, the pumping test performed at PW1-3 was simulated in each of these

realizations and compared to the observed drawdowns. The realization that performed the best was selected for comparison purposes that we describe later. Figure G8 shows a scatterplot of observed versus simulated drawdowns at 5 different times throughout the duration of the pumping tests for all observation ports. While all realizations appear to perform poorly, realization R3 performs marginally better than the rest based on the best fit line through the data points. Realization R3 had a slope of linear model fit closest to 1, and the R^2 and intercept values were comparable to the other tests. Thus, this case (R3) was selected for comparison purposes.

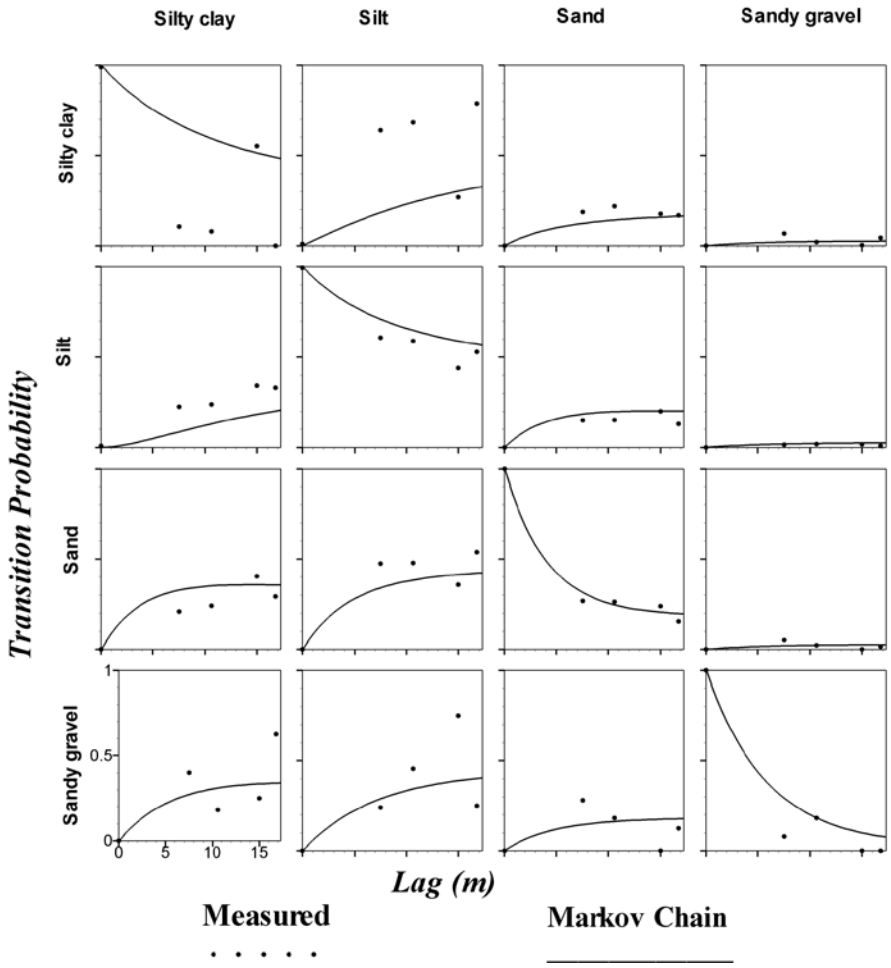


Figure G5: Transition probability matrix for the horizontal plane. The dots are the measured transition probabilities and the solid line of the data fit by the Markov chain.

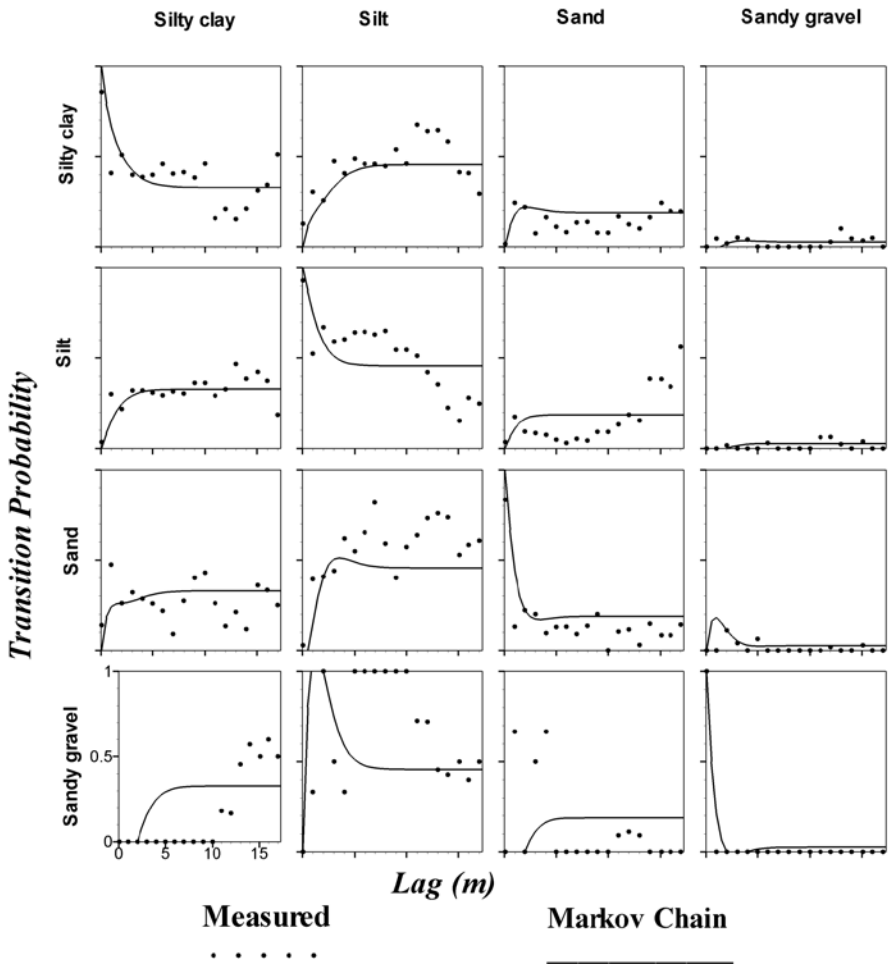


Figure G6: Transition probability matrix in the z-direction. The dots are the measured transition probabilities and the solid line of the data fit by the Markov chain.

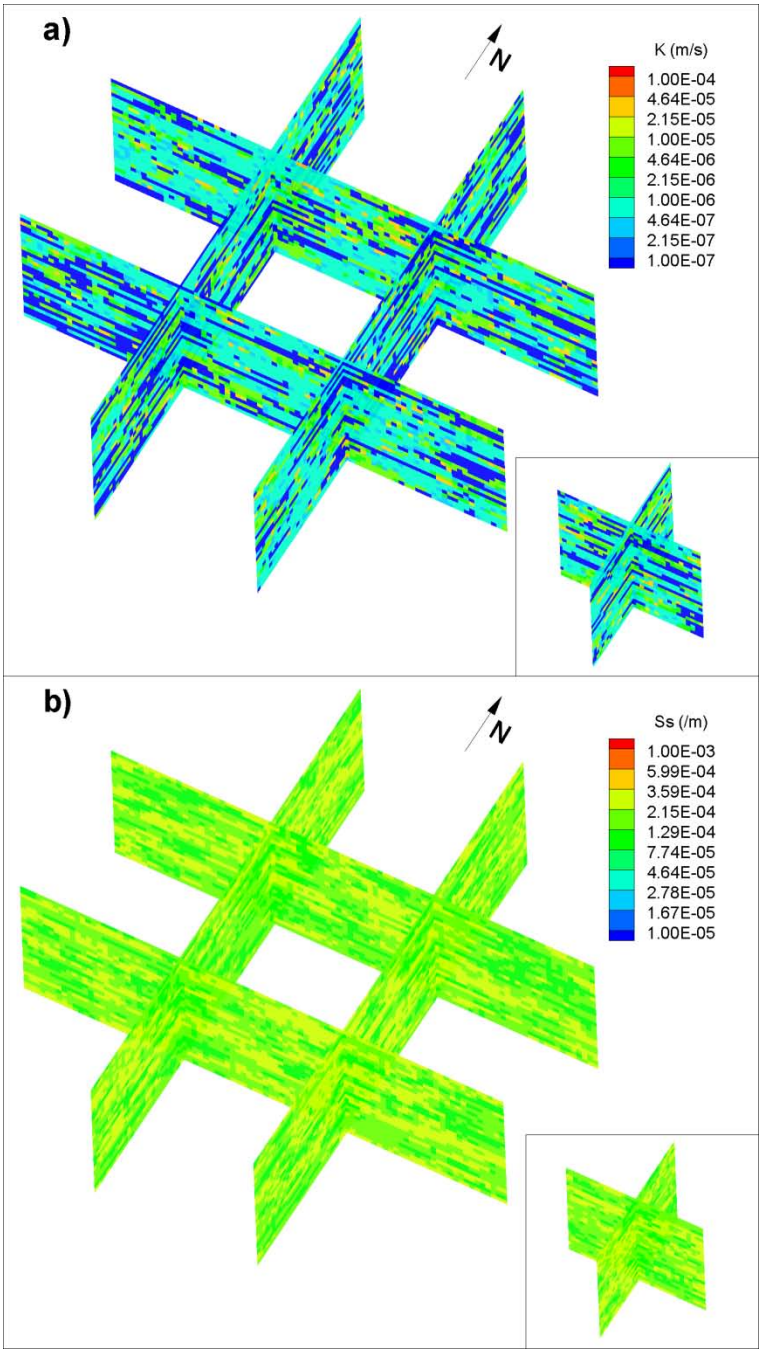


Figure G7: Three-dimensional perspective view of a single realization of K and S_s generated using the TPROGS code. The square created by the slices represents the outer edge of the 15 m x 15 m plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

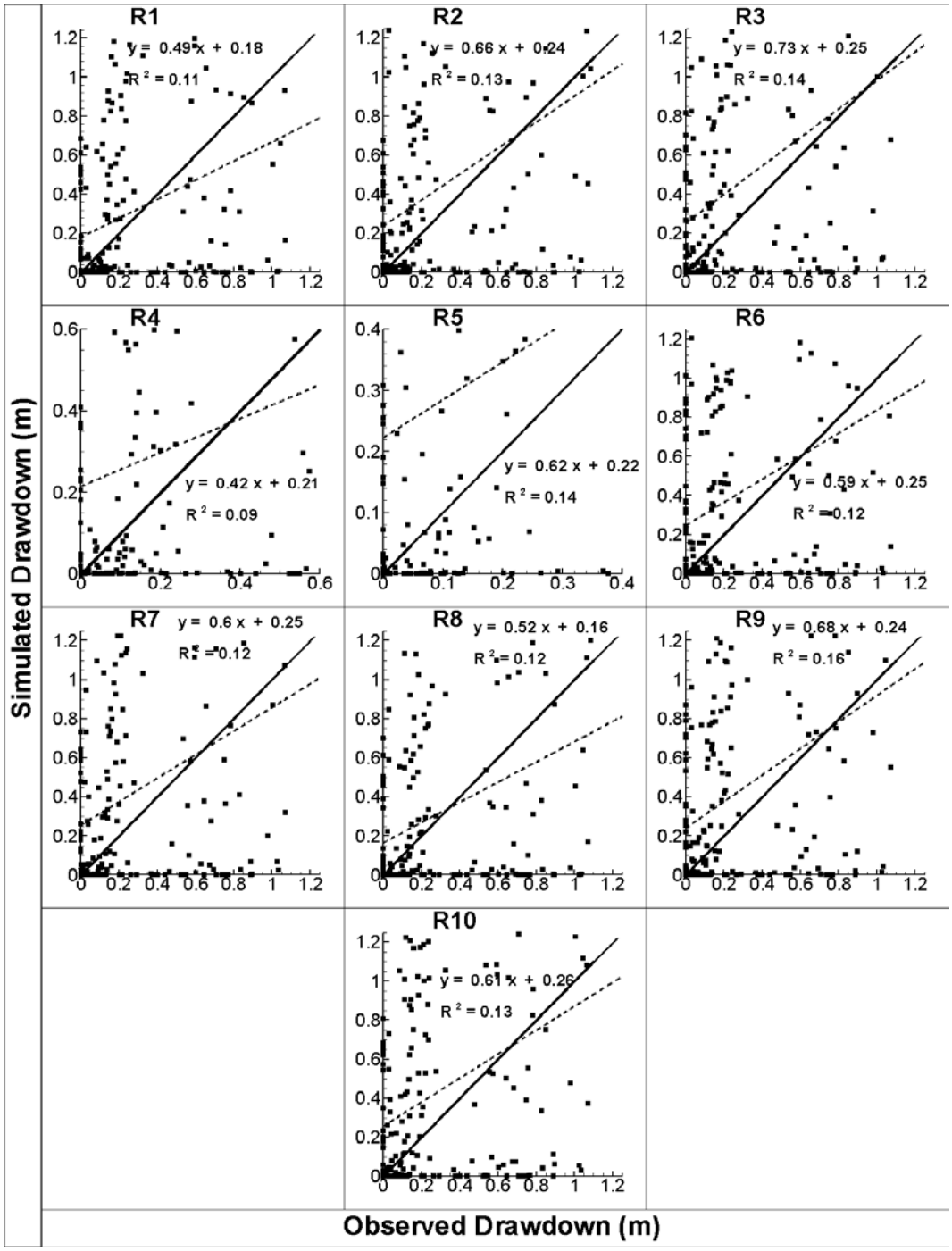


Figure G8: Observed vs. simulated drawdown for each of the 10 TPROGs realizations. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

G3.4.Calibrated Stratigraphic Model

A stratigraphic model was created based on the soil classification performed by *Alexander et al.* [2011] by analyzing the soil cores collected along PW1, CMT1, 2, 3, and 4. In total, 8 soil classifications were assigned, ranging from silty clay to sandy gravel. The interpolation between boreholes was performed using an inverse distance method built into RockWorks 2006 [RockWare, 2007]. The model domain for the inverse distance estimation was identical to that for sections 3.2 and 3.3. We then automatically calibrated this stratigraphic model using HGS [*Therrien et al.*, 2005] coupled with PEST [*Doherty*, 2005] by matching drawdown responses observed at the CMT wells during the pumping test at PW1-3. The drawdown response recorded at PW2, PW3, PW4, and PW5 were omitted from the model calibration so that the analysis would be consistent with the previous cases which were populated with data only from PW1 and CMTs 1-4. In order for the model calibration to complete in a reasonable period of time, the stratigraphic model was transferred to the coarser model domain described in section G3.1.

A total of 20 layers were assigned in the stratigraphic model, and values of K and S_s were estimated for each. The PEST simulation started with initial estimates of K (from permeameter data) and S_s (estimated in a similar fashion to that presented in section 3.2 for the transition probability/Markov chain case). In total approximately 1200 model runs and 1 week was required for the parameter estimation. The simulations were performed on a Lenovo Thinkpad T61p with a 2.2 GHz processor and 2GB of RAM. Figures G9a and G9b show the calibrated K and S_s fields for the calibrated stratigraphic model respectively. Figure G10 shows the calibration results for this case. While the calibrated fields were able to reproduce the observed drawdown average sense, there is considerable scatter, possibly suggesting that the layers are not accurately represented. We also note that as the approach is not based on a geostatistical approach, a map of variance is unavailable.

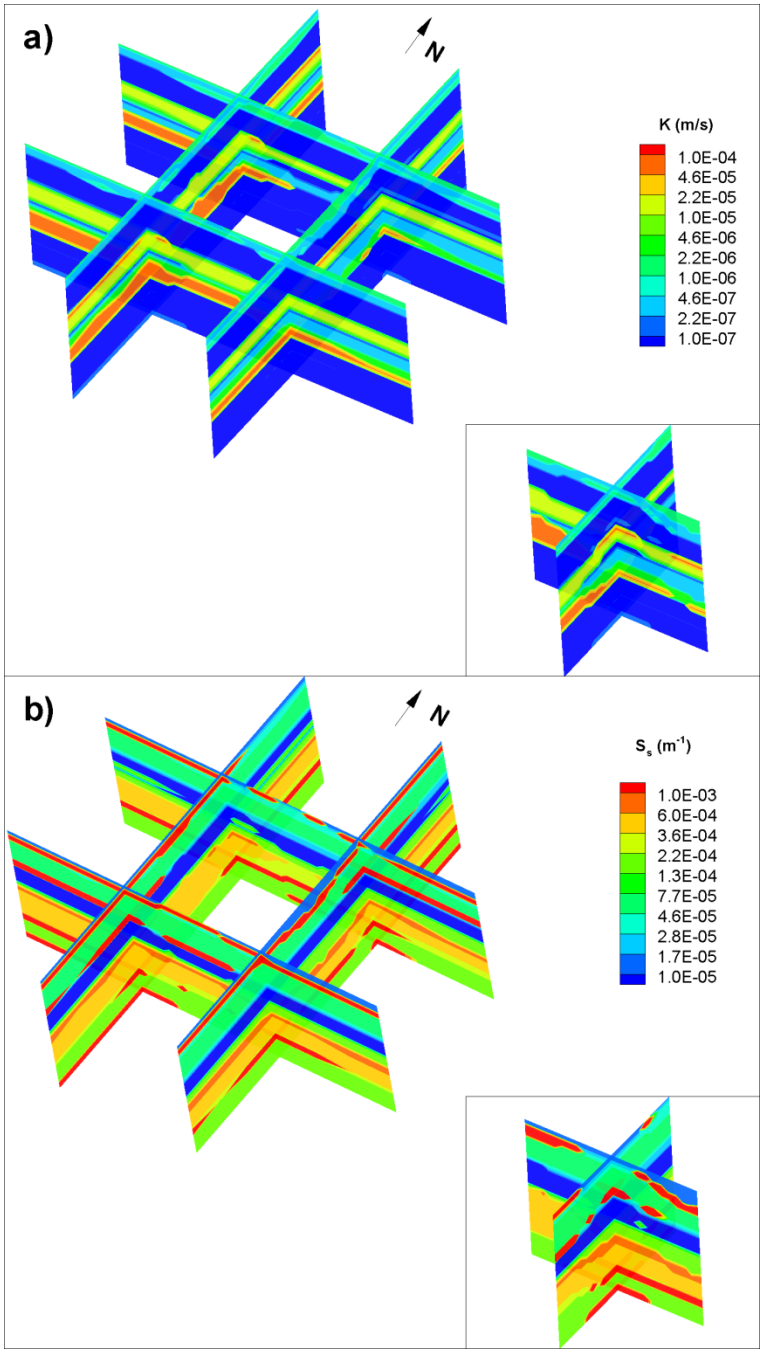


Figure G9: Three-dimensional perspective view of a calibrated K (a) and S_s (b) fields for the stratigraphic model. Note that the square formed by the slices in the main figure corresponds to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

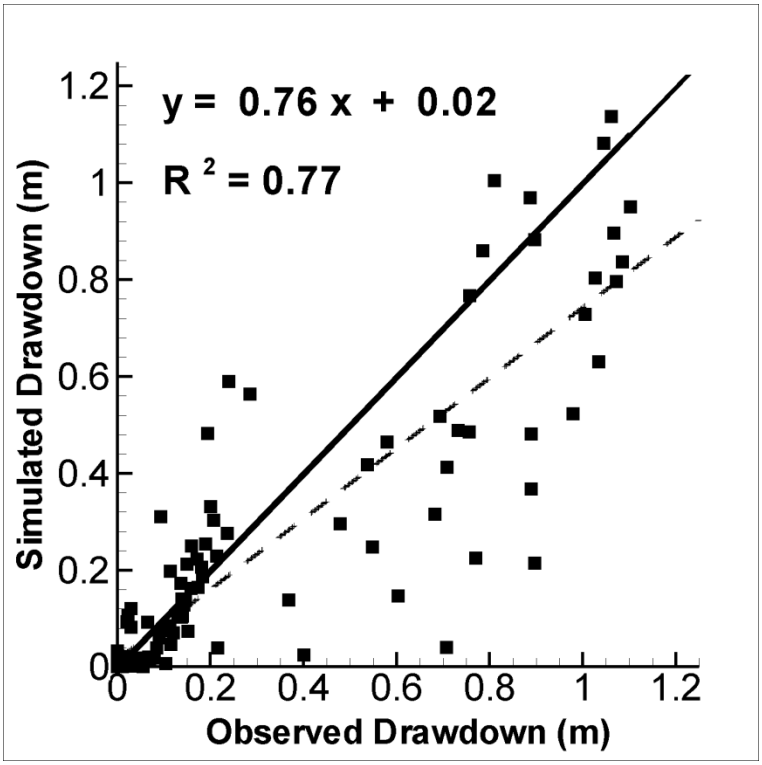


Figure G10: Calibration scatterplot for calibrated stratigraphic model.

G3.5. Transient Hydraulic Tomography

In a recent study, *Berg and Illman* [*in review (b)*] demonstrated the ability of THT to map aquifer heterogeneity in 3D at the field scale at the NCRS. Using 4 of the pumping tests performed at the site (PW4-3, PW1-3, PW5-3, and PW3-3) they evaluated the ability of THT to map a heterogeneous aquifer, as well as the impact of *a priori* information on the inversion. In particular, *Berg and Illman* [*in review (b)*] examined three different cases: a) starting the inversion with effective K and S_s fields estimated from pumping test data (similar to the estimation procedure described in section G3.1); b) starting the inversion with 3-layer K and S_s fields with each layer being homogeneous based on parameter estimates from type-curve analysis using the *Hantush* [1960] solution; and c) starting the inversion with homogeneous K and S_s fields and populating the inverse

model with local-scale K obtained via permeameter analysis of core samples along 5 boreholes. Based on this study, they concluded that Case b) was the best performing case. Similarly for this study, all three different levels of *a priori* information were considered and similar conclusions were reached. As such, only case b will be presented in this study, and will be referred to as Case 5 for the remainder of this manuscript.

The THT analysis performed as part of this study is identical to Case 4b of *Berg and Illman* [*in review NCRS (b)*], expect for this case, only the observed drawdowns recorded at PW1, and CMTs 1-4 were included in the inversion. In *Berg and Illman* [*in review (b)*], 4 more boreholes and 16 observation points were available for their analysis. In this study, we have 5 boreholes and 33 observation points so that we can make a fair comparison with the other heterogeneity modeling approaches that we describe in this paper. A brief description of the model setup and the inversion is provided here and for more details see *Berg and Illman* [*in review (b)*].

Pumping tests performed at locations PW4-3, PW1-3, PW5-3, and PW-3 were selected for THT analysis and were included in the SSLE model in the order just presented. The model domain described for the effective parameter estimation (section G3.1) was used for this case. All THT cases were executed using between 8 and 40 processors on a PC-cluster consisting (of 1 master and 12 slaves each with Intel Q6600 Quad Core CPUs running at 2.4 GHz with 16 GB of RAM per slave) at the University of Waterloo. The operating system managing the cluster was CentOS 5.3 based on a 64-bit system. The total computational time was approximately 1 week.

For this case, the aquifer system was conceptualized as a main aquifer system bounded above and below by aquitards (each 5 m thick). Initial values of K and S_s for these layers were estimated by performing a *Hantush* [1960] type-curve analysis on the observation points located within the central aquifer unit during pumping tests performed at PW1-3, PW1-4, and PW 1-5. Each drawdown response was individually matched using *Aqtesolv Pro 4.5* [*Duffield*, 2007] to estimate K and S_s for

the 'aquifer zone' as well as leakage parameters for the adjacent aquitards which were used to estimate K and S_s for the aquitards.

Values assigned to the aquifer as starting conditions for THT were 8.5×10^{-6} m/s and 2.2×10^{-5} m^{-1} for K and S_s respectively; and 4.5×10^{-7} m/s and 1.0×10^{-4} m^{-1} for the aquitard K and S_s respectively. Figure G11 shows the estimated K (Figure G11a) and S_s (Figure G11c) tomograms and the variance (Figures G11b and G11d respectively) associated with their estimates. This THT inversion captures the presence of the double aquifer zone while honouring the low K material near the top and bottom of the model domain. The variance in estimated parameters (Figures G11b, and G11d) indicate that parameter estimates are best in the upper and central portion of the model domain. This is consistent with the findings of *Berg and Illman* [*in review (b)*]. This zone of lower variance (less error) is associated with the main aquifer zone where a strong hydraulic response was observed during pumping tests. Little to no drawdown was observed for the bottom portion of the model domain where there is greater uncertainty in the estimated parameters.

3.5. Estimation of relevant geostatistical parameters

A geostatistical analysis was then performed to estimate various statistical properties of the different heterogeneous modeling approaches (cases 2-5). Tables 3a and 3b summarize the statistical parameters of the K and S_s fields, including geometric mean, variance, and the variogram parameters. Geometric mean and variance were calculated from the raw data sets without considering spatial location. The variogram parameters were estimated using SGeMS [*Remy et al.,* 2008]. SGeMS was used to generate directional (x, y, and z) experimental variograms for ln K and ln S_s to which either exponential or Gaussian variograms were fit. Many of the vertical variograms showed the 'hole effect' which is a result of the layered nature of the parameter fields.

Examination of Tables G3a and G3b reveals that the variance of ln *K* ranges from 4.0 to 7.4, and the variance of ln S_s ranges from 0.1 to 3.4 for the various approaches. For ln *K*, the horizontal correlation lengths were approximately 2 to 3 times greater than the vertical correlation length for all cases examined with horizontal correlation lengths ranging from 1.4 for case 2 to 18 for case 1. This variability in correlation lengths is more of a reflection of the various modeling approach than the aquifer system itself. For ln S_s, the horizontal correlation lengths were also 2 to 3 times greater than the vertical.

Table G3a. Geometric mean, variance, and correlation lengths of ln *K* for each case.

	ln *K* (m/s)						
	Variance*	Geometric Mean*	Sill	λ_x	λ_y	λ_z	Model
Case 2	5.5	4.0×10^{-8}	4.2^{xy}, 8^{z}	18.0	18.0	8.1	Exponential
Case 3	7.4	2.3×10^{-7}	7.3	1.4	1.4	0.5	Exponential
Case 4	6.5	2.1×10^{-7}	6.5	9.0	9.0	2.3	Exponential
Case 5	4.0	1.0×10^{-6}	4.6^{xy}, 5.5^{z}	5.0	5.0	1.8	Exponential

*Measured from raw data

^{xyz} direction sill was measured in

Table G3b. Geometric mean, variance, and correlation lengths of ln S_s for each case.

	ln S_s (m^{-1})						
	Variance*	Geometric Mean*	Sill	λ_x	λ_y	λ_z	Model
Case 2	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Case 3	0.1	1.6×10^{-4}	0.1	3.2	3.2	0.9	Exponential
Case 4	3.4	9.6×10^{-5}	3.3^{xy}, 4^{z}	11.3	11.3	4.5	Exponential
Case 5	1.5	8.4×10^{-5}	1.7^{xy}, 3.3^{z}	10.8	10.8	5.9	Gaussian

*Measured from raw data

^{xyz} direction sill was measured in

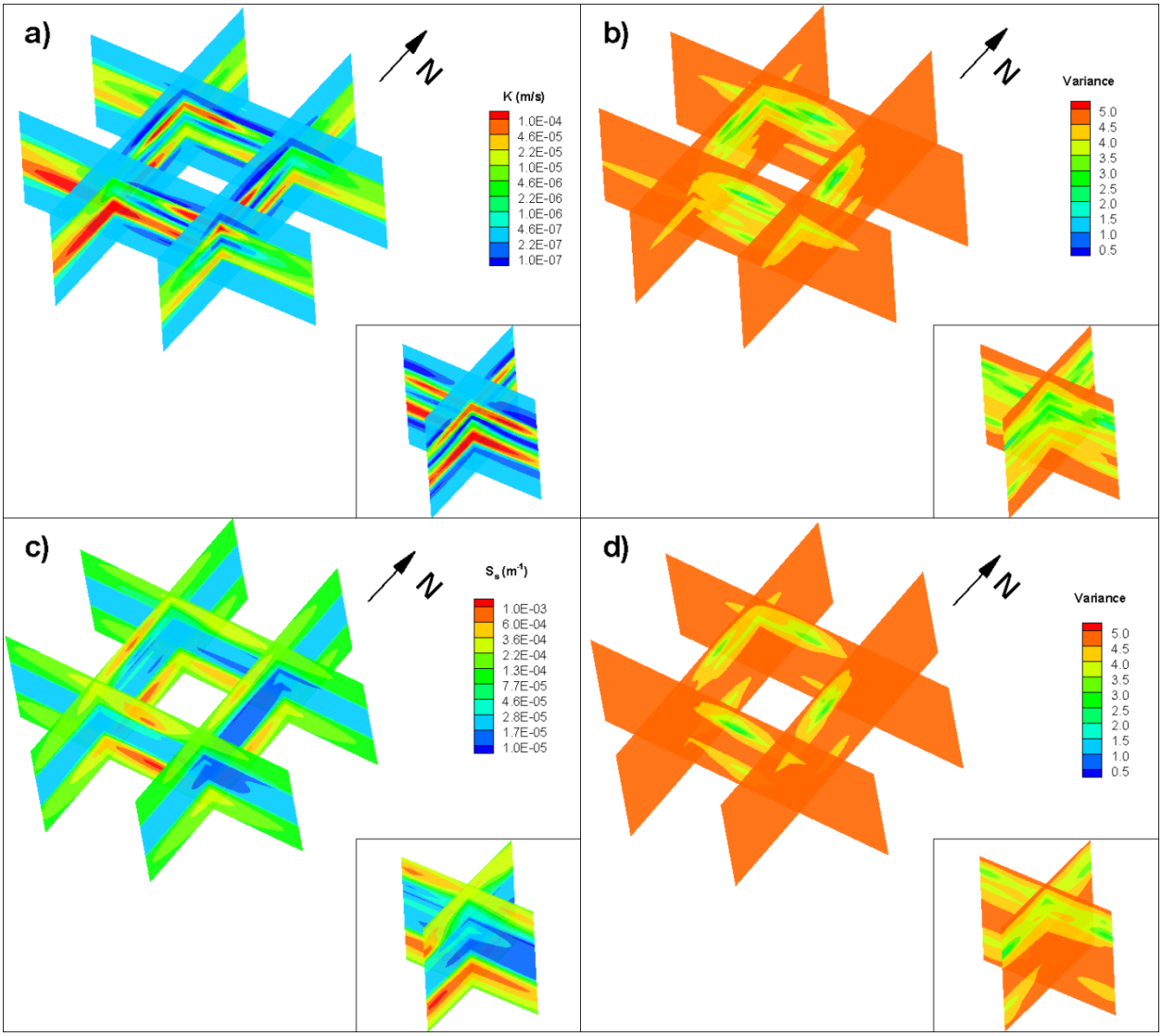


Figure G11: Three-dimensional perspective view of the THT *K*- and S_s-tomograms after the inclusion of 4 pumping tests (PW4-3, PW1-3, PW5-3, and PW3-3): a) *K*-tomogram; b) variance associated with the estimated *K* tomogram; c) S_s-tomogram; and d) variance associated with the estimated S_s-tomogram. Note that the square formed by the slices in the main figure corresponds to the outer edges of the field plot. The inset image for each figure is a cross-section through the middle of the central square. This corresponds to cross-sections through CMT2 to CMT1 (section oriented N-S), and CMT4 to CMT3 (section oriented E-W).

G4. PERFORMANCE COMPARISON OF SUBSURFACE HETEROGENEITY MODELING APPROACHES

A number of approaches have been proposed in the literature to validate forward and inverse groundwater flow models. For example, *Foglia et al*. [2007] tested alternative forward groundwater models using the corrected Akaike information criterion, Bayesian information criterion, and generalized cross-validation. *Poeter and Anderson* [2005] used a Kullback-Leibler information criterion that selects parsimonious groundwater models that provided a more realistic measure of precision than evaluation of any one model or evaluation based on other commonly referenced model selection criteria. Most recently, *Ye et al*. [2008] provided a comprehensive test of how various information criteria weigh and rank alternative models in light of the models' predictive performance under cross validation with real hydrologic data. In terms of the validation of inverse modeling results, *Painter et al*. [2007] compared three separate transmissivity estimation approaches and compared the resulting heads from forward simulation and actual head data from the Edwards aquifer in Texas. *Illman et al*. [2007] conducted laboratory sandbox study to validate the steady state hydraulic tomography approach, and in a separate study, *Liu et al*. [2007] studied the robustness of transient hydraulic tomography. Both studies used various data (core, slug, and single-hole test results) and statistical moments of these small scale data, as well as an independent cross-hole pumping test data to validate their K and S_s tomograms. In both studies, *Illman et al*. [2007] and *Liu et al*. [2007] found that the core, slug, single-hole tests could yield local estimates that could potentially be useful for validating their inverse modeling results. However, the comparison of local estimates to those estimates from hydraulic tomography yielded inconsistent results. These authors found that the inconsistencies are due to scale disparity, different flow conditions, as well as various (interpretation, sampling, experimental) errors and perhaps the small scale data may not be entirely suitable for large scale validation of groundwater flow models.

Illman et al. [2007] then tested their K tomogram by simulating a single independently conducted pumping test not used in the construction of the tomogram. They found excellent

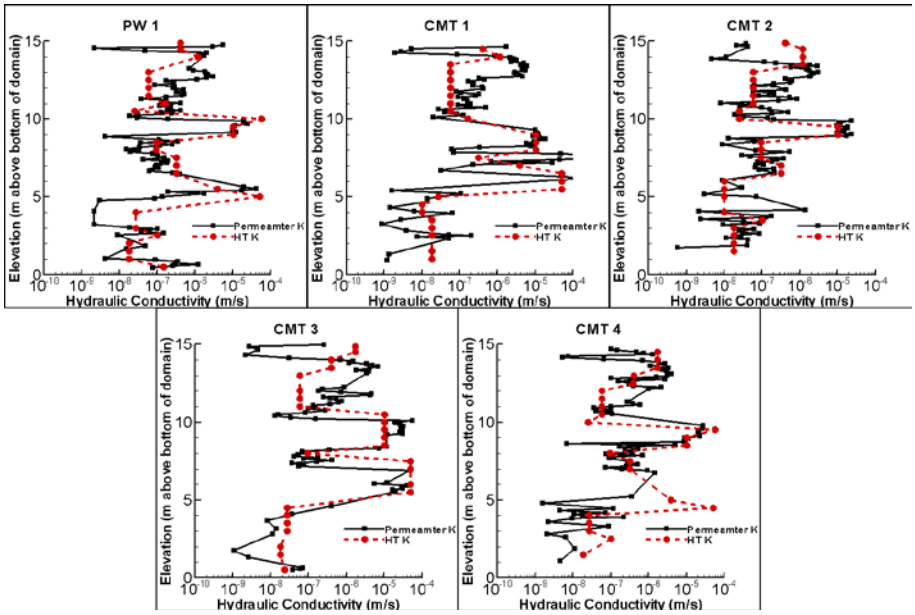
agreement of the simulated and observed heads giving further confidence that steady state hydraulic tomography can provide an accurate *K* distribution. *Liu et al*. [2007] then conducted a similar exercise and found that *K* and S_s tomograms provide the best transient prediction of an independently conducted pumping test. More recently, the results from steady-state hydraulic tomography [*Illman et al*., 2010] and transient hydraulic tomography [*Berg and Illman*, in review] were further tested using a different sandbox and heterogeneous aquifer as well as relying on 16 independent tests not used in the inverse modeling effort.

Here, the performance of the 5 subsurface heterogeneity modeling approaches presented are assessed by: 1) comparing *K* profiles along boreholes to permeameter values; and 2) simulating all 9 pumping tests using the various *K* and S_s fields from each case and comparing the simulated drawdown to the observed drawdown.

G4.1 Comparison to permeameter *K* data

One method for assessing the performance of heterogeneous modeling approaches is to compare the modeled fields to known permeameter data. To do this, we compare the *K* values estimated along 5 boreholes (PW1 and CMTs 1-4) to the permeameter data collected by *Alexander et al.* [2011]. This comparison is only made for cases 4 and 5 as cases 1 to 3 are either homogeneous, or are directly based on the permeameter data. Figures G12a and G12b show this comparison for cases 4 and 5, respectively. For this comparison, the estimated *K* values (red dashed line in Figure G12) are compared against measured permeameter values obtained by *Alexander et al.* [2011] along 5 boreholes (PW1, CMT1, CMT2, CMT3 and CMT4).

a)



b)

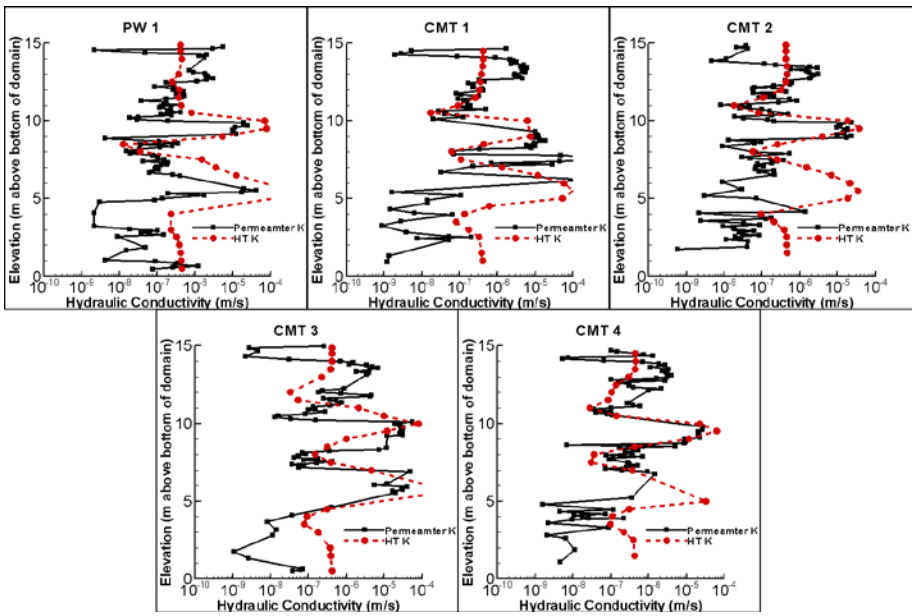


Figure G12: Vertical *K* profile along boreholes for the results from cases 4 (a) and 5 (b).

Examination of Figure G12a shows that the agreement between the *K*'s estimated by case 4 and the permeameter data are quite good. It should be reiterated at this point that the permeameter *K* data was used to provide the initial estimates of *K* for the parameter estimation procedure for the

stratigraphic model. This helps to constrain the estimation procedure in the lower portion of the model domain where very little hydraulic data is available, and thus, parameter estimates for these layers would be uncertain. Figure G12b shows the comparison for case 5 (THT). For this case, the *K*-tomogram captures the presence of the double aquifer zones located at approximately 5 and 10 m above the base of the model domain. However the *K* estimates made by case 5 for the aquifers is noticeably higher than the permeameter *K*.

G4.2 Prediction of Independent Pumping Tests

Simulating pumping tests not used in the inversion procedure is an approach for validating heterogeneous models that has been employed in other HT studies [*Illman et al.*, 2010; *Berg and Illman, in review (a);* and *Berg and Illman, in review (b)*]. For this comparison, all 9 pumping tests performed at the NCRS were simulated using the heterogeneous models created for each case. The forward simulations were conducted using HGS [*Therrien et al.*, 2005] and the model domains described for each case. We note that the inner 15 m x 15 m section of the model (representing the site plot) were nearly identical for all cases and that it is only the region outside of this that has a coarser discretization for cases 1, 4, and 5. As such, the mesh refinement has a negligible effect on the simulated results.

The *K* and S_s tomograms generated for case 5 were created using only the 1st hour of data recorded during each pumping test, whereas for this validation the entire test duration was simulated for each pumping test (6.5 - 22.5 hrs). A similar comparison was made by *Berg and Illman* [*in review (b)*] who found that for this particular site, the *K* and S_s tomograms generated using the first hour of data did a reasonable job predicting the full pumping test durations. We are currently investigating the use of more data from each drawdown curve and its effects on the results. Those results will be reported in the future.

For each case, all 9 pumping tests were simulated and the observed and simulated results were compared using scatterplots. (see Figures G13, G14, G15, G16, and G17 for cases 1, 2, 3, 4, and 5 respectively). These scatterplots compare observed and simulated drawdowns at 5 times throughout the duration of the pumping tests (these times are different for each test, and range from early time to late time). The solid line is the 1:1 line indicating a perfect match between observed and simulated drawdown, the dashed line is a linear model fit to the data (the equation for this line is presented on each graph) and, R^2 is the coefficient of determination.

Figure G13 shows the scatterplots for all 9 pumping tests generated using the effective K and S_s fields of case 1. There is clearly a significant amount of scatter in these plots, suggesting that these homogeneous fields, which do not honour the heterogeneous nature for the site, are poor representations of the true field, even in an average sense.

The scatterplots for cases 2 and 3 are quite similar to each other with the simulated drawdown generally being greater than the observed. This suggests that the geometric mean K of these model domains was too low. Since both of these model domains were populated with K data obtained from the permeameter analysis of soil cores, the most likely explanation is that the aquifer material was under represented in the sampling process. In fact, during the collection of soil cores, *Alexander et al.* [2011] reported that recovery averaged 80% for 4 of the wells (PW1, CMT1, CMT2, and CMT4), and 69% for CMT3. The lost material was often associated with the aquifer units [*Alexander et al.*, 2011], thus resulting in a biased estimate of K. Additionally, for case 3, the selection of an appropriate realization is a challenge as each realization is equally probable.

The drawdown scatterplots for the calibrated stratigraphic model (case 4) show a significant improvement in comparison to previous cases. The best matches are for the pumping tests at PW1-3 (the test to which the stratigraphic model was calibrated), and PW1-5. For the remaining 7 pumping tests, matches are improved in comparison to the previous cases. This indicates that for a layered

system such as this, calibration of a stratigraphic model to a single pumping test may be able to reasonably predict the response of a system to pumping at other pumping locations especially when the stratigraphy (i.e., geology) is accurately captured.

Case 5 shows a slight improvement over case 4, with the best matches occurring for the pumping tests to which it was calibrated (PW4-3, PW1-3, PW3-3, and PW5-3). Poorer matches are seen for pumping tests not included in the inversion, in particular, PW1-4 and PW5-4 show considerable scatter.

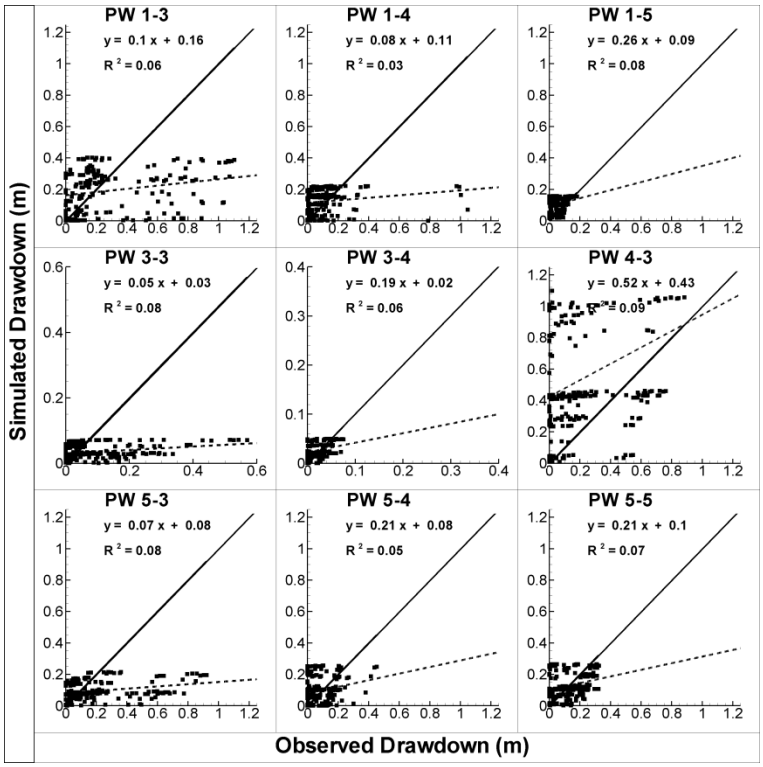


Figure G13: Scatterplots for all 9 pumping tests using the effective K and S_s values generated from Case 1. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

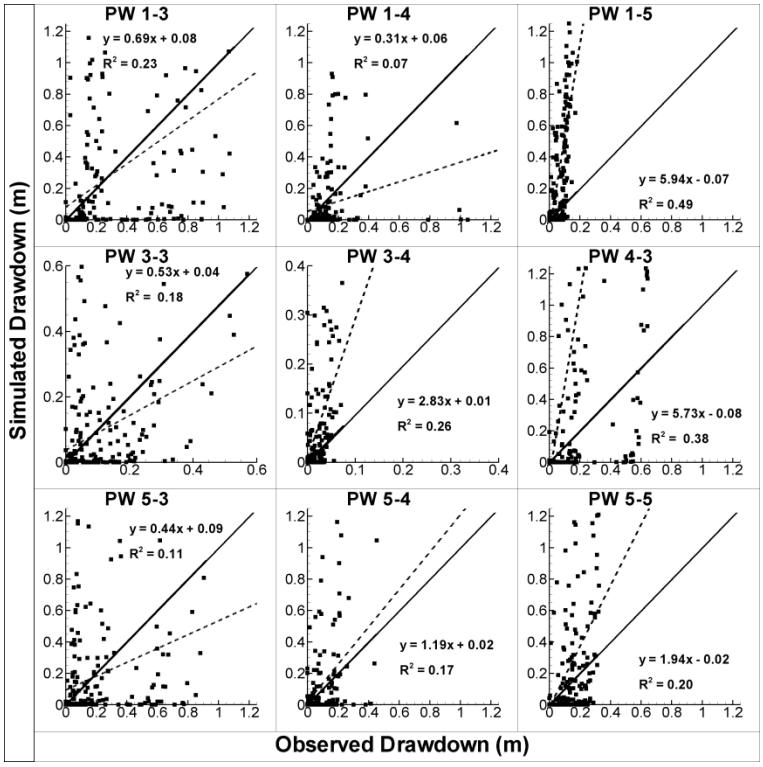


Figure G14: Scatterplots for all 9 pumping tests using the K and S_s distributions generated from Case 2. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

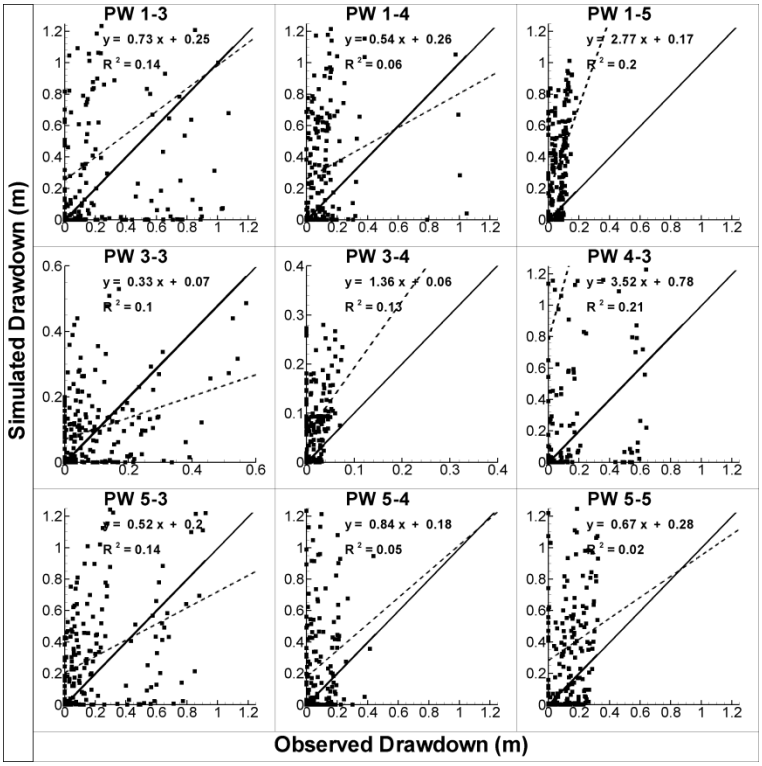


Figure G15: Scatterplot for all 9 pumping tests using the K and S_s distributions generated from Case 3. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

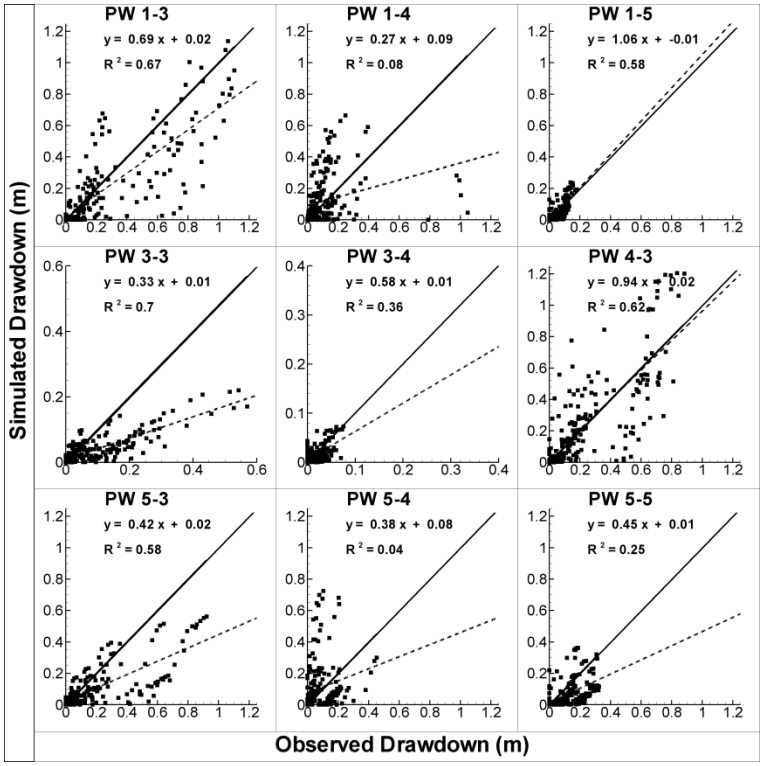


Figure G16: Scatterplot for all 9 pumping tests using the K and S_s distributions from Case 4. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

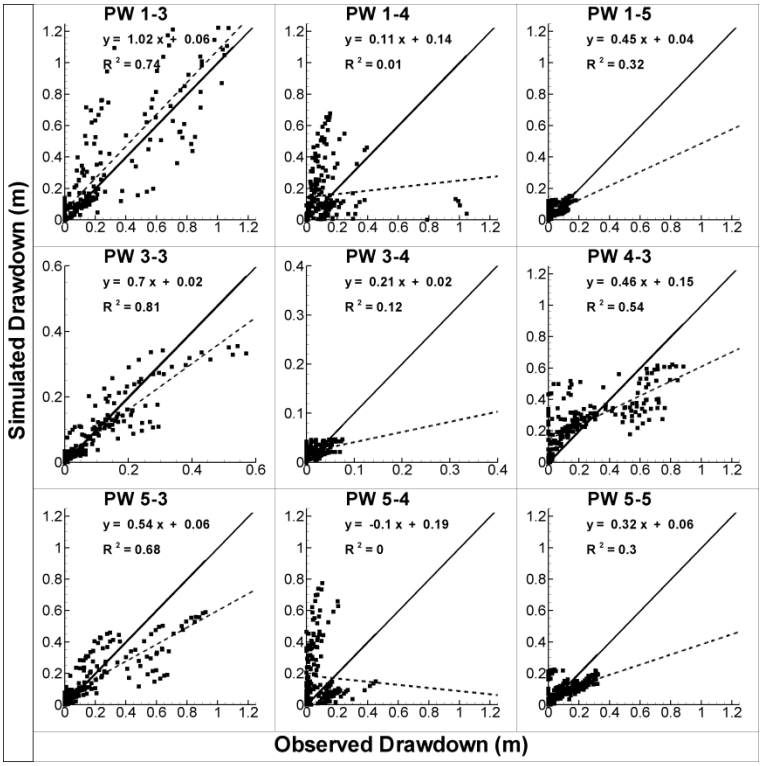


Figure G17: Scatterplot for all 9 pumping tests using the *K* and S_s tomograms from Case 5. The solid line is a 1:1 line indicating a perfect match. The dashed line is a best fit line, and the parameters describing this line are on each plot.

Finally, to further quantitatively assess the correspondence between the simulated and observed drawdown values (same data as that presented in the scatterplots, Figures G13-G17), we compute the mean absolute error (L_1) and the mean square error (L_2) norms of all cases examined. The L_1 and L_2 norms are computed as:

$$L_1 = \frac{1}{n}\sum_{i=1}^{n}\left|\chi_i - \hat{\chi}_i\right| \qquad [1]$$

$$L_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\chi_i - \hat{\chi}_i\right)^2 \qquad [2]$$

where n is the total number of drawdown data, i indicates the data number, and χ_i and $\hat{\chi}_i$ represent the estimates from the simulated and measured drawdowns, respectively. The L_1

and L_2 norms were calculated for each case by evaluating the observed and simulated drawdowns at 3-5 times at each port. Thus, the number of data points for each L_1 and L_2 norm represents 151 to 205 observations.

Table G4 summarizes the L_1 norm, while Table G5 summarizes the L_2 norm calculated for all the cases. The cells of each entry in the table are color-coded. The minimum value in the table is assigned a color of dark green, the maximum value a color of dark red, and the median value a color of yellow. Values intermediate to these anchor points are assigned appropriate intermediate colors. Tables G4 and G5 show that forward simulations using Cases 4 and 5 perform the best, which is in agreement if the visual assessment of the scatterplots.

Table G4. L_1 norms of observed versus simulated drawdowns from pumping tests PW1-3, PW1-4, PW1-5, PW3-3, PW3-4, PW4-3, PW5-3, PW5-4, PW5-5.

	PW1-3	PW1-4	PW1-5	PW3-3	PW3-4	PW4-3	PW5-3	PW5-4	PW5-5
Case 1	0.23	0.10	0.05	0.08	0.02	0.40	0.16	0.08	0.09
Case 2	0.25	0.12	0.27	0.09	0.06	1.18	0.21	0.12	0.20
Case 3	0.39	0.28	0.30	0.10	0.07	1.48	0.25	0.23	0.31
Case 4	0.11	0.11	0.03	0.06	0.01	0.12	0.11	0.11	0.08
Case 5	0.13	0.14	0.03	0.03	0.02	0.13	0.09	0.17	0.07
Max	1.48								
Min	0.01								

Table G5. L_2 norms of observed versus simulated drawdowns from pumping tests PW1-3, PW1-4, PW1-5, PW3-3, PW3-4, PW4-3, PW5-3, PW5-4, PW5-5.

	PW1-3	PW1-4	PW1-5	PW3-3	PW3-4	PW4-3	PW5-3	PW5-4	PW5-5
Case 1	0.10	0.03	0.005	0.02	0.0005	0.30	0.06	0.01	0.01
Case 2	0.17	0.05	0.19	0.02	0.01	6.04	0.11	0.06	0.16

Case 3	0.35	0.17	0.16	0.02	0.01	5.37	0.12	0.14	0.24
Case 4	0.04	0.04	0.002	0.01	0.0003	0.04	0.03	0.03	0.01
Case 5	0.04	0.05	0.002	0.003	0.0004	0.03	0.02	0.06	0.01
Max	6.04								
Min	0.0003								

G5. SUMMARY AND CONCLUSIONS

The main purpose of this paper was to compare commonly used approaches for modeling subsurface heterogeneity in K and S_s fields. In particular we created five transient groundwater flow models consisting of different K and S_s fields that incorporate heterogeneity in various degrees of detail to represent the heterogeneous aquifer-aquitard system at the North Campus Research Site situated on the University of Waterloo. Methods compared include: 1) effective parameters; 2) ordinary kriging; 3) conditional simulation based on the Transition Probability/Markov Chain Geostatistical method; 4) geological modeling, and 5) transient hydraulic tomography. Site characterization data used to construct the various parameter fields included 471 estimates of K from the permeameter analyses of core samples and 4 pumping tests conducted in a tomographic manner. The performance of these cases were then assessed by comparing the estimated hydraulic parameter fields to known geology (permeameter data) and through the numerical simulation of 9 pumping tests performed at the site. Drawdown values from these simulated pumping tests were then compared to the observed drawdowns using scatterplots. For each of the scatterplots, a linear model was fit through the data and the coefficient of determination (R^2) was calculated. These parameters give an indication of the goodness-of-fit between the observed and simulated drawdowns (R^2) and an indication of bias (the slope and intercept of the best fit line). In addition, we calculated the L_1 and L_2 norms.

This study leads to the following major findings and conclusions:

1) While estimating effective parameters (case 1) for relatively homogeneous systems may be useful and simpler to implement, the effective parameter approach when applied to the highly heterogeneous NCRS was unable to predict the response of the system to pumping. This point is particularly important for contaminated sites where an accurate understanding of heterogeneity is crucial to predict the migration of contaminants or design remediation systems;

2) The interpolation of point data (case 2) through the use of ordinary kriging is a useful way of identifying key aquifer features and obtaining a 3D understanding of the aquifer system. However, when using this modeling approach for groundwater simulations, the data must be complete and accurate, else, significant bias and error can result. For this particular study, the lack of samples associated with the key aquifer zones resulted in biased *K* estimates, and thus, caused the simulated drawdowns to be significantly greater than the observed. Additionally, estimates of specific storage are routinely not obtained from core samples thus making the prediction of transient head changes using a groundwater model difficult unless these estimates are available;

3) When a large amount of data are a available, the use of geostatistical interpolators, such as ordinary kriging, is an attractive option. For this site, despite the availability of 471 *K* samples, the approach was unsuccessful. This is likely a result of only having data along 5 boreholes, thus, the vertical characterization was relatively accurate, however, the horizontal extent of the heterogeneity was not well defined. This is a problem common to many sites, where a large number of boreholes are rarely available, resulting in a poor understanding of the lateral extent of aquifer materials;

4) While ordinary kriging honours the observed data, the transition between data points is smooth and may not accurately reflect the geology. Categorical interpolators such as the transition probability/Markov Chain approach can produce abrupt material changes. However, this approach produces many realizations and selecting the best case can pose a challenge. Additionally, these

categories must be assigned parameter values, which, if not accurate (e.g., obtained from incomplete core logs) may result in biased simulation results;

5) If parameter values estimated from soil cores are questionable, a stratigraphic model can be created and the layers calibrated to observed drawdowns. For case 4, this was done by creating a stratigraphic model based on the geological information at the site, and calibrating each layer for K and S_s to the observed drawdown during pumping at PW1-3. The resulting parameter fields were able to better predict pumping tests at the site than the previous cases. However, assigning layers in a model can be restrictive to the estimation process as the layer is assumed to be homogeneous, which may not be an accurate reflection of reality; and

6) For this study, the data requirements of THT were significantly less than cases 2-4 which required detailed core data (cases 2 and 3), as well as hydraulic data for calibration purposes (case 4). The data requirements for THT were a basic understanding of the geology and the hydraulic data from 4 pumping tests. In terms of data collection, THT requires significantly less effort. Additionally, since THT is calibrated to more pumping tests than the calibrated stratigraphic model it is better able to represent the heterogeneity of the system. However, for this particular system, THT was unable to accurately predict all of the independent pumping tests unlike previous findings by *Illman et al.* [*2010*] and *Berg and Illman* [*in review* (a)] which suggests that, although hydraulic tomography is a powerful approach for imaging subsurface heterogeneity, additional work is necessary.

Bibliography

Alexander, M., S. J. Berg, and W. A. Illman (2011), Field Study of Hydrogeologic Characterization Methods in a Heterogeneous Aquifer, *Ground Water*, no. doi: 10.1111/j.1745-6584.2010.00729.x.

Alexander, M. (2009), Evaluation of traditional hydrogeologic characterization approaches in a highly heterogeneous glaciofluvial aquifer/aquitard system. M.Sc. thesis, Dep. of Earth and Environmental Sci., Univ. of Waterloo, Waterloo, Ontario, Canada.

American Society for Testing and Materials (ASTM) (1985), *Annual Book of ASTM Standards, Classification of Soil for Engineering Purposes*. Philadelphia, Pennsylvania: ASTM.

American Society for Testing and Materials (ASTM) (2003), *Annual Book of ASTM Standards, ASTM D 5084 Standard Test Methods for Measurement of Hydraulic Conductivity of Saturated Porous Materials Using a Flexible Wall Permeameter*. Philadelphia, Pennsylvania: ASTM.

Annable, M.D., P.S.C. Rao, K. Hatfield, W.D. Graham, A.L. Wood, and C.G. Enfield (1998), Partitioning tracers for measuring residual NAPL: Field-scale test results, *J. Environ. Eng.*, *124*(6), 498-503.

Benson D. A., S. W. Wheatcraft, M. M. Meerschaert (2000), The fractional-order governing equation of Levy motion. Water Resour. Res., 36(6): 1413–23.

Berg, S.J. and W.A. Illman (2009), Transient hydraulic tomography at the North Campus Research Site, University of Waterloo. *Eos. Trans. AGU*, 90 (52), Fall Meet. Suppl., Abstract H21C-0865.

Berg, S. J. and W. A. Illman (2010), Three-dimensional transient hydraulic tomography and comparison to other heterogeneity imaging methods, Abstract H123D-0996 poster presented at Fall Meeting, AGU, San Francisco, Calif., 13-17 Dec.

Berg, S.J., P.A. Hsieh, and W.A. Illman, (2011), Estimating hydraulic parameters when poroelastic effects are significant, *Ground Water,* no. doi: 10.1111/j.1374-6584.2010.00781.x.

Berg, S. J. and W. A. Illman (*in review (a)*), Capturing heterogeneity in groundwater flow parameters: Comparison of approaches through controlled sandbox experiments.

Berkowitz, B., A. Cortis, M. Dentz, H. Scher (2006), Modeling non-Fickian transport in geological formations as a continuous time random walk. Rev. Geophys., 44:RG2003.

Bevan, M. J., A. L. Endres, D. L. Rudolph, and G. Parkin (2005), A field scale study of pumping-induced drainage and recovery in an unconfined aquifer, *J. Hydrol*., 315, 52– 70, doi:10.1016/j.jhydrol.2005.04.006.

Biot, M.A. (1941), General theory of three-dimensional consolidation, *Journal of Applied Physics* 12: 155-164.

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

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

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

Bohling, G. C. and J. J. Butler Jr. (2010), Inherent Limitations of Hydraulic Tomography, Groundwater, 48(6), 809–824, doi: 10.1111/j.1745-6584.2010.00757.x

Boman, G.K., Molz, F.J., and K.D. Boone (1997), Borehole flowmeter application in fluvial sediments: Methodology, results, and assessment, *Ground Water*, 35(3), 443-450.

Boulton, N. S. (1954), Unsteady radial flow to a pumped well allowing for delayed yield from storage, Publ. 37, pp. 472–477, Int. Assoc. of Sci. Hydrol., Rome.

Boulton, N. S. (1963), Analysis of data from non-equilibrium pumping tests allowing for delayed yield from storage, *Proc. Inst. Civ. Eng*., 26, 469–482.

Bouwer, H., and R. C. Rice (1976), Slug Test for Determining Hydraulic Conductivity of Unconfined Aquifers with Completely or Partially Penetrating Wells, *Water Resour. Res.*, *12*(3), 423-428.

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.

Brauchler, R., R. Hu, P. Dietrich, and M. Sauter (2011), A field assessment of high resolution aquifer characterization based on hydraulic travel time and hydraulic attenuation tomography, *Water Resour. Res.*, *in press*.

Bunn, M.I., J. P. Jones, A.L. Endres and D. L. Rudolph (2010), Effects of hydraulic conductivity heterogeneity on vadose zone response to pumping in an unconfined aquifer, *J. Hydrol.*, 387, 90-104.

Burbey, T.J. (2003), Use of time-subsidence data during pumping to characterize specific storage and hydraulic conductivity of semi-confining units. *Journal of Hydrology* 281: 3-22.

Burbey T.J., S.M. Warner, G. Blewitt, J.W. Bell, and E. Hill (2006), Three-dimensional deformation and strain induced by municipal pumping, part 1: analysis of field data. *Journal of Hydrology* 319: 123-142.

Burbey, T.J. (2006), Three-dimensional deformation and strain induced by municipal pumping, part 2: numerical analysis. *Journal of Hydrology* 330: 422-434.

Butler, J. J., Jr., P. Dietrich, V. Wittig, and T. Christy (2007), Characterizing hydraulic conductivity with the direct-push permeameter, *Ground Water*, 45(4), 409– 419.

Cacas, M. C., E. Ledoux, G. de Marsily, A. Barbreau, P. Calmels, B. Gaillard, and R. Margritta, (1990a), Modeling fracture flow with a stochastic discrete fracture network: Calibration and validation, 1, The flow model, *Water Resour. Res.*, *26*(3), 479-489.

Cacas, M. C., E. Ledoux, G. de Marsily, A. Barbreau, P. Calmels, B. Gaillard, and R. Margritta, (1990b), Modeling fracture flow with a stochastic discrete fracture network: Calibration and validation, 2, The transport model, *Water Resour. Res.*, *26*(3), 491-500.

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

Caers J (2001) Geostatistical reservoir modelling using statistical pattern recognition. *J Petroleum Sci Eng* 29:177–188

Carle, S. F., and G. E. Fogg (1997), Modeling spatial variability with one and multi-dimensional continuous Markov Chains, *Math Geol.*, 29(7), 891-918.

Carle, S. F. (1999), *T-PROGS: Transition Probability Geostatistical Software*, Version 2.1.

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

Cooper, H.H, Jr., and C.E. Jacob (1946), A generalized graphical method for evaluating formation constants and summarizing well-field history. *Transactions American Geophysical Union* 27: 526-534.

Cooper, H., Jr., J. Bredehoeft, and I. Papadopulos (1967), Response of a Finite-Diameter Well to an Instantaneous Charge, *Water Resour. Research*, *3*(1), 263-269.

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

Cushman, J.H. and T. R. Ginn. (1993), Nonlocal dispersion in media with continuously evolving scales of heterogeneity. Transp Porous Media,13(1): 123–38.

Dagan, G. (1982), Stochastic modeling of groundwater flow by unconditional and conditional probabilities: 2. The solute transport, Water Resour. Res., 18(4), 835– 848, doi:10.1029/WR018i004p00835.

Dagan, G. (1984), Solute transport in heterogeneous formations, J. Fluid Mech., 145, 151– 177, doi:10.1017/S0022112084002858.

Dagan, G. (1967), A method of determining the permeability and effective porosity of unconfined anisotropic aquifers, *Water Resour. Res*., 3, 1059– 1071, doi:10.1029/WR003i004p01059.

Datta-Gupta, A., S. Yoon, D.W. Vasco, and Pope, G.A (2002), Inverse modeling of partitioning interwell tracer tests: A streamline approach, *Water Resour. Res.*, *38*(6), 1079, doi:10.1029/2001WR000597.

De Marsily, G., F. Delay, J. Gonçalvès, Ph. Renard, V. Teles, and S. Violette (2005), Dealing with spatial heterogeneity, *Hydrogeol. J*., 13:161-183.

de Vries, L. M., J. Carrera, O. Falivene, O. Gratacos, and L. J. Slooten (2009), Applicaiton of multiple point geostatistics to non-stationary images, *Mathematical Geosci*., 41(1), 29-42.

Dershowitz, W., G. Lee, J. Geier, S. Hitchcock, and P. LaPointe, (1997), Fracman Interactive Discrete Feature Data Analysis, Geometric Modeling, and Exploration Simulation, User Documentation, Version 2.6, Golder Associates Inc. Seattle, WA.

Deutsch, C. V., and A. G. Journel (1998), *GSLIB Geostatistical Software Library and User's Guide*, 2nd ed., 369 pp.

Dietrich, P., and C. Leven (2005), Direct Push Technologies, In Groundwater Geophysics, edited by R. Kirsch, Springer, 321– 340.

Dietrich, P., J. J. Butler Jr., and K. Faiß (2008), A rapid method for hydraulic profiling in unconsolidated formations, Ground Water, 46(2), 323– 328, doi:10.1111/j.1745-6584.2007.00377.x.

Doherty, J. (2005), *PEST Model-Independent Parameter Estimation User Manual: 5th Edition.* Watermark Numerical Computing, Brisbane, Australia.

Domenico, P.A., and F.W. Schwartz (1997), *Physical and Chemical Hydrogeology*. New York: John Wiley and Sons, 506.

Duffield, G. M. (2007), *Aqtesolv Pro 4.5. Aquifer Test Analysis Software*. HydroSOLVE Inc.

Dugan, P.J., J.E. McCray, and G.D. Thyne (2003), Influence of a solubility-enhancing agent (cyclodextrin) on NAPL-water partition coefficients, with implications for partitioning tracer tests, *Water Resour. Res.*, *39*(5), 1123, doi:10.1029/2002WR001672.

Dverstorp, B., J. Andersson, and W. Nordqvist, (1992), Discrete fracture network interpretation of field tracer migration in sparsely fractured rock, *Water Resour. Res*., *28*(9), 2327-2343.

Ferrante, A.F.M (2007), Temporal moment analysis of partitioning tracer tests in synthetic heterogeneous geologic media. M.Sc. thesis, Dep. of Civil and Environmental Engineering, University of Iowa, Iowa City, Iowa, U.S.A.

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

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

Foglia, L., S.W. Mehl, M.C. Hill, P. Perona, P. Burlando (2007), Testing Alternative Ground Water Models Using Cross-Validation and Other Methods, *Ground Water*, 45(5), 627-641.

Gelhar, L. W., (1993), *Stochastic Subsurface Hydrology*. Prentice-Hall. Englewood Cliffs, New Jersey.

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

Gottlieb, J., and P. Dietrich (1995), Identification of the permeability distribution in soil by hydraulic tomography, *Inverse Probl*., 11(2), 353–360.

Guardiano F, Srivastava RM (1993) Multivariate geostatistics: beyond bivariate moments. In: Soares A (ed) *Geostatistics-Troia*. Kluwer, Dordrecht, pp 133–144

Gutjahr, A. L., and J. L. Wilson (1989), Cokriging for stochastic models, *Transp. Porous Media*, 4(6), 585-598.

Haldorsen H. H. and D. M. Chang (1986), Notes on stochastic shales from outcrop to simulation models. In: Lake LW, Carol HB Jr (eds) Reservoir characterization. Academic, New York, pp 152–167.

Hantush, M. S. (1960), Modification of the theory of leaky aquifers, *J Geophys. Res.65*(11), 3713-3725.

Harvey, C. and S. M. Gorelick (1995), Mapping hydraulic conductivity: Sequential conditioning with measurements of solute arrival time, hydraulic head, and local conductivity, *Water Resour. Res.*, 31(7), 1615-1626.

Hassanizadeh, S. M. and J. Carrera (1992), Editorial, *Adv. Water Resour*., 15(1): 1-3.

Heinz, J., S. Kleineidam, G. Teutsch, and T. Aigner (2003), Heterogeneity patterns of Quaternary glaciofluvial gravel bodies (SW-Germany): Application to hydrogeology, *Sediment. Geol*., 158, 1–23, doi:10.1016/S0037-0738(02)00239-7.

Hibbitt, K., and Sorenson-Inc. (1998), *ABAQUS, Standard Users Manual*. Pawtucket, Rhode Island.

Hsieh, P.A. 1996. Deformation-induced changes in hydraulic head during ground-water withdrawal, *Ground Water 34*: 1082-1089.

Hoeksema, R. J., and P. K. Kitanidis (1984), An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling, *Water Resour. Res.*, 20(7), 1003-1020.

Hoeksema, R. J., and P. K. Kitanidis (1989), Prediction of transmissivities, heads, and seepage velocities using mathematical modeling and geostatistics, *Adv. Water Resour.*, 12, 90-101.

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

Hughson, D., and T.-C. Yeh (2000), An Inverse Model for Three-Dimensional Flow in Variably Saturated Porous Media, *Water Resour. Res*., 36(4), 829-839.

Hvorslev, M. J. (1951), Time lag and soil permeability in ground-water observations*Rep.*, 1-50 pp, Waterways Exper. Sta. Corps of Engrs, U.S. Army.

Illman, W., and S. Neuman (2001), Type Curve Interpretation of a Cross-Hole Pneumatic Injection Test in Unsaturated Fractured Tuff, Water Resour. Res., 37(3), 583-603.

Illman, W. A., and S. P. Neuman (2003), Steady-state analysis of cross-hole pneumatic injection tests in unsaturated fractured tuff, *J Hydrol*, *281*(1-2), 36-54.

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

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

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

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

Illman, W. A., S. J. Berg, X. Liu, A. J. Craig and A. Massi (2010b), Hydraulic/partitioning tracer tomography for trichloroethene source zone characterization: Small-scale sandbox experiments, Environ. Sci. Technol., 44(22), pp. 8609-8614, doi: 10.1021/es101654j.

Jacob, C.E. 1940. The flow of water in an elastic artesian aquifer. *Transactions American Geophysical Union* 21: 574-586.

James, A.I., W.D. Graham, K. Hatfield, P.S.C. Rao, and M.D. Annable. 2000. Estimation of spatially variable residual nonaqueous phase liquid saturations in nonuniform flow fields using partitioning tracer data, *Water Resour. Res., 36*(4), 999–1012.

Jin, M, M. Delshad, V. Dwarakanath, D.C. McKinney, G.A. Pope, K. Sepehrnoori, C.E. Tillburn (1995), Partitioning tracer tests for detection, estimation, and remediation performance assessment of subsurface nonaqueous phase liquids. *Water Resour. Res., 31*(5), 1201-1211.

Journel, A. G. (1983), Nonparametric estimation of spatial distribution. *Math. Geol.* 15(3): 445-468.

Journel, A. G. and F. G. Alabert (1990). New method for reservoir mapping., *J. Pet. Technol.* February 42(2):212-218.

Journel, A. G. and J. Gomez-Hernandez (1993). Stochastic imaging of the Wilmington clastic sequence. *Soc. Pet. Eng. Form. Eval*. March 8(1): 33-40.

Journel, A. G. and E. K. Isaacs (1984). Conditional indicator simulation: application to a Saskatcheouan uranium deposit. *Math. Geol.* 16)7):685-718.

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

Karrow, P. F. (1993), Quaternary Geology, Stratford-Conestogo Area., Geological Survey Report, Ontario, Canada.

Karrow, P. F., Geology of the University of Waterloo Campus. In Department of Earth Sciences, University of Waterloo, Waterloo, Ontario, Canada: 1979.

Kim, J.M. (1995), COWADE123D: A finite element model for fully coupled saturated-unsaturated water flow in deforming one-, two-, and three-dimensional porous fractured media. *Technical Report HGL-1995-9*. Hydrogeology Laboratory, Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania.

Kim, J.M. and R.R. Parizek (1997), Numerical simulation of the Noordbergum effect resulting from groundwater pumping in a layered aquifer system. *Journal of Hydrology* 202: 231-243.

Kim, J.M. and R.R. Parizek (2005), Numerical simulation of the Rhade effect in layered aquifer systems due to groundwater pumping shutoff. *Advances in Water Resources* 28: 627-642.

Kitanidis, P. K., and E. G. Vomvoris (1983), A geostatistical approach to the inverse problem in groundwater modeling and one-dimensional simulations, *Water Resour. Res.,* 19(3), 677-690.

Klingbeil, R., S. Kleineidam, U. Asprion, T. Aigner, and G. Teutsch (1999), Relating lithofacies to hydrofacies: Outcrop-based hydrogeological characterization of Quaternary gravel deposits, *Sediment. Geol*., 129, 299– 310, doi:10.1016/S0037-0738(99)00067-6.

Klise, K. A., G. S. Weissmann, S. A. McKenna, E. M. Nichols, J. D. Frechette, T. F. Wawrzyniec, and V. C. Tidwell (2009), Exploring solute transport and streamline connectivity using lidar-based outcrop images and geostatistical representations of heterogeneity, *Water Resour. Res*., 45, W05413, doi:10.1029/2008WR007500.

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

Koltermann, C. E., and S. M. Gorelick (1992), Paleoclimatic signature in terrestrial flood deposits. *Science* 256:1775-1782.

Koltermann, C. E., and S. M. Gorelick (1996), Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches, *Water Resour. Res.*, *32*(9), 2617–2658.

Konikow L. F. and J. D. Bredehoeft (1992), Ground-water models cannot be validated, *Adv. Water Resour*., 15(1): 75-83.

Kozeny, J. (1927), Ueber kapillare leitung des Wassers im boden, *Akademie der Wissenschaften in Wien, 137*(2a).

Kram M.L., A.A. Keller, J. Rossabi, and L.G. Everett (2001), Characterization methods and approaches, part 1: Performance comparison. *Groundwater Monitoring Review*, 109-123.

Kroszynski, U. I., and G. Dagan (1975), Well pumping in unconfined aquifers: The influence of the unsaturated zone, *Water Resour. Res*., 11, 479–490, doi:10.1029/WR011i003p00479.

Langguth, H.R., and C. Treskatis. (1989), Reverse water level fluctuations in semiconfined aquifer systems – Rhade effect. *Journal of Hydrology* 109: 79-93.

LeBlanc, D. R., S. P. Garabedian, K. M. Hess, L. W. Gelhar, R. D. Quadri, K. G. Stollenwerk, and W. W. Wood (1991), Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 1. Experimental design and observed tracer movement, Water Resour. Res., 27(5), 895– 910, doi:10.1029/91WR00241.

Li, B. and T.-C. J. Yeh (1999), Cokriging estimation of the conductivity field under variably saturated flow conditions, Water Resour. Res., 35(12), 3663-3674.

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

Li, W., A. Englert, O. A. Cirpka, J.Vanderborght, and H.Vereecken (2007),Two dimensional characterization of hydraulic heterogeneity by multiple pumping tests, *Water Resour. Res*., 43, W04433, doi:10.1029/2006WR005333.

Liu, G., J. J. Butler Jr., G. C. Bohling, E. Reboulet, S. Knobbe, and D. W. Hyndman (2009), A new method for high-resolution characterization of hydraulic conductivity, *Water Resour. Res*., 45, W08202, doi:10.1029/2009WR008319.

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

Liu, X., and P. K. Kitanidis (2011), Large-scale inverse modeling with an application in hydraulic tomography, *Water Resour. Res.*, *47*(2), W02501.

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

Long, J.C.S., P. Gilmour, and P.A. Witherspoon (1985), A model for steady fluid flow in random 3-D network of disc shaped fractures, *Water Resour. Res*., *21*(8), 1105-1115.

Mackay, D. M., D. L. Freyberg, P. V. Roberts, and J. A. Cherry (1986), A Natural Gradient Experiment on Solute Transport in a Sand Aquifer 1. Approach and Overview of Plume Movement, Water Resour. Res., 22(13), 2017–2029.

Mao, D., L. Wan, T.-C. J. Yeh, C.-H. Lee, K.-C. Hsu, and J.-C. Wen (2011), A revisit of drawdown behaviour during pumping in unconfined aquifers, *Water Resour. Res*., (in review).

Maji, R. (2005). Conditional stochastic modelling of DNAPL migration and dissolution in a high-resolution aquifer analog. Ph.D. diss., Dept. of Earth Sciences, University of Waterloo.

Maji, R. and E.A. Sudicky (2008), Influence of mass transfer characteristics for DNAPL source depletion and contaminant flux in a highly characterized glaciofluvial aquifer, *J. Con. Hyd.*, 102(1-2), 105-119.

Martin, P.J. and E.O. Frind (1998), Modeling a complex multi-aquifer system: The Waterloo Moraine, *Ground Water 36*(4), 679-690.

Mathias, S. A., and A. P. Butler (2006), Linearized Richards' equation approach to pumping test analysis in compressible aquifers, *Water Resour. Res*., 42, W06408, doi:10.1029/2005WR004680.

McCarthy, J., B. Gu, L. Liang, J. Mas-Pla, T. Williams, and T.-C. Yeh. (1996), Field Tracer Tests on the Mobility of Natural Organic Matter in a Sandy Aquifer, Water Resour. Res., 32(5), 1223-1238.

McDermott C. I., M. Sauter, and R. Liedl (2003), New experimental techniques for pneumatic tomographical determination of the flow and transport parameters of highly fractured porous rock samples, *J. of Hydrol.*, *278* (1-4), 51-63.

Meerschaert, M. M. , D. A. Benson, and B. Baumer. (1999), Multidimensional advection and fractional dispersion. Phys. Rev. E., 59(5): 5026–8.

Meinzer, O.E. (1928), Compressibility and elasticity of artesian aquifers. *Economic Geology* 23: 263-291.

Mishra, P. K., and S. P. Neuman (2010), Improved forward and inverse analyses of saturated-unsaturated flow toward a well in a compressible unconfined aquifer, *Water Resour. Res*., 46, W07508, doi:10.1029/2009WR008899.

Moench, A. F. (1997), Flow to a well of finite diameter in a homogeneous, anisotropic water table aquifer, *Water Resour. Res*., 33, 1397–1407, doi:10.1029/97WR00651.

Moench, A.F., and S.P. Garabedian, and D.R. LeBlanc (2001), Estimation of hydraulic parameters from an unconfined aquifer test conducted in a glacial outwash deposit, Cape Cod, Massachusetts. *U.S. Geological Survey Professional Paper, 1629*, 69 pp.

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

Moreno-Barbero, E. and T.H. Illangasekare (2006), Influence of dense nonaqueous phase liquid pool morphology on the performance of partitioning tracer tests: Evaluation of the equilibrium assumption, *Water Resour. Res*., 42, W04408, doi:10.1029/2005WR004074.

Mualem, Y. (1976), A new model for predicting hydraulic conductivity of unsaturated porous media, *Water Resour. Res*., 12, 513–522, doi:10.1029/WR012i003p00513.

Nelson, N.T., and M.L. Brusseau (1996), Field study of the partitioning tracer method for detection of dense nonaqueous phase liquid in a trichloroethene-contaminated aquifer, *Environmental Science & Technology*, 30, 2859-2863.

Nelson, N.T., M. Oostrom, T.W. Wietsma, and M.L. Brusseau (1999), Partitioning tracer method for the in situ measurement of DNAPL saturation: Influence of heterogeneity and sampling method. *Environmental Science & Technology, 33*, (22), 4046-4053.

Neuman, S. P. (1972), Theory of flow in unconfined aquifers considering delayed response of the water table, *Water Resour. Res*., 8, 1031–1045, doi:10.1029/WR008i004p01031.

Neuman, S. P. (1974), Effects of partial penetration on flow in unconfined aquifers considering delayed aquifer response, *Water Resour. Res*., 10, 303–312, doi:10.1029/WR010i002p00303.

Neuman, S. P. (1987), Stochastic continuum representation of fractured rock permeability as an alternative to the REV and fracture network concepts, pp. 533-561, Rock Mechanics: Proceedings of the 28th U.S. Symposium, Tucson, Arizona, edited by I.W. Farmer, J.J.K. Daemen, C.S. Desai, C.E. Glass, and S.P. Neuman, A.A. Balkema, Rotterdam, 1240 p.

Neuman, S. P. (1993), Eulerian–Lagrangian theory of transport in space–time nonstationary velocity fields: exact nonlocal formalism by conditional moments and weak approximation. Water Resour Res., 29(3): 633–45.

Neuman, S. P. (2005), Trends, prospects and challenges in quantifying flow and transport through fractured rocks, *Hydrogeol. J*., 13: 124-147.

Neuman, S.P., and P. Witherspoon (1969), Theory of flow in a confined two aquifer system. *Water Resources Research* 5, no. 4: 803-816.

Neuman, S.P., and P Witherspoon. (1972), Field determination of the hydraulic properties of a leaky multiple aquifer system. *Water Resources Research* 8, no. 5: 1284-1298.

Neuman, S. P., C. L. Winter, and C. M. Newman (1987), Stochastic theory of field-scale Fickian dispersion in anisotropic porous media, *Water Resour. Res.*, 23(3), 453– 466, doi:10.1029/WR023i003p00453.

Neuman, S. P. and D. M. Tartakvosky (2009), Perspective on theories of non-Fickian transport in heterogeneous media, *Adv. Water Resour*. 32, 670-680.

Ni, C.-F. and T.-C. J. Yeh (2008), Stochastic inversion of pneumatic cross-hole tests and barometric pressure fluctuations in heterogeneous unsaturated formations, *Adv Water Resour*., 31(12), 1708-1718.

Ni, Chuen-Fa, T.-C. J. Yeh, and J.-S. Chen (2009), Cost-effective hydraulic tomography surveys for predicting flow and transport in heterogeneous aquifers, Environ. Sci. Technol., 43(10), 3720-3727.

Nwankwor, G.I., R.W. Gillham, G. van der Kamp, and F.F. Akindunni (1992), Unsaturated and saturated flow in response to pumping of an unconfined aquifer: field evidence of delayed drainage. *Ground Water*, 30(5), 690–700.

Oldenborger, G. A., R. A. Schincariol, and L. Mansinha (2002), Spacelocal spectral texture segmentation applied to characterizing the heterogeneity of hydraulic conductivity, *Water Resour. Res*., 38(8), 1154, doi:10.1029/2001WR000496.

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

Painter, S. L. A. D. Woodbury, and Y. Jiang (2007), Transmissivity estimation for highly heterogeneous aquifers: comparison of three methods applied to the Edwards Aquifer, Texas, USA, *Hydrogeol. J.*, 15: 315-331.

Pankow, J.F., and J.A. Cherry (1996), *Dense chlorinated solvents and other DNAPLs in ground water: History, behaviour, and remediation*, Waterloo Press, 522p.

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

Refsgaard, J. C., and H. J. Henriksen (2004), Modelling guidelines-terminology and guiding principles, *Adv. Water Resour*., 27(1), 71-82.

Refsgaard, J. C. (1997), Parameterisation, calibration and validation of distributed hydrological models, *J. Hydrol*., 198 (1-4): 69-97.

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

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

Remy, N., A. Boucher, and J. Wu (2008),. *Applied Geostatisticswith SGeMS: A Users Guide.* New York: Cambridge University Press.

Renard, P. and G. De Marsily (1997), Calculating equivalent permeability: a review, *Adv. Water Resour*., 20(5-6) 253-278.

Rodrigues, J.D. (1983), The Noordbergum effect and characterization of aquitards at the Rio Maior mining project. *Ground Water* 21: 200-207.

Rubin, Y., and G. Dagan (1987), Stochastic identification of transmissivity and effective recharge in steady groundwater flow, 1, Theory, *Water Resour. Res.*, 23(7), 1185-1192.

Sanchez-Vila, X., A. Guadagnini, and J. Carrerra (2006), Representative hydraulic conductivities in saturated groundwater flow. *Reviews of Geophysics, 44*, (3).

Scher H. and M. Lax. (1973), Stochastic transport in a disordered solid: 1. Theory. Phys. Rev. B., 7:4491–502.

Sebol, L. A. (2000), Determination of groundwater age using CFC's in three shallow aquifers in Southern Ontario, University of Waterloo, Waterloo.

Shepherd, R. G. (1989), Correlations of permeability and grain-size, *Ground Water*, *27*(5), 633-638.

Snodgrass, M.F. and P.K. Kitanidis (1998), Transmissivity identification through multi- directional aquifer stimulation. *Stochastic Hydrology and Hydraulics*, *12*(5), 299-316.

Stephens, D.B. (1995), *Vadose Zone Hydrology*. Boca Raton, Florida: CRC Press.

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

Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. Math Geol 34:1–22.

Streltsova, T. (1973), Flow near a Pumped Well in an Unconfined Aquifer under Nonsteady Conditions, *Water Resour. Res*., 9(1), 227-235.

Sudicky, E. A. (1986), A Natural Gradient Experiment on Solute Transport in a Sand Aquifer: Spatial Variability of Hydraulic Conductivity and Its Role in the Dispersion Process, *Water Resour. Res*., 22(13), 2069–2082.

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

Sun, N.-Z., and W. W.-G. Yeh (1992), A stochastic inverse solution for transient groundwater flow: parameter identification and reliability analysis, *Water Resour. Res.,* 28(12), 3269-3280.

Tartakovsky, G. D., and S. P. Neuman (2007), Three-dimensional saturated-unsaturated flow with axial symmetry to a partially penetrating well in a compressible unconfined aquifer, *Water Resour. Res.*, 43, W01410, doi:10.1029/2006WR005153.

Teles V., J. P. Bravard, G. de Marsily, and E. Perrier (2001), Modelling of the construction of the Rhône alluvial plain since 15,000 years BP. *Sedimentology* 48: 1209-1224.

Teles V., F. Delay, and G. de Marsily (2004), Comparison between different methods for characterizing the heterogeneity of alluvial media: groundwater flow and transport simulations. *J. Hydrol.* 294(1-3): 103-121.

Theis, C.V. (1935), The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using ground water storage. *Transactions-American Geophysical Union, 16*, 519-524.

Therrien, R., R. G. McLaren, E. A. Sudicky, and S. M. Panday (2005), Hydrogeosphere: A Three-dimensional Numerical Model Describing Fully-integrated Subsurface and Surface Flow and Solute Transport, edited, Groundwater Simulations Group, Waterloo, Ontario, Canada.

van Genuchten, M. T. (1980), A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, 44, 892–898.

Verruijt, A. (1969), Elastic storage of aquifers, in *Flow Through Porous Media*, ed. R.J.M. de Wiest, 331-376. New York: Academic Press.

Vesselinov, V., S. Neuman, and W. 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., S. Neuman, and W. 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.

Webb, E. K. (1995), Simulation of braided channel topology and topography, *Water Resour. Res.* 31(10), 2603-2612.

Webb, E. K. and M. P. Anderson (1996), Simulation of preferential flow in three-dimensional, heterogeneous conductivity fields with realistic internal architecture, *Water Resour. Res*., 32(3), 533-546

Weissmann, G., S. Carle, and G. Fogg (1999), Three-Dimensional Hydrofacies Modeling Based on Soil Surveys and Transition Probability Geostatistics, *Water Resour. Res*., 35(6), 1761-1770.

Wen, J.C., C. Wu., T.C.J. Yeh., and C.M.Tseng (2010), Estimation of effective aquifer hydraulic properties from an aquifer test with multi-well observations (Taiwan). *Hydrogeology Journal, 18*, (5), 1143-1155.

Whittaker, J., and G. Teutsch (1999), Numerical simulation of subsurface characterization methods: Application to a natural aquifer analogue, *Adv. Water Resour*., 22(8), 819–829, doi:10.1016/S0309-1708(98)00056-6.

Wolf, S.H. (1988), Spatial variability of hydraulic conductivity in a sand and gravel aquifer, Eng. thesis, 118 pp., Dep. of Civ. Eng., Mass. Inst. of Technol., Cambridge.

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

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

Ye, M., P. D. Meyer, and S. P. Neuman (2008), On model selection criteria in multimodel analysis, Water Resour. Res., 44, W03428, doi:10.1029/2008WR006803.

Yeh, T.-C. J., R. Srivastava, A. Guzman, and T. Harter (1993), A numerical model for water flow and chemical transport in variably saturated porous media, *Ground Water*, 31(4), 634-644.

Yeh, T.-C. J., A.L. Gutjahr, and M. Jin (1995), An iterative cokriging-like technique for groundwater flow modeling, *Ground Water*, 33(l), 33-41.

Yeh, T.-C.J., J. Mas-Pla, T.M. Williams, and J.F. McCarthy (1995), Observation and three-dimensional simulation of chloride plumes in a sandy aquifer under forced-gradient conditions, *Water Resources Research*, 31(9), 2141-2157.

Yeh, T.-C., M. Jin, and S. Hanna (1996), An iterative stochastic inverse method: conditional effective transmissivity and hydraulic head fields, *Water Resour. Res.*, 32(l), 85-92.

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

Yeh, T.-C.J., and J. Zhu (2007), Hydraulic/partitioning tracer tomography for characterization of dense nonaqueous phase liquid source zones, *Water Resour. Res.*, 43, W06435, doi:10.1029/2006WR004877.

Yeh, T.-C. J., et al. (2008), A view toward the future of subsurface characterization: CAT scanning groundwater basins, Water Resour. Res., 44, W03301, doi:10.1029/2007WR006375.

Yeh, T.-C. J., J. Xiang, R. M. Suribhatla, K.-C. Hsu, C.-H. Lee, and J.-C. Wen (2009), River stage tomography: A new approach for characterizing groundwater basins, *Water Resour. Res.*, 45, W05409, doi:10.1029/2008WR007233.

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

Zhang D. and S. P. Neuman (1996), Effect of local dispersion on solute transport in randomly heterogeneous media. *Water Resour. Res.*, 32(9): 2715–23.

Zheng, C., and S.M. Gorelick (2003), Analysis of solute transport in flow fields influenced by preferential flowpaths at the decimeter scale, *Ground Water*, *41*(2), 142-155, doi:10.1111/j.1745-6584.2003.tb02578.x.

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

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

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