

## Research papers

## On the importance of considering specific storage heterogeneity in hydraulic tomography: Laboratory sandbox and synthetic studies

Zhanfeng Zhao <sup>a,b,\*</sup>, Walter A. Illman <sup>b</sup><sup>a</sup> Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China<sup>b</sup> Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario, Canada

## ARTICLE INFO

This manuscript was handled by C. Corradini, Editor-in-Chief, with the assistance of Okke Batelaan, Associate Editor

## Keywords:

Aquifer heterogeneity  
Specific storage  
Hydraulic tomography  
Model calibration and validation

## ABSTRACT

Over the last two decades, various studies on Transient Hydraulic Tomography (THT) have shown that it is an effective approach to characterize subsurface heterogeneity. Typically, high-resolution hydraulic conductivity ( $K$ ) distributions were recovered, while the spatial variability of specific storage ( $S_s$ ) was found to be smooth. In some studies,  $S_s$  heterogeneity has been intentionally ignored due to the belief that  $S_s$  is less variable than  $K$ . Therefore, one may question the importance of considering  $S_s$  heterogeneity during THT and its impact on the reliability of estimated hydraulic parameters. To investigate these issues, three modeling approaches (i.e., effective parameters, geological, and geostatistics-based) were used to obtain  $K$  and  $S_s$  estimates of varying spatial resolutions. The reliability of  $K$  and  $S_s$  estimates were evaluated by comparing their drawdown prediction performances. The values of using different prior  $K$  and  $S_s$  information for THT analyses were investigated. Our results revealed that: (1) the  $K$  distribution estimated from the geostatistics-based steady-state HT analysis accurately predicted the late time drawdowns, while further improvements in transient drawdowns were obtained only after jointly treating the  $S_s$  field as heterogeneous; (2)  $S_s$  heterogeneity should be considered in addition to  $K$  for THT inversions, even when the estimated  $S_s$  field is smooth; and (3) using  $K$  and  $S_s$  estimates from the calibrated geological model as initial mean distributions for the geostatistical inversion approach were helpful in capturing both interlayer and intralayer heterogeneity of  $K$  and  $S_s$ . These findings suggested that  $S_s$  heterogeneity should be properly considered during the implementation of THT.

## 1. Introduction

The accurate knowledge of hydraulic properties such as hydraulic conductivity ( $K$ ) and specific storage ( $S_s$ ) is critical in groundwater flow modeling, contaminant transport, and water resource management issues. Traditional analytical solutions (e.g., [Theis \(1935\)](#) or [Cooper and Jacob \(1946\)](#)) treat the aquifer to be homogeneous and yield equivalent estimates for hydraulic properties over the support volume of a test, whereas in reality, hydraulic parameters are inherently heterogeneous at multiple scales. In order to deal with subsurface heterogeneity, different laboratory and field characterization approaches, as well as modeling techniques have been developed ([Koltermann and Gorelick, 1996](#); [de Marsily et al., 2005](#); [Hendricks Franssen et al., 2009](#); [Linde et al., 2015](#)). For example, a large number of core samples have been subjected to laboratory permeameter tests ([Sudicky, 1986](#); [Sudicky et al., 2010](#); [Alexander et al., 2011](#); [Smith et al., 2016](#)), slug tests have

been conducted at multiple locations ([Yeh et al., 1995](#); [Aristodemo et al., 2018](#)), flowmeter tests have been performed along multiple wells ([Molz et al., 1989](#); [Rehfeldt et al., 1992](#); [Klepikova et al., 2013](#); [Tamayo-Mas et al., 2018](#)), and direct push measurements ([Dietrich et al., 2008](#); [Liu et al., 2009](#); [Bianchi and Zheng, 2016](#)) have been utilized to delineate the heterogeneous distribution of  $K$  in sedimentary deposits.

In contrast, the distribution of  $S_s$ , which has been typically estimated from methods such as single-hole tests or consolidation experiments (e.g., [Neuman and Witherspoon, 1972](#); [Keller et al., 1989](#); [Shaver, 1998](#); [Quinn et al., 2016](#)), is less frequently considered in many field investigations, as its heterogeneity has been assumed to be much less than  $K$  (e.g., [Dagan, 1989](#); [Gelhar, 1993](#); [Tartakovsky and Neuman, 1998](#); [Lu and Robinson, 2006](#); [Kuang et al., 2020](#)). Based on the linear relationship between solid matrix compressibility and  $K$  proposed by [Freeze \(1975\)](#), [Dagan \(1989\)](#) suggested that  $S_s$  is one order of magnitude less variable than  $K$ . [Hoeksema and Kitanidis \(1985\)](#) found that both  $K$  and

\* Corresponding author.

E-mail address: [zhaozhanfeng@igsnrr.ac.cn](mailto:zhaozhanfeng@igsnrr.ac.cn) (Z. Zhao).

storage coefficient ( $S$ ) values from 31 aquifers can be generally considered as log-normally distributed. In stochastic analyses of transient flow problems, early studies treated  $K$  and  $S_s$  (or  $S$ ) as statistically homogeneous (stationary) variables (e.g., Dagan, 1982; Sun and Yeh, 1992; Zhang, 1999) and  $S_s$  (or  $S$ ) were sometimes taken to be constant for simplicity (e.g., Dagan and Rubin, 1988; Tartakovsky and Neuman, 1998; Panzeri et al., 2013). These assumptions seem plausible when the aquifer is composed of a single geological unit, while can be violated for aquifers with complicated geological structures and facies. To evaluate the impacts of structural heterogeneity on flow, Winter and Tartakovsky (2000, 2002) proposed the composite medium model, in which hydraulic parameters were represented through pairing a random boundary process that determines block geometry, with a stationary process that defines hydraulic properties within a material block. Winter et al. (2002) concluded that the structural variability of  $K$  had a more significant effect on flow prediction than local scale variability within the material block.

In aquifer systems composed of multiple geological materials,  $S_s$  values determined for the same field can vary several orders of magnitude (Younger, 1993; Shaver, 1998; Alexander et al., 2011; Acworth et al., 2017). Wu et al. (2005) stated that, when the aquifer is composed of contrasting storage property blocks, the estimated  $S$  value from traditional aquifer test analyses correlated well with the  $S$  value of the block where the observation point was located in. Thus far, studies have shown that,  $S_s$  or  $S$  as well as their spatial variability play important roles in hydrogeological research such as groundwater flow predictions (Alexander et al., 2011; Quinn et al., 2016), groundwater storage and depletion (Konikow and Neuzil, 2007; Ojha et al., 2018), land subsidence (Domenico and Mifflin, 1965; Hung et al., 2012; Erban et al., 2013; Zhuang et al., 2017), and solute transport under transient flow conditions (Goode and Konikow, 1990; Elfeki et al., 2012; Pool et al., 2016). For example, Elfeki et al. (2012) reported that an increase in  $S$  led to a decrease in plume spatial moments and longitudinal macrodispersion. Inouchi et al. (1990) found that longitudinal and transverse mixing induced by tides in confined coastal aquifers strongly depended on the value of  $S_s$ . Through numerical experiments, Pool et al. (2014) revealed that large storativity caused fluctuations in the flow boundary conditions to propagate through the aquifer with a spatially variable speed, which enhanced mixing and was especially important for contaminant migration. At the regional scale, appropriate identification of aquifer storage parameters is critical for groundwater resource management (Wu et al., 2005; Rodell et al., 2007; Sun et al., 2010; Konikow, 2011; Béjar-Pizarro et al., 2017). Since groundwater availability analyses are directly affected by aquifer storage parameters, inaccurate estimates of  $S_s$  or  $S$  could lead to serious consequences to regional groundwater assessment. In the study of Mississippi River basin, Rodell et al. (2007) stated that uncertainty in specific yield ( $S_y$ ) would significantly change the amplitude of computed groundwater fluctuations, which highlighted the importance of using distributed aquifer storage parameters instead of assuming global mean values. In aquifer systems experiencing significant land subsidence, the rate and magnitude of consolidation relies heavily on the inelastic storage coefficient, while the elastic storage coefficient only has a smaller effect on the rate of subsidence (Don et al., 2005), suggesting that quantification of inelastic storativity is essential for the sustainable water resources management (Erban et al., 2014; Rezaei et al., 2020).

Over the past two decades, hydraulic tomography (HT) has been proposed and developed (e.g., Gottlieb and Dietrich, 1995; Yeh and Liu, 2000), to characterize both  $K$  and  $S_s$  heterogeneity at high resolutions. The fundamental idea of HT is to obtain high-density hydraulic response data from transient state (THT) or steady state (SSHT). Compared to SSHT which only estimates  $K$  distributions, THT that properly considers  $K$  and  $S_s$  heterogeneity enables the simultaneous estimation of  $K$  and  $S_s$  values (Zhu and Yeh, 2005). Thus far, the performance of HT has been demonstrated via laboratory experiments (e.g., Liu et al., 2002, 2007; Illman et al., 2010; Berg and Illman, 2011a; Zhao et al., 2016) and field

studies (e.g., Straface et al., 2007; Bohling et al., 2007; Brauchler et al., 2011, 2013; Berg and Illman, 2011b; Castagna et al., 2011; Cardiff et al., 2012, 2013; Hochstetler et al., 2016; Paradis et al., 2016; Mao et al., 2018; Fischer et al., 2020).

For a number of THT studies, geostatistics-based inverse modeling approaches have been adopted. Typically, high-resolution  $K$  distributions have been recovered through THT inversions, while the spatial variability of  $S_s$  has been found to be smooth or lacking of structural features (e.g., Liu et al., 2007; Berg and Illman, 2011a, 2011b; Zhao and Illman, 2018). In particular, Liu et al. (2007) presented the first validation study of THT in a laboratory sandbox with a synthetic aquifer packed in a deterministic layered pattern. They found that the recovered structural pattern noticeable in the  $K$  tomogram was not mapped for the  $S_s$  tomogram when visually compared with the image of sandbox aquifer. Instead, a general decreasing trend in  $S_s$  values with depth was observed. The smooth nature of the  $S_s$  tomogram was due to various size sands used in the sandbox experiments, while silts and clays were not used. Similar smooth  $S_s$  results have also been reported for other THT studies conducted in different sandboxes (Berg and Illman, 2011a; Luo et al., 2017) and also under field conditions (Straface et al., 2007; Berg and Illman, 2011b).

The lower degree of  $S_s$  heterogeneity observed by various researchers in comparison to  $K$  may have caused some researchers to ignore  $S_s$  variability in their THT analyses (e.g., Li et al., 2005; Cardiff and Barrash, 2011; Tiedeman and Barrash, 2020). In particular, Li et al. (2005) and Cardiff and Barrash (2011) stated that, treating  $S_s$  and/or  $S_y$  as constants during inverse modeling even when these fields are variable, would not lead to dramatic deviations in the estimated  $K$  or  $T$  fields. In contrast, Castagna et al. (2011), through synthetic experiments, concluded that incorrectly treating the spatially heterogeneous  $S$  field to be homogeneous yielded a consistent representation of  $T$  field, but entirely misrepresented the  $S$  field and deteriorated the quality of the inversion. Moreover, Jiménez et al. (2015) showed that the consideration of  $S_s$  heterogeneity in addition to  $K$  in the inversion process could improve the fitting of model results to transient data, while the structural features were only slightly changed. These controversial findings lead to the question whether one should consider  $S_s$  heterogeneity in inverse groundwater modeling or not and whether the reliability of estimated values is impacted or not.

Thus far, only a few published works have tried to address the above issues. Through numerical studies, Sun et al. (2013) suggested a temporal sampling strategy of pressure head data from HT survey, to improve the estimation of heterogeneous  $T$  and  $S$  fields. Based on the cross-correlation analyses between the pressure head and  $S$  values, they suggested using early time " $t_m$ " data from drawdown curves, when the observed head and  $S$  values are most correlated. The time  $t_m$  is approximately equal to the intercept time " $t_0$ " at which the extrapolated drawdown from the first straight portion of the drawdown-log time curve becomes zero. Based on a well-studied laboratory sandbox aquifer initially constructed by Illman et al. (2010), Luo et al. (2017) compared model performances of different conceptualizations for THT analyses and investigated the potential benefits of using both  $K$  and  $S_s$  estimates from calibrated geological models as initial distributions for the geostatistics-based THT inversion approach. They found that geological features were preserved in the estimated  $K$  tomograms, while the estimated  $S_s$  tomogram almost remained as smooth as the case without using any geological information. The work by Luo et al. (2017), together with the findings from other studies (e.g., Wu et al., 2005; Liu et al., 2007; Xiang et al., 2009; Berg and Illman, 2011b; Zha et al., 2019), highlighted that it is more difficult to estimate  $S_s$  heterogeneity than those of  $K$  during THT analyses. Studies have also demonstrated the performances of HT in reconstructing  $S_s$  or  $S_y$  distributions through inverting different types of information such as head and self-potential data (Soueid Ahmed et al., 2016), head and time-lapse gravity data (Tsai et al., 2017), which would require additional experimental efforts.

In one recent THT study conducted at the highly heterogeneous

North Campus Research Site (NCRS), Zhao and Illman (2018) found that estimated  $K$  distributions showed the very heterogeneous double-layer aquifer features of a multi aquifer-aquitard system, but the jointly estimated  $S_s$  tomograms were quite smooth despite the large number of pressure head data used for model calibrations. This is so despite the fact that the  $S_s$  of aquifers and aquitards typically varies over several orders of magnitude. Reflecting this highly variable nature of the aquifer-aquitard system at the NCRS, Alexander et al. (2011), reported  $S_s$  estimates ranging up to five orders of magnitude (from  $2.6 \times 10^{-8}/\text{m}$  to  $3.8 \times 10^{-3}/\text{m}$  with a geometric mean of  $3.1 \times 10^{-5}/\text{m}$ ) based on the type curve analysis of a pumping test. Such different results suggest that  $S_s$  heterogeneity could have been potentially overlooked during the inversion process and additional efforts may be needed in order to obtain  $S_s$  estimates at finer resolutions for the geostatistics-based THT analysis.

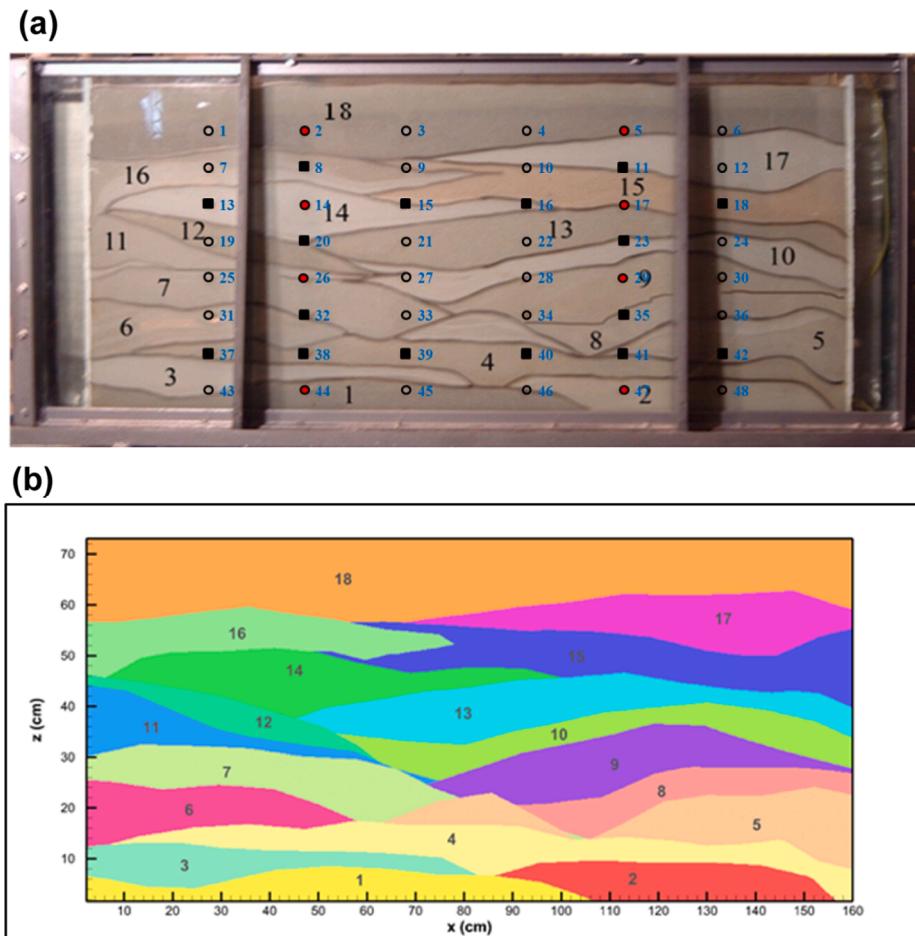
The main purposes of this study are to systematically investigate the importance of  $S_s$  heterogeneity for HT and whether  $S_s$  heterogeneity can be reliably mapped or not, in addition to  $K$  heterogeneity through THT. We first examine the performance of different inverse modeling approaches in a multi-layered laboratory sandbox with known stratigraphy, following the temporal data sampling strategy suggested by Sun et al. (2013), to evaluate whether heterogeneous structures of  $S_s$  could be more clearly mapped. Then, we systematically investigate the reliability of  $K$  and  $S_s$  estimates from various modeling approaches. The impacts of different kinds of prior information were evaluated for THT analyses through synthetic experiments, which contain composite materials similar to the NCRS site. Findings from this study should yield further insights for THT implementations in the field.

## 2. Laboratory sandbox study

### 2.1. Sandbox settings and experimental data

A synthetic 18-layer heterogeneous sandbox aquifer was constructed by Illman et al. (2010), through cyclically depositing sediments of different sizes under varying water flow and sediment feeding rates, to mimic an interfingering deposition of natural fluvial aquifers, as shown in Fig. 1a. The sandbox aquifer is 160.0 cm long, 78.0 cm high, and 10.2 cm wide. At each end of the sandbox, a series of 4-inch perforated plate/mesh combination was installed to serve as a sand/water interface and to provide hydraulic control. Forty-eight fully-penetrating horizontal wells along six columns were installed on the back wall of the sandbox, which can also be utilized for pumping and injection purposes. Each well was instrumented with a 0 to 1 psig (pounds per square inch gauge) Setra model 209 pressure transducer.

To obtain local  $K$  values, constant head permeameter tests were conducted on core samples extracted at 48 port locations. The extracted materials were quite uniform based on visual inspection and showed no obvious layering. Therefore, the anisotropy in  $K$  within each layer was neglected. Table 1 lists the sand types and  $K$  estimates for each layer of the sandbox aquifer. When multiple ports intersect a given layer, a geometric mean of the calculated  $K$  values was provided. It is important to note that  $K$  values can be different for the same sand type due to the impacts of sediment deposition, compaction, in situ coring, and testing. For example, average  $K$  values are different for 20/30 sand layers deposited at various sections of the sandbox aquifer. In particular, the particle diameter for which 50% of the weight is finer ( $d_{50}$ ) is adopted



**Fig. 1.** (a) Front view of the sandbox and (b) “perfect” geological model of the synthetic heterogeneous aquifer constructed by delineating layer boundaries based on the photograph of the laboratory sandbox (after Zhao et al., 2016).

**Table 1**

Sand types, their  $d_{50}$  (mm) and  $K$  (cm/s) estimates obtained through core permeameter measurements and single-hole tests as well as  $S_s$  (1/cm) estimates from single-hole tests for each layer in the sandbox aquifer.

Layer	Sand type	$d_{50}$ (mm)	Core permeameter <sup>a</sup>	Single-Hole $K$ (cm/s) <sup>b</sup>	Single-Hole $S_s$ (1/cm) <sup>b</sup>
1	20/30	0.750	$3.2 \times 10^{-2}$	$5.32 \times 10^{-02}$	$2.12 \times 10^{-04}$
2	4030	0.355	$5.3 \times 10^{-2}$	$5.67 \times 10^{-02}$	$2.60 \times 10^{-04}$
3	F-85	0.151	$7.1 \times 10^{-2}$	$5.70 \times 10^{-02}$	$5.00 \times 10^{-04}$
4	20/40	0.578	$5.7 \times 10^{-2}$	$5.10 \times 10^{-02}$	$2.22 \times 10^{-04}$
5	mix	0.411 <sup>c</sup>	$6.2 \times 10^{-2d}$	$3.88 \times 10^{-02d}$	$5.56 \times 10^{-04d}$
6	mix	0.411 <sup>c</sup>	$8.2 \times 10^{-2}$	$5.00 \times 10^{-02}$	$4.00 \times 10^{-04}$
7	#12	0.525	$1.3 \times 10^{-1}$	$7.35 \times 10^{-02}$	$4.20 \times 10^{-04}$
8	F32	0.504	$1.3 \times 10^{-1}$	$4.50 \times 10^{-02}$	$1.75 \times 10^{-04}$
9	20/40	0.578	$8.7 \times 10^{-2}$	$4.60 \times 10^{-02}$	$2.15 \times 10^{-04}$
10	F-65	0.204	$1.1 \times 10^{-1}$	$8.25 \times 10^{-02}$	$1.14 \times 10^{-03}$
11	#12	0.525	$1.4 \times 10^{-1}$	$2.05 \times 10^{-01}$	$2.15 \times 10^{-04}$
12	16/30	0.872	$3.4 \times 10^{-2}$	$4.95 \times 10^{-02}$	$6.32 \times 10^{-04}$
13	20/30	0.750	$2.6 \times 10^{-1}$	$1.05 \times 10^{-01}$	$9.80 \times 10^{-04}$
14	F-75	0.174	$9.8 \times 10^{-2}$	$5.70 \times 10^{-02}$	$9.80 \times 10^{-04}$
15	20/40	0.578	$8.6 \times 10^{-2}$	$7.50 \times 10^{-02}$	$2.00 \times 10^{-03}$
16	mix	0.411 <sup>c</sup>	$4.2 \times 10^{-2}$	$2.68 \times 10^{-02}$	$7.11 \times 10^{-04}$
17	F-85	0.151	$4.5 \times 10^{-2}$	$4.47 \times 10^{-02}$	$1.14 \times 10^{-03}$
18	20/30	0.750	$1.5 \times 10^{-1}$	$1.16 \times 10^{-01}$	$3.38 \times 10^{-03}$

Note: The layers labelled “mix” consisted of equal volumes of #14, F75, and 16/30 sands.

<sup>a</sup> These core permeameter test results are adopted from Illman et al. (2010).

<sup>b</sup> These single-hole test values are adopted from Berg and Illman (2011a).

<sup>c</sup> The  $d_{50}$  value is estimated as the mean of #14, F75, and 16/30 sands.

<sup>d</sup> The  $K$  and  $S_s$  of layer 5 are estimated as the mean values of layers 6 and 16 with the same sand type.

from Craig (2005) for each different sand type, provided in Table 1.

Single-hole pumping tests were performed in the heterogeneous sandbox aquifer to estimate  $K$  and  $S_s$  values at locations of the 48 ports. The tests were conducted by pumping water from each port at a constant rate (1.25 ml/s) and monitoring the transient head changes at the pumped locations using a pressure transducer. For each test, data collection was started before pumping to collect the initial hydraulic heads at all monitoring ports. A peristaltic pump was then activated and run at a constant flow rate until the development of steady state conditions. The entire transient hydrograph was manually matched using VSAFT2 (Yeh et al., 1993) by treating the aquifer to be homogeneous. The estimated  $K$  and  $S_s$  for different layers are also listed in Table 1. If multiple ports are in the same layer, then the geometric mean of  $K$  or  $S_s$  is presented.

## 2.2. Pumping tests and data selections

We utilized 24 cross-hole tests conducted at different ports with constant pumping rates that ranged from 2.50 to 3.17 ml/s for this study. During each test, the left, right and top faces of the sandbox were hydraulically connected to produce constant head boundaries of 77.5 cm (Illman et al., 2010). Meanwhile, hydraulic head changes were observed at all 48 ports until steady state conditions. For all pumping tests, steady state conditions were reached within 20 s since the initiation of pumping. The pump was then shut off and hydraulic heads were monitored to full recovery.

Eight pumping tests (ports 2, 5, 14, 17, 32, 35, 44, and 47) along two well columns indicated on Fig. 1a were chosen for model calibration, while the other 16 independent pumping tests (ports 8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42) were kept for model validation purposes. Prior to extracting data points from hydrographs, pressure head data from observation ports were fitted with a fifth- or sixth-order polynomial curve (Luo et al., 2017) or denoised using the Wavelet toolbox in Matlab to remove pressure transducer noise. The denoised drawdown curves at selected observation ports are plotted in Figs. S1–S8. The signal-to-noise ratios (SNR) were calculated for each

drawdown curve using the method provided in Xiang et al. (2009) to examine the quality of denoised data. Previously, Xiang et al. (2009) stated that low SNR pumping test data could lead to erroneous estimation of the  $S_s$  field, and the denoised data could improve the  $S_s$  field, although not as good as using noise-free data. They suggested that SNR should be larger than 1 in order to provide useful information of drawdown. As shown in Fig. S9 of the Supplementary Information section, SNR values of the denoised data are generally larger than 1 except for several data points in the very early time period and/or from the observation ports (i.e., ports 1, 3, 4, and 6) at the upper part of the sandbox. For these ports, data were selected at time points having SNR larger than 1.

In studies of the asymptotic analysis approach (e.g., Illman and Tartakovsky, 2006), pressure transients at the intermediate to late time were used to estimate  $K$  and  $S_s$  values. Previous studies, like Leven and Dietrich (2006) through sensitivity analyses, Sun et al. (2013) and Mao et al. (2013a) through cross-correlation analyses, Yeh et al. (2015) through numerical model, have also highlighted the importance of using early time data for accurate  $S$  (or  $S_s$ ) estimation. Therefore, data points that cover the early, intermediate, and late time of aquifer responses were extracted from each drawdown curve. Unlike the data sampling schemes used in previous studies (e.g., Zhu and Yeh, 2005; Liu et al., 2007; Luo et al., 2017), we newly selected three to four data points at and around the time “ $t_0$ ” representing the early time response, as shown in Figs. S1–S8. During the early time, noticeable drawdown develops and propagates to observation wells, which leads to early changes in drawdown-log time curves. The time “ $t_0$ ” was defined as the time at which the extrapolated drawdown from the first straight portion of an observed drawdown-log time curve becomes zero (Sun et al., 2013), which were reported to contain the most information about  $S_s$  heterogeneity in the region between the pumping and observation locations (Wu et al., 2005; Mao et al., 2013a; Sun et al., 2013). During the intermediate and late times, the impact of  $S_s$  on head change diminishes and groundwater flow reaches steady state conditions. Pressure head during this time contains more information about  $K$  heterogeneity (Sun et al., 2013). Thus, one to two data points from intermediate and late time were extracted. In total, 1514 data points were selected from eight pumping tests for model calibration.

## 3. Analysis of laboratory sandbox data

### 3.1. Groundwater flow modeling

In order to examine the necessity of capturing  $S_s$  heterogeneity for THT analyses, we characterized the  $K$  and  $S_s$  distributions of the sandbox aquifer using four different approaches, namely the effective parameter approach, the geological modeling approach, and the geostatistics-based SSHT and THT approaches. For all groundwater flow models considered, the synthetic aquifer was discretized into 741 elements and 1600 nodes with average element dimensions of 4.1 cm  $\times$  4.1 cm  $\times$  10.2 cm. The left, right, and top faces of the sandbox were set as constant head boundaries, while the bottom, front, and back faces were set as no-flow boundaries.

#### 3.1.1. Effective parameter modeling approach

The sandbox aquifer is first treated as homogeneous and isotropic to estimate the effective  $K$  ( $K_{eff}$ ) and  $S_s$  ( $S_{eff}$ ) values, through coupling the forward groundwater flow model MMOC3 (Yeh et al., 1993) with the parameter estimation code PEST (Doherty, 2005). The effective parameter model can be used to describe the overall behaviour of the aquifer when being calibrated to multiple pumping tests. More importantly, this model provides the “baseline scenario” for model calibration and validation performance comparisons among more sophisticated groundwater flow models. Furthermore, the estimated  $K_{eff}$  and  $S_{eff}$  values can be used as initial guesses of hydraulic parameters to facilitate the calibration of other groundwater flow models.

### 3.1.2. Geological modeling approach

Construction of a geological model is the standard approach in building groundwater flow and transport models. Typically, geological models are built by interpolating stratigraphy information obtained from borehole logs. For this study, we constructed a geological model utilized for forward and inverse groundwater flow modeling consisting of multiple zones (Fig. 1b) based on the layering information shown on the glass (Fig. 1a). That is, we assume perfect knowledge of stratification for the entire sandbox aquifer, which is not available with current field technology. Unlike the geological models constructed by Luo et al. (2017) through interpolations of borehole logs of varying accuracy, this geological model can be considered as the “best-case scenario” in terms of knowledge of stratification. In total,  $K$  and  $S_s$  of 18 layers were estimated by coupling MMOC3 (Yeh et al., 1993) with PEST (Doherty, 2005).

### 3.1.3. Geostatistical inverse modeling approaches: SSHT and THT

In order to characterize the heterogeneous  $K$  and  $S_s$  fields at a resolution higher than the effective and geological modeling approaches, we performed geostatistical inverse analyses of eight pumping tests using the Simultaneous Successive Linear Estimator (SimSLE) code (Xiang et al., 2009). THT analysis was performed to simultaneously estimate  $K$  and  $S_s$  values, while SSHT analysis was used to estimate the  $K$  distribution only. Both analyses were run using the same pumping and observation densities. However, unlike inverting all 1514 data points for THT analysis, only 374 steady state pressure heads were chosen for the SSHT analysis.

The geostatistical inversion approach implemented in SimSLE treats the natural log values of hydraulic parameters (e.g.,  $\ln K$  and/or  $\ln S_s$ ) of the heterogeneous porous media as stochastic processes. With given prior information, like the unconditional means, variances ( $\sigma^2_{\ln K}$ ,  $\sigma^2_{\ln S_s}$ ), and correlation scales ( $\lambda_x$ ,  $\lambda_y$ , and  $\lambda_z$ ), SimSLE creates the initial parameter fields through conditioning on available measurements of hydraulic parameters with cokriging. The parameter field is then used to solve the transient or steady state flow equation to obtain simulated pressure heads. Based on the differences between simulated and observed pressure heads, SimSLE iteratively improves the estimated hydraulic parameter fields, until the difference between the two continuous estimated hydraulic parameter fields or the largest difference between simulated and observed heads is smaller than a specified tolerance. We assumed that the hydraulic parameter fields can be described by the exponential covariance model. The initial values of  $K$  and  $S_s$  were set as  $K_{eff}$  and  $S_{eff}$ , respectively, while  $\sigma^2_{\ln K} = \sigma^2_{\ln S_s} = 3.0$ ,  $\lambda_x = 50$  cm,  $\lambda_y = 10.2$  cm, and  $\lambda_z = 10.0$  cm in order to be consistent with a previous investigation (Zhao et al., 2016).

### 3.2. Model calibration and validation

All calibration runs for groundwater flow models were performed on the same PC with a quad-core CPU and 16 GB of Random Access Memory. Calibrations of the effective parameter and geological models were performed by minimizing an objective function that represents a weighted sum of squared differences between simulated and observed pressure heads. Model parameters were iteratively optimized by PEST based on the derivatives of all observations with respect to all adjustable parameters. PEST is a versatile and model-independent parameter estimation program, which could be readily coupled with other types of forward models. The Levenberg-Marquardt method implemented in PEST was utilized to minimize the objective function. The optimization process converged when the lowest four objective function values were within a relative distance of 0.05, defined as:

$$(\Phi_i - \Phi_{min}) / \Phi_i \leq 0.05 \quad (1)$$

where  $\Phi_i$  is the objective function value at the end of the  $i$ th optimization iteration,  $\Phi_{min}$  is the lowest objective function achieved during  $i^{\text{th}}$  to

$(i-3)^{\text{th}}$  optimization iteration. The lowest objective functions during the calibrations of the effective parameter and geological models were summarized in Table S1.

The effective parameter model calibration took five optimization iterations (28 PEST model calls), since only one pair of  $K$  and  $S_s$  was estimated. Each “model call” consisted of one forward simulation of eight pumping tests run by MMOC3 (Yeh et al., 1993) and parameter estimation run by PEST (Doherty, 2005). With further conceptualizing the sandbox aquifer as a stratigraphy-constrained geological model, it took 10 optimization iterations (576 PEST model calls) to converge for 18 pairs of  $K$  and  $S_s$  values. On the other hand, the geostatistical inversion approach based on SimSLE does not rely on geological structures. It instead uses pixel-by-pixel correlation relationships between hydraulic parameters (e.g.,  $K$  or  $S_s$ ) and pressure heads at observed locations, represented as cross-covariance coefficients that evolve dynamically to reflect the incorporation of head data. For the highly parameterized geostatistical inversion models, we selected inversion results from the iteration step at which the  $L_2$  norms have stabilized, indicating the convergences of the inversion processes as suggested by Xiang et al. (2009).

To validate the estimated  $K$  and  $S_s$  distributions from all models, the 16 pumping tests not used in model calibration efforts were simulated. The quality of matches between simulated and observed drawdowns was assessed through scatterplots and the best-fit lines and coefficient of determination ( $R^2$ ). The mean absolute error ( $L_1$ ) and mean square error ( $L_2$ ) were also calculated for each pumping test to further evaluate the fits between observed and simulated drawdowns. The  $L_2$  norm magnifies large discrepancies and allows one to better assess the different models. Those quantities were computed as:

$$L_1 = \frac{1}{n} \sum_{i=1}^n |\chi_i - \hat{\chi}_i| \quad (2)$$

$$L_2 = \frac{1}{n} \sum_{i=1}^n (\chi_i - \hat{\chi}_i)^2 \quad (3)$$

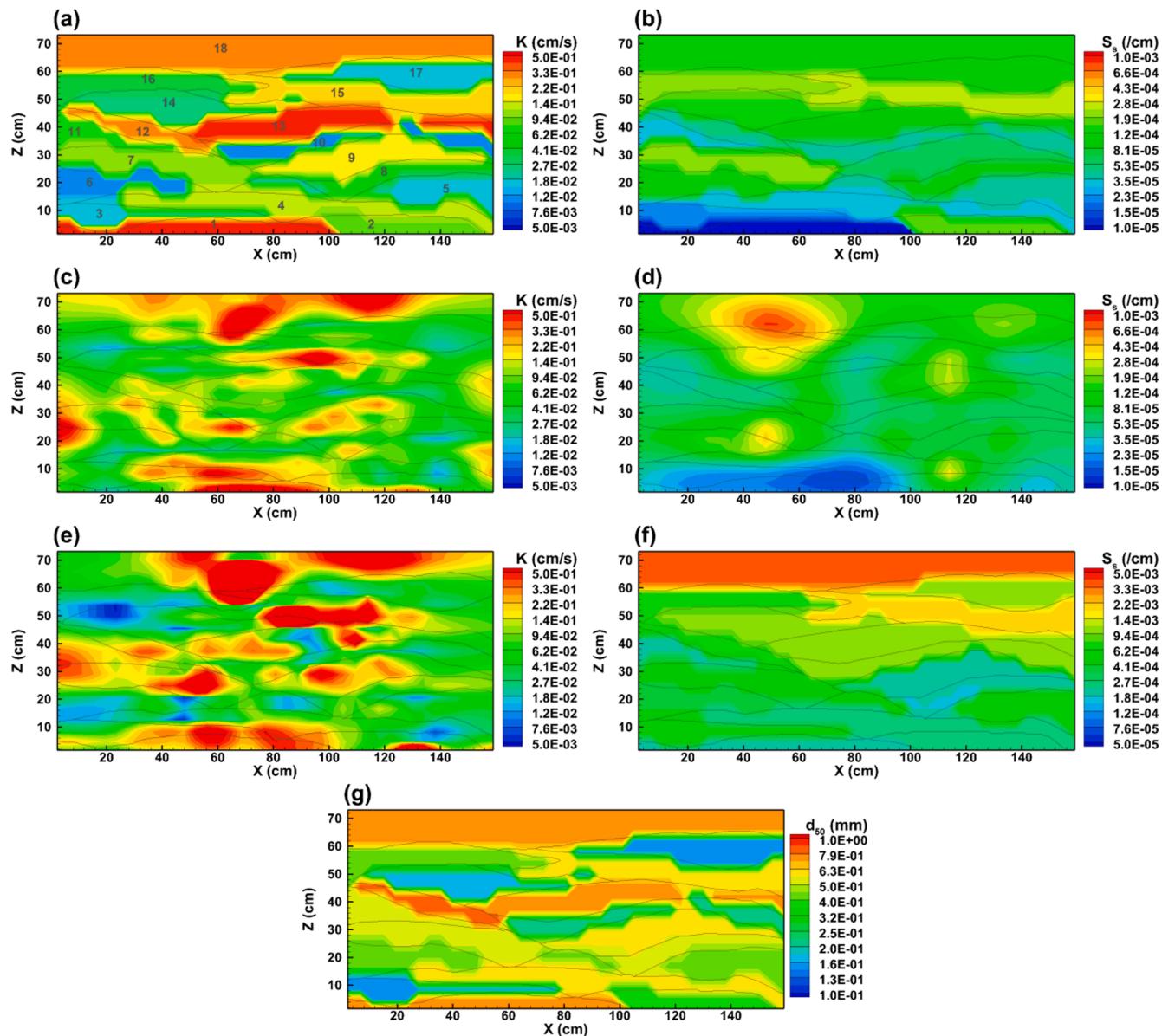
where  $n$  is the total number of drawdowns,  $i$  indicates the data number,  $\chi_i$  and  $\hat{\chi}_i$  represent the simulated and measured drawdowns, respectively.

### 3.3. Results from the laboratory sandbox

#### 3.3.1. $K$ And $S_s$ estimates of different modeling approaches

The estimated  $K$  value of the effective parameter approach is  $8.81 \times 10^{-2}$  cm/s, with lower and upper 95% confidence limits of  $8.64 \times 10^{-2}$  cm/s and  $9.00 \times 10^{-2}$  cm/s, respectively. The estimated  $S_s$  value is  $9.37 \times 10^{-5}$ /cm, with lower and upper 95% confidence limits of  $8.95 \times 10^{-5}$ /cm and  $9.81 \times 10^{-5}$ /cm, respectively. The effective parameters have previously been determined by taking geometric means of 48 single-hole  $K$  and  $S_s$  estimates ( $K = 6.0 \times 10^{-2}$  cm/s and  $S_s = 6.1 \times 10^{-4}$ /cm) (Berg and Illman, 2011a), from calibration of a transient flow model ( $K = 9.57 \times 10^{-2}$  cm/s and  $S_s = 6.32 \times 10^{-5}$ /cm) (Luo et al., 2017), and from calibration of steady state flow model ( $K = 8.9 \times 10^{-2}$  cm/s) (Illman et al., 2015). Our estimated value of  $K$  is close to that from Illman et al. (2015) and  $S_s$  value is slightly higher than that from Luo et al. (2017). Given the large number of transient data points selected from eight pumping tests, especially the increased number of early time data around the time point “ $t_0$ ” as defined by Sun et al. (2013), the  $K$  and  $S_s$  estimates from this study should be sufficient for predicting the averaged spatial trend of observed heads in a heterogeneous aquifer (Yeh et al., 2015).

The  $K$  and  $S_s$  distributions estimated from the calibration of the perfect geological model were plotted as Figs. 2a and b, and corresponding values were listed in Table S2. Figs. 2c and d were the estimated  $K$  and  $S_s$  fields from the THT analysis, respectively, while the estimated  $K$  distribution from the SSHT analysis was plotted as Fig. 2e.



**Fig. 2.** Contour plots of hydraulic parameters for: (a)  $K$  and (b)  $S_s$  from the calibration of the “perfect” geological model; (c)  $K$  and (d)  $S_s$  from THT analysis; (e)  $K$  obtained from SSHT analysis; (f) and (g) geological model populated with single-hole  $S_s$  and  $d_{50}$  values, respectively.

We also populated the geological model (Fig. 1b) with  $S_s$  values estimated from single-hole tests (shown as Fig. 2f), for visual comparisons of  $S_s$  values obtained from different approaches. Black lines in Fig. 2 indicate the true layer boundaries delineated in the photograph of the sandbox aquifer (Fig. 1a).

We see from Figs. 2a, f and Table S2 that, the estimated  $K$  values of the calibrated geological model ranged from a low  $K$  value of  $9.69 \times 10^{-3}$  cm/s in layer 6 to a high  $K$  value of  $4.55 \times 10^{-1}$  cm/s in layer 1, while the estimated  $S_s$  values ranged from a low  $S_s$  value of  $6.46 \times 10^{-6}$  cm in layer 1 to a high  $S_s$  value of  $2.56 \times 10^{-4}$  /cm in layer 15. Specifically, the estimated  $K$  values generally increased with increasing grain sizes ( $d_{50}$  in Table 1, plotted as Fig. 2g), showing a low value range of  $1.06 \times 10^{-2}$  cm/s to  $2.46 \times 10^{-2}$  cm/s for layers (3, 10, 14, and 17) with fine sands ( $d_{50} = 0.151\text{--}0.204$  mm), a middle range of  $4.25 \times 10^{-2}$  cm/s to  $1.99 \times 10^{-1}$  cm/s for layers (2, 4, 7, 8, 9, 11, and 15) with medium sands ( $d_{50} = 0.355\text{--}0.578$  mm), and a high value range of  $2.93 \times 10^{-1}$  cm/s to  $4.55 \times 10^{-1}$  cm/s for layers (1, 13, 12, and 18) with coarse sands ( $d_{50} = 0.750\text{--}0.872$  mm). This is reasonable since that coarse sand is more permeable than the fine sand. On the other hand, such a trend had

not been obviously identified when comparing the estimated  $S_s$  values in Table S2 with grain size parameter  $d_{50}$  in Table 1, and comparing Fig. 2b with Fig. 2g. However, clear structural features have been captured for the  $S_s$  tomogram in Fig. 2b, which had not been previously identified by Luo et al. (2017). The improved structural distinctions in Fig. 2b could be attributed to the inclusion of increased number of early time response data points at and around the time “ $t_0$ ” into the model calibration procedure for the current study.

By definition,  $S_s$  is the volume of water released from storage of a unit volume of aquifer due to the expansion of the water and compression of the aquifer skeleton under a unit change in average head, which has been mathematically expressed as  $S_s = \rho g(\alpha + n\beta)$ , where  $\rho$  is water density [M/L<sup>3</sup>],  $g$  is acceleration of gravity [L/T<sup>2</sup>],  $n$  is matrix porosity [-],  $\alpha$  and  $\beta$  are compressibility of aquifer skeleton and water [LT<sup>2</sup>/M], respectively. Although the estimated  $S_s$  values in the calibrated geological model (Table S2) showed no clear correlation with  $d_{50}$  values of different sands (Table 1) in our sandbox study, we can still see that the estimated  $S_s$  values for layers with the same sand type generally showed a decreasing trend with increasing depth. Specifically, for sand type 20/

30 (in layers 1, 13, and 18), the estimated  $S_s$  value is  $1.13 \times 10^{-04}/\text{cm}$  for the top layer 18,  $9.45 \times 10^{-05}/\text{cm}$  for layer 13 in the center, and  $6.46 \times 10^{-06}/\text{cm}$  for layer 1 at the bottom of the sandbox. Similar trends were also shown for sand types F-85 (in layers 3 and 17) and 20/40 (in layers 4, 9, and 15). This trend was also evident in single-hole test  $S_s$  values (Table 1 and Fig. 2f), although  $S_s$  estimates from single-hole tests ( $S_s = 1.75 \times 10^{-04}$ – $3.38 \times 10^{-03}/\text{cm}$ ) were nearly one order of magnitude larger than those from the calibrated geological model ( $S_s = 6.46 \times 10^{-06}$ – $2.56 \times 10^{-04}/\text{cm}$ ).

Note that the color scale in Fig. 2f was changed to better show the distribution pattern of data. This decreasing trend of  $S_s$  with depth can be physically plausible as sands in the upper part of the sandbox were less compressed by overlying materials than deeper sands (e.g., Liu et al., 2007; Brauchler et al., 2013; Paradis et al., 2016). Given the fact that uncertainty in the  $S_s$  estimate is substantial for different characterization approaches (Rau et al., 2018; Kuang et al., 2020), rigorous derivations of  $S_s$  values based on measurable physical properties like  $\alpha$ ,  $n$ , and  $\beta$  from laboratory consolidation tests would certainly provide useful insights, to examine our estimates from the interpretation of pumping tests and single-hole tests data. Yet a detailed study of quantifying spatial  $S_s$  heterogeneity through sampling and laboratory testing has never been performed for both laboratory and field hydraulic tomography analyses. This is partly due to the fact that results obtained from disturbed samples were deemed to be questionable in representing in situ conditions (e.g., Smith et al., 2013). Moreover, subjecting the miniature samples from this sandbox for consolidation tests was also not possible. Instead, single-hole tests were conducted to obtain local estimates of  $S_s$  by Illman et al. (2010). Therefore, we further justify our  $K$  and  $S_s$  tomograms in the following sections through predictions of 16 independent pumping tests not used in model calibration efforts. In particular, accurate predictions of early to intermediate time drawdowns most sensitive to  $S_s$ , should logically lead to the conclusion that  $S_s$  estimates are more reliable, than when the predictions are poor.

The estimated  $K$  tomograms from THT (Fig. 2c) and SSHT (Fig. 2e) analyses showed distribution patterns similar to Fig. 2a, although the high and low  $K$  layer boundaries and connectivities were not as consistent as those of the calibrated geological model when being compared to the black lines. Compared to the  $S_s$  tomogram estimated through the calibration of the geological model (Fig. 2b), the  $S_s$  tomogram obtained from THT analysis (Fig. 2d) was still smooth and nearly same to the results in Luo et al. (2017), both showing an obviously low value zone at the bottom of the aquifer. Results of Figs. 2b and d suggested that impacts of using an increased number of early time data on revealing  $S_s$  heterogeneity could be different for different modeling approaches. Thus, using a simple geological model instead of a geostatistical model as a first-cut approach is likely to be more favourable in capturing structural features of  $S_s$  fields if a geological model could be reliably constructed.

These results suggested that information other than early time response data (e.g., geological and geophysical information, local  $S_s$  data from slug and/or single-hole tests, information on aquifer/aquitard compressibility) might have to be jointly considered for the geostatistical inversion approach in order to yield improved  $S_s$  estimates. Meanwhile, when comparing Figs. 2b and d with Fig. 2f, we found that  $S_s$  distributions of the calibrated geological model and the THT analysis were quite different from the geological model populated with single-hole test  $S_s$  values.

### 3.3.2. Reliability of $K$ and $S_s$ estimates for drawdown predictions

In order to further evaluate the reliability of estimated  $K$  and  $S_s$  values, we systematically compared the drawdown prediction performance of 16 independent pumping tests by six groundwater models using combinations of  $K$  and  $S_s$  fields obtained from different characterization approaches (the effective parameter model, the geological model populated with single-hole test data, the calibrated geological model, and the geostatistical models). These six models were named as:

(1) Case 1a: Homo, using estimated  $K_{eff}$  and  $S_{s,eff}$  values from the effective parameter modeling approach; (2) Case 1b: SSHT +  $S_{s,eff}$ , using the high resolution  $K$  field from SSHT and  $S_{s,eff}$  value from the effective parameter modeling approach; (3) Case 1c: SSHT +  $S_{s,sgh}$ , using  $K$  field from SSHT and  $S_s$  field of the geological model populated with  $S_s$  values from single-hole tests; (4) Case 1d: SSHT +  $S_{s,geo}$ , using  $K$  field from SSHT and  $S_s$  field from the calibrated geological model; (5) Case 1e: GeoModel, using both  $K$  and  $S_s$  fields from the calibrated geological model; and (6) Case 1f: THT, using both  $K$  and  $S_s$  fields from THT. By comparing the results of these models, we could examine the performance and benefits of conceptualizing the heterogeneous  $K$  and  $S_s$  fields of groundwater models to varying resolutions.

The simulated versus observed drawdowns at several time points (0.25, 0.5, 1, 2, 5, and 15 s) from the drawdown curves of 16 pumping tests were selected and shown as Fig. 3. The linear fit results were also included in each subfigure. In particular, the simulated drawdown curves for Cases 1a, 1b, 1d, and 1f were plotted for all 16 pumping tests at 15 selected observation ports to compare with the observed drawdowns (Fig. 4 and Figs. S11–S25 in the supplementary information). Comparing Fig. 3a with Fig. 3b, the coefficient of determination ( $R^2$ ) value increased from 0.83 for Case 1a to 0.95 for Case 1b. From Fig. 4, we see that the homogeneous model in Case 1a poorly predicted the drawdown curves for early, intermediate, and late time for most of the selected observation ports, while using the heterogeneous  $K$  field of high resolution in Case 1b accurately simulated the late time drawdowns. Since  $S_s$  heterogeneities were ignored in both cases, these results suggested that  $K$  heterogeneity have to be considered in order to accurately predict steady state groundwater flow fields.

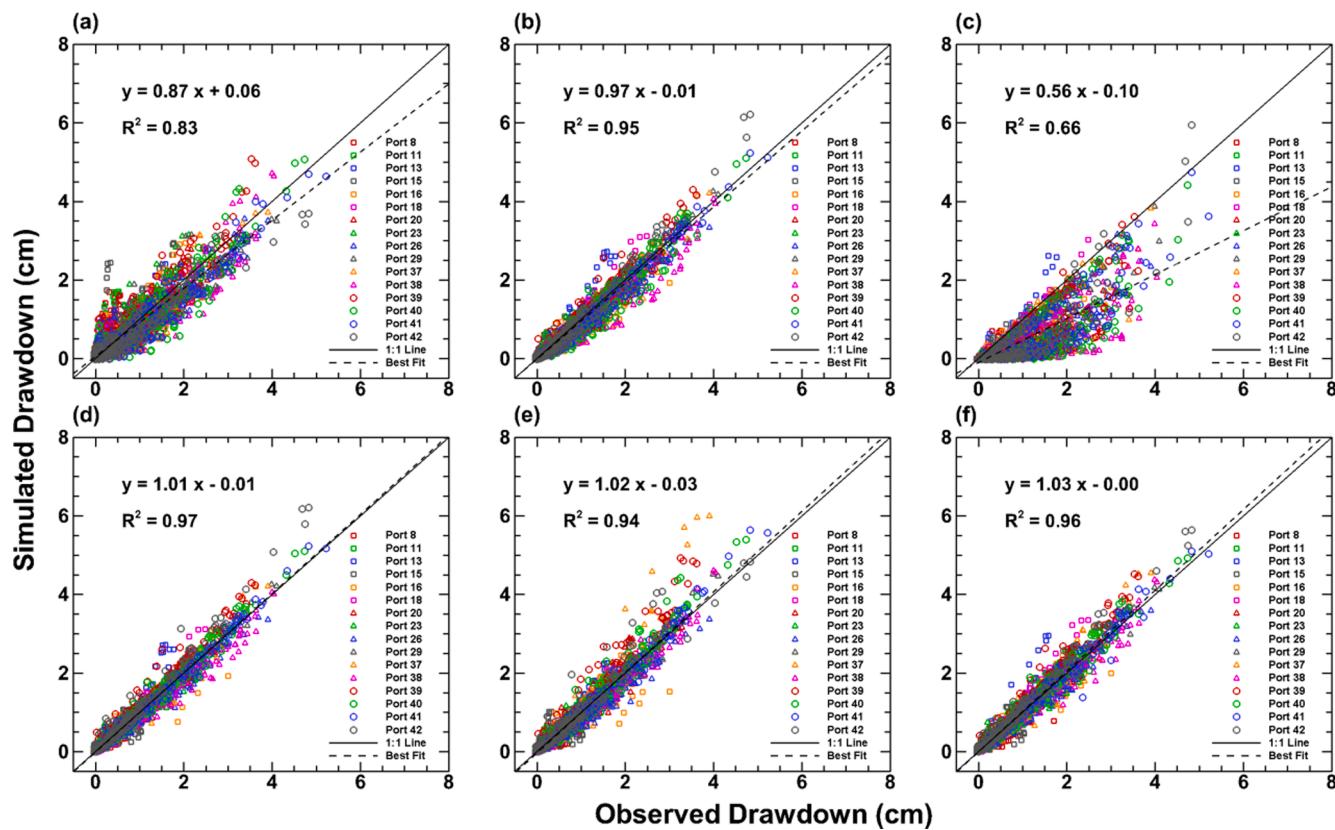
Comparisons among Figs. 3b–d revealed that: (1) drawdown predictions were improved when treating the  $S_s$  values as a geologically distributed field (Fig. 3d), instead of as a homogeneous field (Fig. 3b), and the improvements originated from the more accurate predictions of the transient drawdowns at some observation ports (e.g., ports 13, 25, 30, 37, 39, 42), as shown in Fig. 4 and Figs. S11–S25; (2) drawdown predictions were significantly biased in Fig. 3c for Case 1c, indicating that  $S_s$  values estimated through single-hole tests should be used with caution for groundwater flow models.

When conceptualizing the sandbox aquifer as perfect geological and geostatistical models, drawdown predictions of both models were satisfactory and similar (Figs. 3e and f), although the latter had a slightly higher  $R^2$  value and a less scattered fit than the former. These findings, on one hand, revealed that reliable  $S_s$  estimates can be obtained through the calibration of an accurate geological model if such a model can be obtained, and on the other hand, suggested that the  $S_s$  field of the sandbox aquifer was not very heterogeneous.

### 3.3.3. Impacts of ignoring $S_s$ heterogeneity in THT

Although the  $K$  and  $S_s$  tomograms of the geostatistical inversion approach had predicted drawdowns of 16 pumping tests most accurately, the estimated  $S_s$  tomogram from THT was smooth (Fig. 2d) and the heterogeneity of  $S_s$  in this laboratory sandbox aquifer was milder in comparison to that of the  $K$ . Based on these results, another THT analysis (Case 2) was performed to investigate whether  $S_s$  heterogeneity of the synthetic sandbox aquifer could be ignored for the inverse modeling procedure. That is, the  $K$  field was treated as heterogeneous to be estimated by the geostatistical inversion approach implemented in SimSLE, while the  $S_s$  field was treated to be homogeneous by fixing the value equalling  $S_{s,eff}$ . All other initial inputs (e.g.,  $\sigma_{lnK}^2$ ,  $\lambda$ ) and data points selected for model calibration were kept the same as in Case 1f.

Fig. 5 showed the estimated  $K$  tomogram for Case 2 and validation scatterplots for 16 pumping tests were presented as Fig. 6. Compared to Case 1f (Fig. 2c) which treated both  $K$  and  $S_s$  fields as spatially heterogeneous, Fig. 5 showed similar distribution patterns of high and low  $K$  value zones in the tomogram. However, some noticeable differences can be found between Figs. 2c and 5 in terms of geometries of high  $K$  value zones around the boundaries of the modeling domain, and connectivities



**Fig. 3.** Validation scatterplots of drawdown predictions for: (a) Case 1a: Homo, using estimated  $K_{eff}$  and  $S_{eff}$  values from the effective parameter modeling approach; (b) Case 1b: SSHT +  $S_{s,eff}$ , using the  $K$  field from SSHT and a  $S_{eff}$  value from the effective parameter modeling approach; (c) Case 1c: SSHT +  $S_{s,sgh}$ , using the  $K$  field from SSHT and a  $S_s$  field of the geological model populated with  $S_s$  values from single-hole tests; (d) Case 1d: SSHT +  $S_{s,geo}$ , using the  $K$  field from SSHT and a  $S_s$  field from the calibrated geological model; (e) Case 1e: GeoModel, using both  $K$  and  $S_s$  fields from the calibrated geological model; and (f) Case 1f: THT, using both  $K$  and  $S_s$  fields from THT.

of high  $K$  zones in the central layers (layers 7, 9, 13, and 15) of the sandbox. Meanwhile, drawdown predictions of 16 pumping tests of Case 2 were slightly deteriorated in terms of  $R^2$  values, revealing a larger scatter, when comparing Fig. 6 with Fig. 3f.

Through cross-correlation analysis, Wu et al. (2005) and others (Mao et al., 2013a; Sun et al., 2013) found that the observed head was only correlated with  $S$  (or  $S_s$ ) within a narrow region between the pumping and observation wells during the early stage of pumping tests, while the observed head was correlated with  $T$  (or  $K$ ) through early, intermediate, and late times in the cone of depression. Giving the fact that the number of unknown parameters was far more than the number of observation locations in our sandbox study, the necessary conditions (Mao et al., 2013b; Yeh et al., 2015) enabling the inverse problem to be well-defined were not met. For such ill-posed estimation problems of THT, the geostatistical inversion scheme implemented in SimSLE (Xiang et al., 2009) seeks the most likely, conditional, effective  $K$  and  $S_s$  fields, honouring the head measurements at sampling locations. The estimated  $K$  and  $S_s$  fields were then iteratively updated by the estimator based on the differences between the simulated and observed heads, relying on cross-correlation coefficients between the heads and parameters to be estimated. In Case 2, when  $S_s$  heterogeneity was intentionally ignored, the head differences caused by  $K$  and  $S_s$  heterogeneity were attributed solely to the  $K$  field, leading to noticeable differences that can be found between Figs. 2c and 5.

Tables 2 and 3 summarized the geometric means ( $K_G$ ,  $S_{sG}$ ), variances ( $\sigma^2_{lnK}$ ,  $\sigma^2_{lnS_s}$ ), and correlation lengths ( $\lambda_x$ ,  $\lambda_z$ ) fitted to the exponential model for the resulting  $K$  and  $S_s$  tomograms of the geological model, SSHT and THT cases. Generally, the values of  $K_G$  and  $S_{sG}$  were very close among different cases, and also close to the  $K$  ( $8.81 \times 10^{-2}$  cm/s) and  $S_s$  ( $9.37 \times 10^{-5}$ /cm) values estimated by the effective parameter model,

suggesting that reliable mean values could be obtained by different model approaches for the sandbox. On the other hand, there were some noticeable differences in the estimated variances and correlation lengths. When the same transient dataset was used for inversions, the estimated  $\sigma^2_{lnK}$  of the geological model was nearly twice the value of the THT case 1f, and the difference between estimated  $\sigma^2_{lnS_s}$  for both modeling cases was smaller. This finding implied that an accurate geological model might be more favourable over the highly parameterized geostatistical model to reveal the interlayer heterogeneity pattern, benefiting from the stratigraphy information for the full simulation domain, while the latter had to rely solely on the hydraulic response data to reconstruct it. However, this is based on the premise that an accurate geological model could be obtained.

When only using the steady state data for model calibration, the SSHT case obtained  $K_G$ ,  $\sigma^2_{lnK}$ ,  $\lambda_x$ , and  $\lambda_z$  values that were similar to the geological modeling approach, while the estimated  $\sigma^2_{lnK}$  was twice the value of the THT case 1f (Table 2). The large differences in  $\sigma^2_{lnK}$  values of SSHT and THT case 1f could be attributed to the fact that the former captured more low  $K$  zones than the latter.

This result suggested that SSHT analyses might be a better choice to estimate  $K$  tomograms, when steady state pumping test data are available, since the number of unknowns to be inversely estimated by SSHT is only half of that of the THT case.

When incorrectly treating the heterogeneous  $S_s$  field to be homogeneous, the  $\sigma^2_{lnK}$  of the estimated  $K$  field for THT case 2 was higher than that of the THT case 1f (Table 2), confirming our computational analysis that the head differences caused by  $K$  and  $S_s$  heterogeneity were solely reflected in the  $K$  field.

The influence of  $S_s$  heterogeneity on model calibrations and validations were quantitatively evaluated through comparing average  $L_1$  and

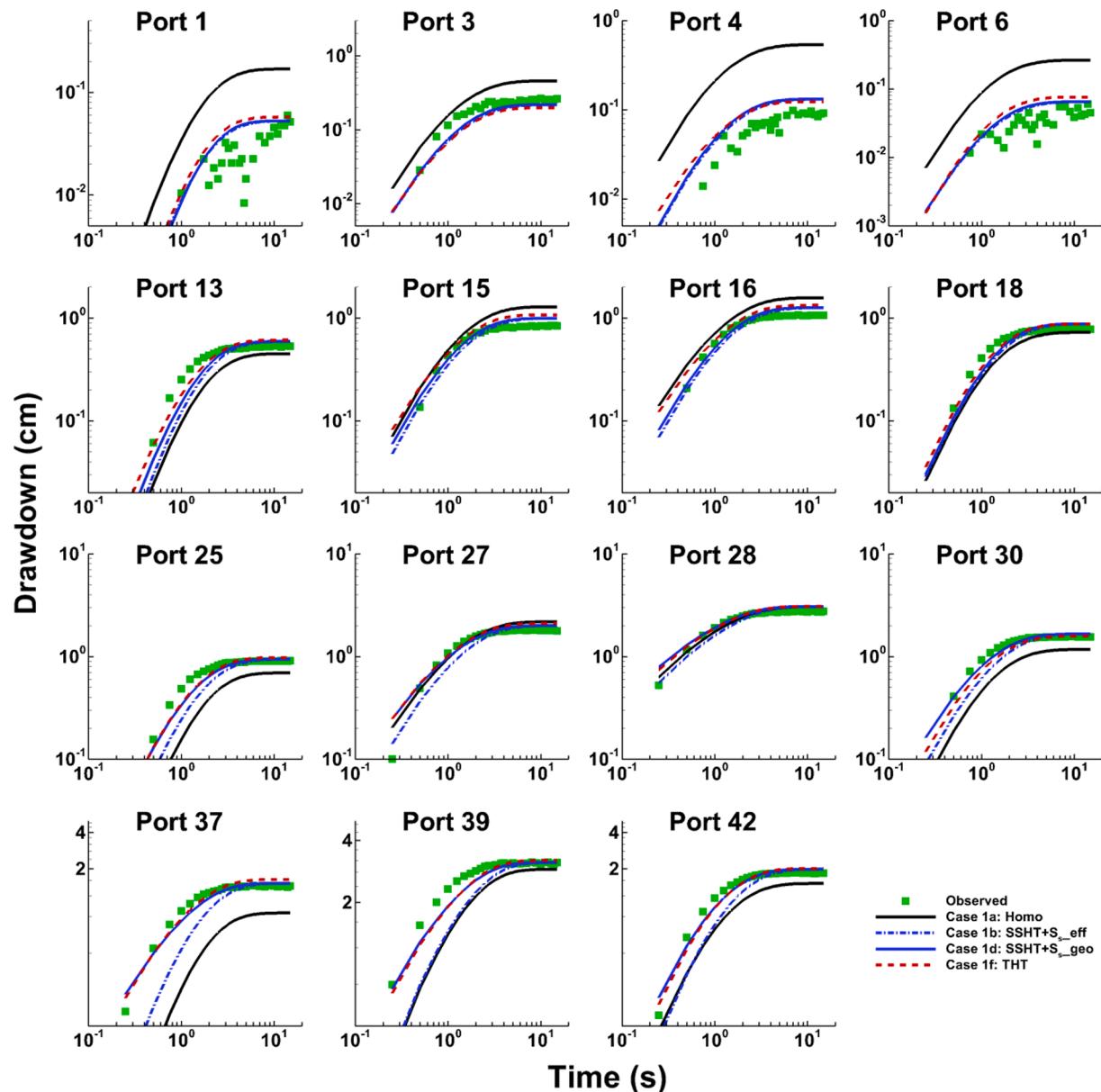


Fig. 4. Simulated versus observed drawdown curves at 15 selected observation ports when conducting a pumping test at port 40.

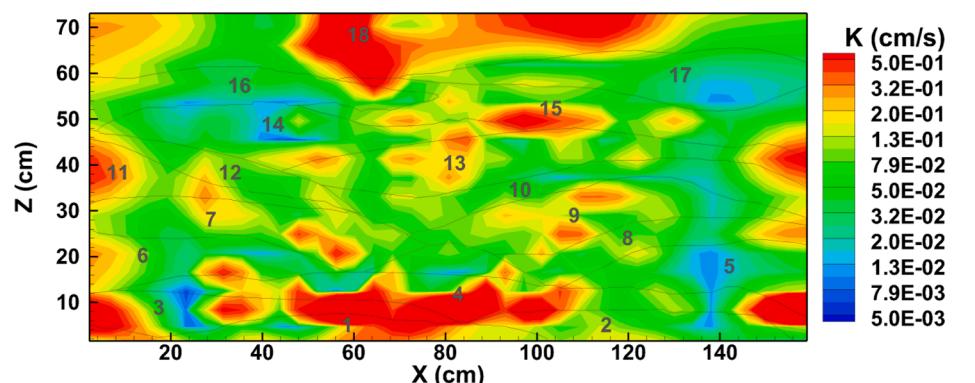
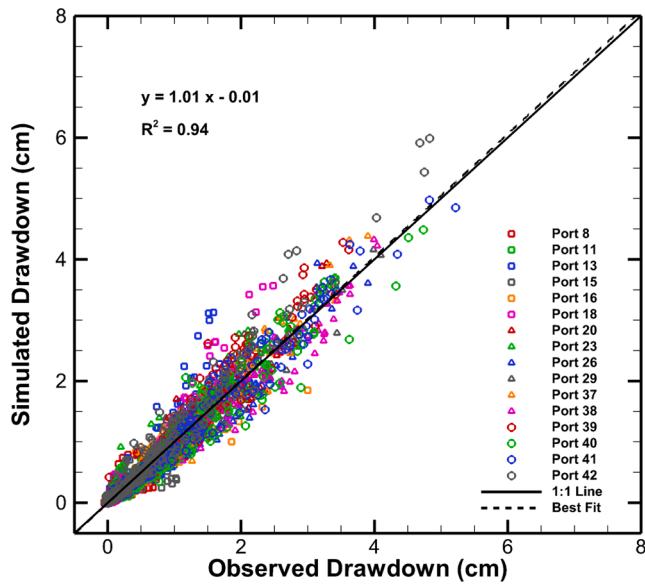


Fig. 5. The estimated  $K$  tomogram from the THT analysis (Case 2) using eight pumping tests with  $S_s$  being treated to be fixed and uniform with a value of  $9.37 \times 10^{-5} / \text{cm}$ . Thin black lines represent layer boundaries.



**Fig. 6.** Validation scatterplots of drawdown predictions for 16 pumping tests in Case 2 which treats  $K$  as heterogeneous and  $S_s$  as a fixed and uniform value of  $9.37 \times 10^{-5}$  cm.

**Table 2**  
Statistical properties of the estimated  $K$  fields.

Case	$K_G$ (cm/s)	$\sigma_{lnK}^2$	$\lambda_x$ (cm)	$\lambda_z$ (cm)	Model
GeoModel	$8.98 \times 10^{-2}$	1.7	24.8	12.3	Exponential
SSHT	$9.07 \times 10^{-2}$	1.8	25.3	13.6	Exponential
THT (Case 1f)	$9.40 \times 10^{-2}$	0.9	25.5	14.3	Exponential
THT (Case 2)	$9.23 \times 10^{-2}$	1.2	16.3	13.6	Exponential

**Table 3**  
Statistical properties of the estimated  $S_s$  fields.

Case	$S_{SG}$ (/cm)	$\sigma_{lnS_s}^2$	$\lambda_x$ (cm)	$\lambda_z$ (cm)	Model
GeoModel	$7.52 \times 10^{-5}$	0.8	26.1	14.1	Exponential
SSHT	N/A	N/A	N/A	N/A	Exponential
THT (Case 1f)	$8.22 \times 10^{-5}$	0.6	21.1	16.8	Exponential
THT (Case 2)	$9.37 \times 10^{-5}$	N/A	N/A	N/A	Exponential

N/A denotes that data are not available for SSHT analysis and Case 2 using homogeneous  $S_s$  field.

$L_2$  values obtained for different pumping tests.  $L_1$  and  $L_2$  values of model calibrations were summarized in Fig. S10 presented in the Supplementary Information, for the effective parameter model (Case 1a: Homo), the “perfect” geological model (Case 1e: GeoModel), and the

geostatistics-based inversion models (Case 1f, THT; Case 2, THT with fixed  $S_s = 9.37 \times 10^{-5}$  cm). Note that  $L_1$  and  $L_2$  values of the SSHT analysis case are not listed since it used a different number of pressure head data for model calibration (i.e., only steady state). We also calculated the  $L_1$  and  $L_2$  norms for the drawdown predictions of each of the 16 pumping tests in Cases 1a–1f and Case 2.

Based on Fig. S10 and Fig. 7, we can see that influences of the  $S_s$  heterogeneity on pressure head simulations are noticeable in our sandbox study. For model calibration results in Fig. S10, comparing Case 2 to Case 1f, average  $L_1$  and  $L_2$  values increase by 24.4% and by 92.9%, respectively. For model validation results in the Fig. 7, comparing Case 2 to Case 1f, average  $L_1$  and  $L_2$  values increase by 19.8% and by 41.8%, respectively. The large differences in  $L_1$  and  $L_2$  values for model calibration can be attributed to our data sampling strategy, in which more early time data are used than data from the intermediate and late time. Examinations of  $L_1$  and  $L_2$  norms (in Fig. 7) of observed versus simulated drawdowns for each of the 16 pumping tests showed that, (1) Case 1d performed the best and Case 1f based on THT came a close second; (2) Case 2 performed worse than Cases 1d and 1f which treated both  $K$  and  $S_s$  as heterogeneous in terms of  $L_1$  and  $L_2$  norm ranks; (3) Case 2 performed even worse than Case 1b which used  $K$  field from SSHT and  $S_{eff}$  value from the effective parameter modeling approach.

These results collectively suggested that it is important to consider  $S_s$  heterogeneity for groundwater flow models when transient drawdown data are inverted to recover the subsurface heterogeneity even for a mildly heterogeneous aquifer. Otherwise, utilizing steady state data based on SSHT analysis instead of ignoring  $S_s$  heterogeneity for THT analysis could estimate  $K$  distribution more reasonably and quickly as well as predicting pumping tests more accurately.

Comparing results of Case 1b to Case 1d, and Case 1f to Case 2 (in Fig. 7), we can see that considering the  $S_s$  heterogeneity for the sandbox aquifer model generally leads to improved model validation performances. When the  $K$  field from SSHT was used, treating  $S_s$  values as a geologically distributed field instead of using an uniform  $S_{eff}$  value yielded decreases in average  $L_1$  and  $L_2$  values by 15.8% and 30.4%, respectively. For THT inversion cases, after considering  $S_s$  heterogeneity, the average  $L_1$  and  $L_2$  values for 16 pumping tests decreased by 16.5% and 29.5%, respectively. We should also note that  $L_1$  and  $L_2$  values for some tests (e.g., port 8, port 16) increased slightly when comparing Case 1d to Case 1b, suggesting the need for further improvement of the heterogeneous  $S_s$  field (Fig. 2b).

Overall, our laboratory sandbox study indicated that  $S_s$  heterogeneity should be considered in addition to  $K$  for the THT analysis even when the estimated  $S_s$  field is smooth. However, for a mildly heterogeneous aquifer, ignoring the spatial heterogeneity of  $S_s$  would not significantly bias the drawdown predictions when the  $K$  field estimated from SSHT is used for groundwater flow modeling purposes, as clearly shown in Figs. 3, 4, and 7.

L1	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	Average	Rank
Case 1a: Homo	0.222	0.113	0.125	0.223	0.175	0.142	0.282	0.193	0.264	0.268	0.227	0.302	0.250	0.280	0.240	0.179	0.218	6
Case 1b: SSHT+ $S_s$ _eff	0.054	0.043	0.102	0.073	0.086	0.068	0.102	0.077	0.134	0.082	0.119	0.193	0.146	0.163	0.102	0.108	0.103	3
Case 1c: SSHT+ $S_s$ _sgh	0.132	0.221	0.191	0.278	0.310	0.237	0.474	0.370	0.532	0.490	0.419	0.644	0.497	0.605	0.538	0.345	0.393	7
Case 1d: SSHT+ $S_s$ _geo	0.059	0.043	0.092	0.080	0.091	0.070	0.095	0.074	0.124	0.062	0.081	0.124	0.131	0.098	0.058	0.111	0.087	1
Case 1e: GeoModel	0.063	0.076	0.065	0.116	0.117	0.068	0.145	0.099	0.145	0.102	0.144	0.125	0.163	0.138	0.106	0.107	0.111	4
Case 1f: THT	0.060	0.047	0.099	0.093	0.096	0.094	0.115	0.103	0.124	0.072	0.100	0.098	0.150	0.110	0.077	0.098	0.096	2
Case 2: THT (Fixed $S_s$ )	0.058	0.055	0.119	0.096	0.092	0.117	0.125	0.126	0.147	0.103	0.119	0.134	0.143	0.158	0.113	0.132	0.115	5
L2	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	Average	Rank
Case 1a: Homo	0.116	0.045	0.043	0.143	0.069	0.042	0.127	0.075	0.107	0.132	0.097	0.172	0.123	0.149	0.109	0.070	0.101	6
Case 1b: SSHT+ $S_s$ _eff	0.007	0.004	0.047	0.013	0.021	0.015	0.020	0.014	0.036	0.015	0.039	0.096	0.048	0.059	0.027	0.040	0.031	3
Case 1c: SSHT+ $S_s$ _sgh	0.038	0.093	0.083	0.131	0.187	0.114	0.420	0.262	0.524	0.455	0.415	0.864	0.528	0.734	0.580	0.298	0.358	7
Case 1d: SSHT+ $S_s$ _geo	0.009	0.004	0.038	0.015	0.025	0.014	0.017	0.012	0.030	0.008	0.018	0.036	0.043	0.019	0.008	0.052	0.022	1
Case 1e: GeoModel	0.011	0.013	0.015	0.032	0.044	0.010	0.040	0.020	0.043	0.023	0.122	0.031	0.092	0.051	0.026	0.047	0.039	5
Case 1f: THT	0.013	0.004	0.049	0.022	0.025	0.031	0.025	0.023	0.030	0.011	0.032	0.025	0.055	0.025	0.016	0.031	0.026	2
Case 2: THT (Fixed $S_s$ )	0.010	0.005	0.067	0.022	0.023	0.053	0.032	0.035	0.044	0.025	0.033	0.042	0.051	0.054	0.032	0.061	0.037	4

**Fig. 7.**  $L_1$  and  $L_2$  norms of observed versus simulated drawdowns for 16 validation tests.

## 4. Synthetic experiments

### 4.1. Experimental design

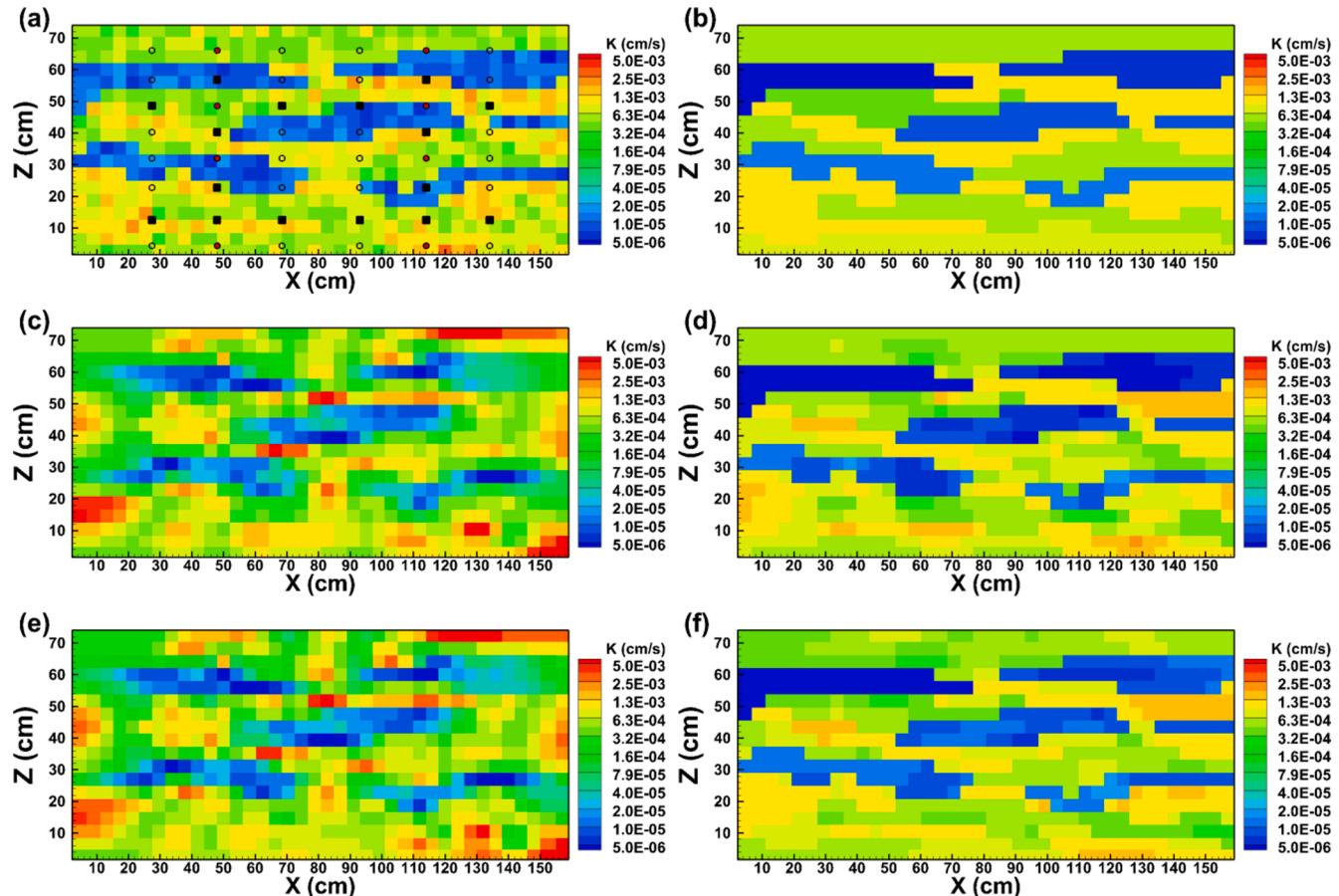
In the previous sections, the true fields of  $K$  and  $S_s$  of the sandbox aquifer were unknown and it is not possible to directly evaluate the accuracy of  $K$  and  $S_s$  estimates from different approaches. Additionally, the sandbox in Fig. 1 was built with sand of various sizes, lacking of less permeable clay layers which were frequently encountered during field investigations. Therefore, a synthetic numerical model, which contained both high and low permeable materials, was constructed to systematically investigate the issue of obtaining  $S_s$  estimates at finer resolutions for geostatistics-based THT analyses.

The synthetic aquifer had a groundwater flow model setting identical to the sandbox aquifer as described in the Section 2, except that sand types,  $K$  and  $S_s$  values for each of 18 layers were assumed to be known and the degree of heterogeneity to be much higher than the actual sandbox. The  $K$  values for the synthetic aquifer, as shown in Fig. 8a, were generated through: (1) constructing a layered mean  $\ln K$  field using values listed in the Table S3; (2) generating a set of normally distributed random numbers with a mean of zero and a variance of 0.10, representing variations of  $\ln K$ ; and (3) variations of  $\ln K$  were added to the mean  $\ln K$  field and then back-transformed, yielding a  $K$  field with both interlayer and intralayer variability. The  $S_s$  field was created following the same procedure and plotted as Fig. 9a. The mean values of  $\ln K$  and  $\ln S_s$  listed in Table S3 were based on  $K$  estimates for sediments typical at the NCRS site on the University of Waterloo campus from Alexander et al. (2011) and on ranges of  $S_s$  values suggested for porous media in Batu (1998). Through this approach, we built a synthetic numerical

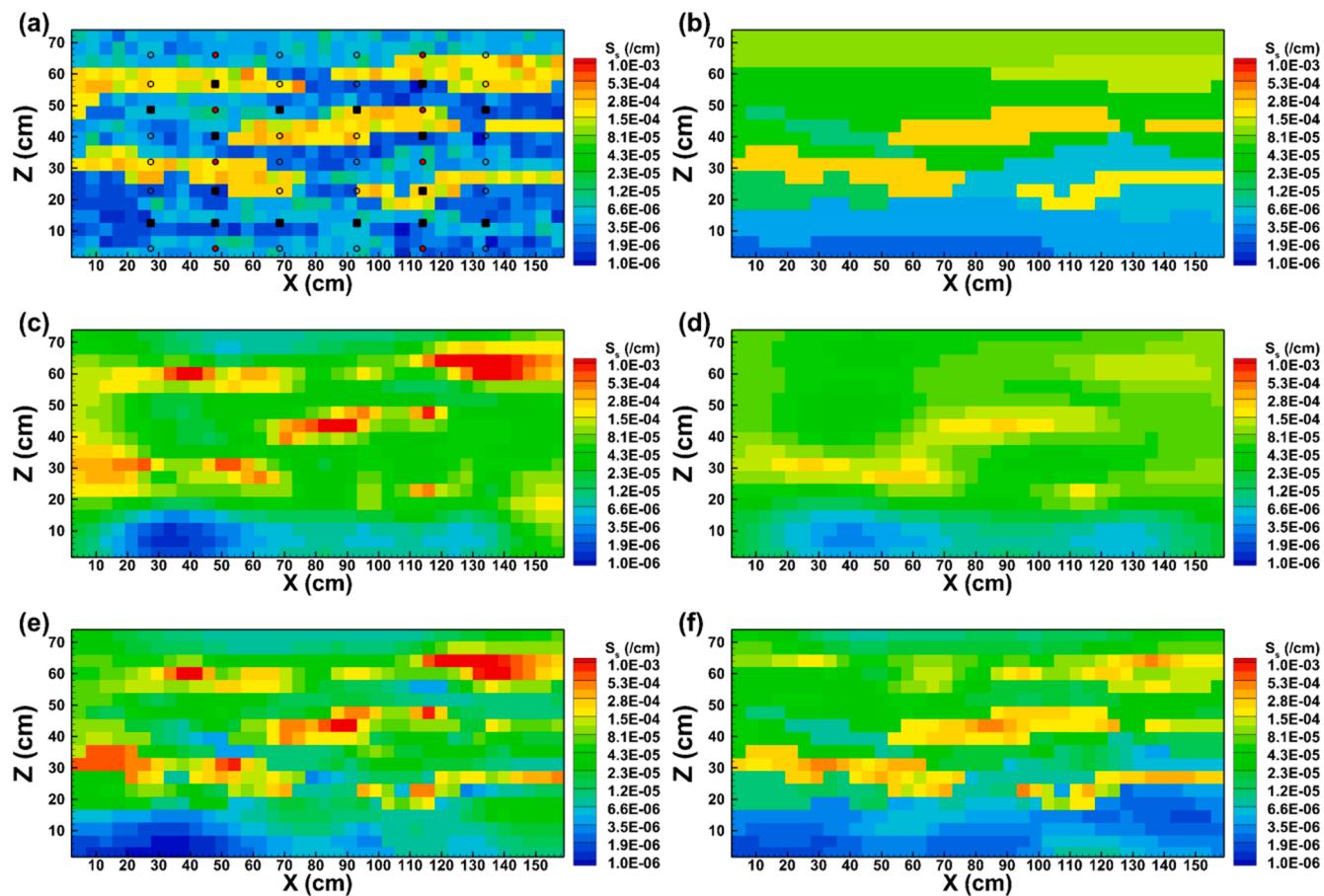
aquifer with geological information as in the laboratory sandbox constructed by Illman et al. (2010), while the hydraulic parameters for each of 18 layers similar to the values estimated for the highly heterogeneous NCRS (Berg and Illman, 2011b).

After generating the true fields for  $K$  and  $S_s$ , pumping tests were simulated at eight ports (ports 2, 5, 14, 17, 32, 35, 44, and 47) with a constant withdrawal rate of 10 ml/min for 200 min to ensure that steady state conditions were reached. During each pumping test, head data were collected at the other 47 observation ports not used for pumping. These pumping and observation locations were the same as Fig. 1a. Five data points were extracted from each drawdown curve for model calibration and the same data sampling strategy as described in Section 2.2 was utilized. In total, 1880 data points were selected from eight pumping tests to calibrate the effective parameter model, the geological model and the geostatistical inversion model. The effective parameter model was calibrated to yield effective hydraulic parameters ( $K_{eff}$  and  $S_{eff}$ ) that were used as the initial guesses of hydraulic parameters for other models. For the geostatistical inversion model, the other prior information ( $\sigma^2_{lnK}$ ,  $\sigma^2_{lnS_s}$ ,  $\lambda_x$ ,  $\lambda_y$ , and  $\lambda_z$ ) were assumed to be the same as those values given in the Section 3.1.3.

To meet our study objectives, five inverse modeling scenarios (Cases 3a, 3b, 3c, 3d, and 3e) were designed and their performances in recovering the true  $K$  and  $S_s$  fields were compared. In Case 3a, we assumed that the stratification of the geological model was perfectly known for the entire simulation domain. In Case 3b, the geostatistical inversion started with homogeneous mean fields of  $K_{eff}$  and  $S_{eff}$ . In Case 3c, the estimated  $K$  values from model calibration of Case 3a and the  $S_{eff}$  values from Case 3b were used as initial parameter fields. In Case 3d, the estimated  $S_s$  values from model calibration of Case 3a and the



**Fig. 8.** True and estimated  $K$  distributions in synthetic Case 3: (a) True  $K$ ; (b) Case 3a: calibrated geological model; (c) Case 3b: THT starting with uniform  $K$  and  $S_s$  mean fields; (d) Case 3c: THT starting with a heterogeneous  $K$  mean field from Case 3a and an uniform  $S_s$  mean field; (e) Case 3d: THT starting with an uniform  $K$  mean field and a heterogeneous  $S_s$  mean field from Case 3a; and (f) Case 3e: THT starting with heterogeneous  $K$  and  $S_s$  mean fields from Case 3a.



**Fig. 9.** True and estimated  $S_s$  distributions in synthetic Case 3: (a) True  $S_s$ ; (b) Case 3a: calibrated geological model; (c) Case 3b: THT starting with uniform  $K$  and  $S_s$  mean fields; (d) Case 3c: THT starting with a heterogeneous  $K$  mean field from Case 3a and an uniform  $S_s$  mean field; (e) Case 3d: THT starting with an uniform  $K$  mean field and a heterogeneous  $S_s$  mean field from Case 3a; and (f) Case 3e: THT starting with heterogeneous  $K$  and  $S_s$  mean fields from Case 3a.

$K_{eff}'$  values from Case 3b were used as the initial parameter fields. In Case 3e, both  $K$  and  $S_s$  values from Case 3a were used as the initial mean fields. The usefulness of initial  $K$  and/or  $S_s$  distributions for geostatistical inversion approach can be evaluated through comparing results from Cases 3b, 3c, 3d, and 3e.

#### 4.2. Results of the synthetic experiments

To facilitate visual comparisons,  $K$  and  $S_s$  estimates for Cases 3a, 3b, 3c, 3d, and 3e were plotted in Figs. 8b-f and 9b-f, respectively. Meanwhile, the estimated hydraulic parameters were quantitatively compared to true fields in Figs. 10a-e and 11a-e, respectively.

Comparing Fig. 8b with Fig. 8a, the calibrated geological model of Case 3a revealed the heterogeneity pattern of  $K$  quite accurately, which could be attributed to the facts that accurate knowledge of stratigraphy information was assumed to be available and a large amount of pumping test data were used for model calibration. The linear fitting results in Fig. 10a also showed that  $R^2$  reached 0.95, suggesting a very good match of estimated  $K$  with the true values. On the other hand, the estimated  $S_s$  field in Fig. 9b captured the main heterogeneity pattern in Fig. 9a, while losing details in the two high  $S_s$  value zones near the upper boundary and overestimating the low value areas (Fig. 11a). These results emphasize that  $S_s$  heterogeneity is more difficult to be accurately estimated than  $K$  (Wu et al., 2005; Liu et al., 2007; Zha et al., 2019).

For the geostatistical inversion Case 3b, we see that Fig. 8c showed the five low  $K$  zones located in the central part, with some details missing from areas around the left and right boundaries. The estimated  $S_s$  tomogram in Fig. 9c showed the featured high value zones in Fig. 9a. However, the low values were also overestimated in Fig. 9c. For Case 3c,

after using  $K$  values of the calibrated geological model (Fig. 8b) as the initial mean  $K$  field for the geostatistical inversion approach, the estimated  $K$  tomogram in Fig. 8d showed significant improvements over Fig. 8c of the Case 3b. However, the estimated  $S_s$  field of the Case 3c showed less details than Figs. 9b (Case 3a) and c (Case 3b), with worse results in terms of the linear model fit in Fig. 11c than those in Figs. 11a and b.

When using  $S_s$  values in Fig. 9b as the initial mean  $S_s$  field for Case 3d, the geostatistical inversion approach yielded the best  $S_s$  estimates among Cases 3a, 3b, 3c, and 3d, with a higher  $R^2$  value of 0.54 in Fig. 11d than those in Figs. 11a ( $R^2 = 0.39$ ), b ( $R^2 = 0.33$ ), and c ( $R^2 = 0.22$ ). However, the  $K$  estimates of Case 3c (Fig. 10d) were generally the same as those of Case 3b (Fig. 10b). For Case 3e, both  $K$  and  $S_s$  fields (Figs. 8f and 9f) were improved compared to those of Case 3b, with increased  $R^2$  values and lowered  $L_1$  and  $L_2$  norms when comparing Figs. 10e and 11e to Figs. 10b and 11b. These results suggested that, using only  $K$  or  $S_s$  values from the calibrated geological model as the initial mean field for the geostatistical inversion approach did not necessarily benefit the accurate estimation of the other parameter ( $S_s$  or  $K$ ) field. The summarized statistical parameters of the true case and Cases 3a-3e (Tables S4 and S5) also suggested similar findings that, (1) geometric means of  $S_s$  fields were generally overestimated in Cases 3a-3e; (2) using geologically distributed  $K$  or  $S_s$  values for the geostatistical inversion approach was helpful for the estimation of corresponding  $K$  or  $S_s$  tomograms. Therefore, in order to capture both interlayer and intralayer heterogeneity of  $K$  and  $S_s$  values for the THT analysis, reliable prior information have to be included.

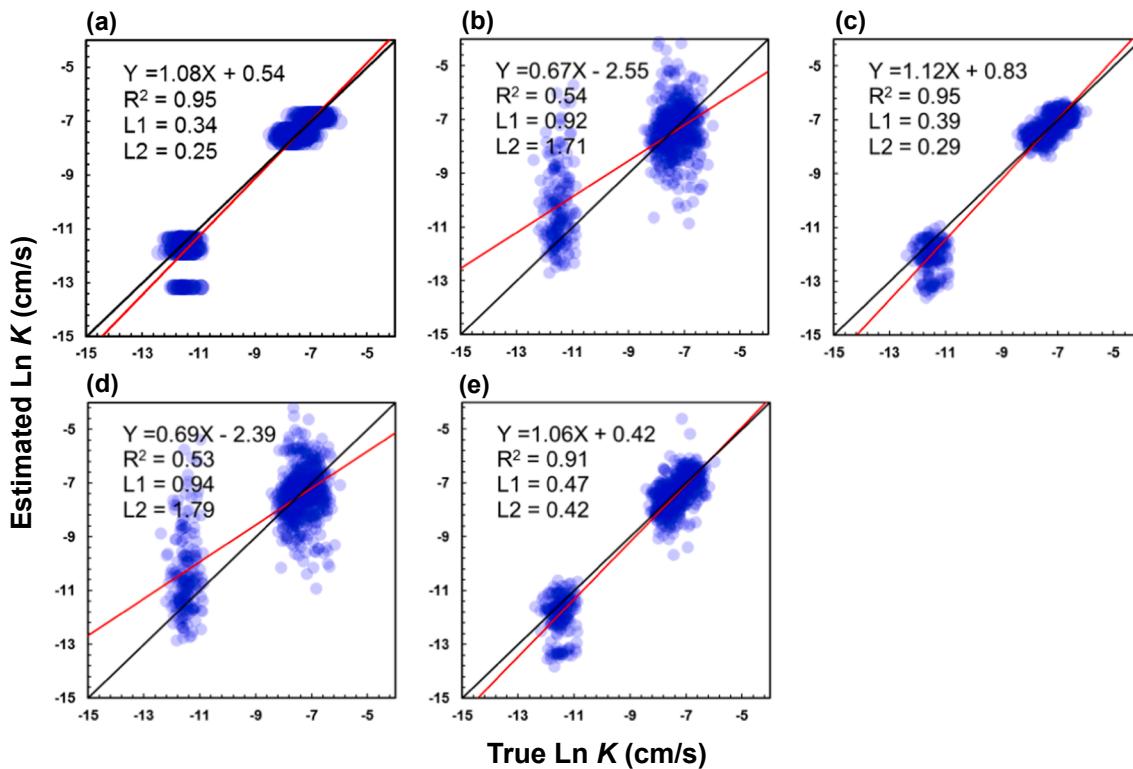


Fig. 10. Scatterplots of the true versus estimated  $\ln K$  values for synthetic Case 3.

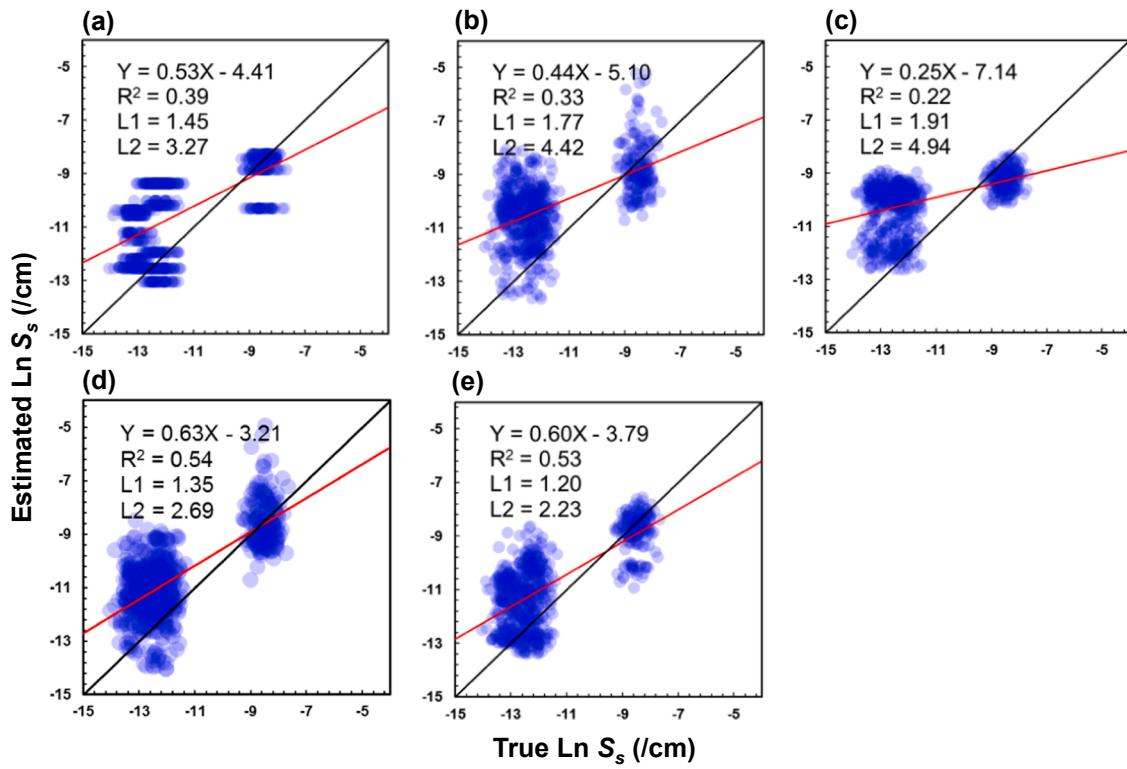


Fig. 11. Scatterplots of the true versus estimated  $\ln S_s$  values for synthetic Case 3.

## 5. Summary and conclusions

In this study, we investigated whether the spatial heterogeneity of  $S_s$  should be considered during the interpretation of transient hydraulic

tomography (THT) data to reliably characterize the heterogeneity of a laboratory sandbox aquifer. We selected three modeling approaches of different complexities and parameterization schemes to map the  $K$  and  $S_s$  heterogeneity of the laboratory sandbox aquifer at varying

resolutions, including: (1) the effective parameter model; (2) the “perfect” geological model; and (3) the geostatistics-based inversion models. In addition, we examined the value of geologically-distributed  $K$  and  $S_s$  values for the THT analysis by using them as prior distributions for the geostatistical inversion approach, with a synthetic case study. Our work resulted in the following major findings and conclusions:

1. The estimated  $K$  values from the calibration of the “perfect” geological model corresponded well with the grain size information of the laboratory sandbox aquifer. When the inversion started with homogeneous parameter estimates, the geostatistics-based THT and SSHT analyses yielded  $K$  tomograms with high and low  $K$  zones that resembled the geological model, but lacked in consistency in terms of layer boundaries. On the other hand, with the inclusion of increased number of early time response data (at and around the time “ $t_0$ ” as suggested by Sun et al. (2013) into the calibration procedure, the geological model revealed clearer structural features for the  $S_s$  field of the laboratory aquifer, while the  $S_s$  tomogram obtained from THT analysis was still smooth and revealed no distinctive geological features. Therefore, to reconstruct the  $S_s$  field more reliably, the combined use of both geological information and pumping tests data seems more favourable for THT analyses.
2. The effective parameter approach that treats both  $K$  and  $S_s$  fields as homogeneous poorly predicted the drawdown curves at early, intermediate, and late times for 16 independent pumping tests used for model validation. When the high resolution  $K$  field estimated from SSHT was used for the groundwater flow model, ignoring the spatial heterogeneity of  $S_s$  could still predict late time drawdowns accurately. Further conceptualizing the  $S_s$  values as a geologically distributed field obtained from model calibration was reliable and led to improved drawdown predictions. At sites with complex geological structures, groundwater models based on the composite media concept (Winter and Tartakovsky, 2000; Clark et al., 2020) or the general statistical framework suggested by Zha et al. (2017), that account for both boundary geometry variability and intralayer heterogeneity, should be more routinely adopted for estimating both  $K$  and  $S_s$  fields.  $S_s$  values estimated from single-hole tests should be used with caution in groundwater flow models, since they could lead to significantly biased drawdown predictions.
3. Although  $S_s$  heterogeneity in this laboratory sandbox aquifer was milder in comparison to that of  $K$  and the estimated  $S_s$  tomogram from THT was found to be smooth,  $S_s$  heterogeneity should still be considered in order to accurately estimate  $K$  values from the inverse analysis of transient pumping test data. This agrees with the conclusions by Castagna et al. (2011). The variance of  $S_s$  field for the sandbox is smaller ( $\sigma^2_{lnS_s} = 0.6\text{--}0.8$ ) than that of the synthetic experiment ( $\sigma^2_{lnS_s} = 3.3$ ), while in the former, we evaluated the reliability of  $S_s$  estimates through validation pumping tests, and in the latter, we directly compared the estimates to the true  $S_s$  values. Results from both experiments suggested that  $S_s$  may be reliably estimated or recovered through model calibration efforts if proper parameterization schemes and modeling approaches were used.
4. Finally, we conclude that the inclusion of reliable geological information of  $K$  and  $S_s$  are helpful for THT analyses to better characterize the inherent heterogeneous hydraulic parameter fields at fine resolutions. Meanwhile, in order to map the  $S_s$  field reliably, it is important to use sufficient non-redundant information regarding aquifer heterogeneity and a reasonable data sampling strategy for THT analyses (Sun et al., 2013; Wen et al., 2020). In this regard, methods involving analyses of slug test data, sea tides, atmospheric loading and earth tides, and seismic waves, etc., may be jointly used for future research to provide valuable in-situ estimates of  $S_s$  (Kuang et al., 2020). Furthermore, we believe the results of this sandbox study could help to guide the interpretation of THT under field conditions in order to help map both  $K$  and  $S_s$  distributions more accurately.

## CRediT authorship contribution statement

Zhanfeng Zhao: Conceptualization, Methodology, Writing - original draft, Writing - review & editing. Walter A. Illman: Conceptualization, Methodology, Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research was supported by the Discovery Grant awarded to Walter A. Illman by the Natural Sciences and Engineering Research Council of Canada (NSERC), as well as grants from the Ontario Research Foundation (ORF) and Canada Foundation for Innovation (CFI). The first author acknowledges the support of the National Natural Science Foundation of China (Grant Number: 41807202) and the Beijing Municipal Natural Science Foundation (Grant Number: 8162043). We thank the Associate Editor (Okke Batelaan) and two anonymous reviewers for their helpful comments in improving the manuscript.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhydrol.2020.125874>.

## References

- Acworth, R.I., Rau, G.C., Halloran, L.J.S., Timms, W.A., 2017. Vertical groundwater storage properties and changes in confinement determined using hydraulic head response to atmospheric tides. *Water Resour. Res.* 53 (4), 2983–2997. <https://doi.org/10.1002/2016WR020311>.
- Alexander, M., Berg, S.J., Illman, W.A., 2011. Field study of hydrogeologic characterization methods in a heterogeneous aquifer. *Ground Water* 49 (3), 365–382. doi:10.1111/j.1745-6584.2010.00729.x.
- Aristodemo, F., Ianchello, M., Falllico, C., 2018. Smoothing analysis of slug tests data for aquifer characterization at laboratory scale. *J. Hydrol.* 562, 125–139. <https://doi.org/10.1016/j.jhydrol.2018.04.056>.
- Batu, V., 1998. *Aquifer Hydraulics: A Comprehensive Guide to Hydrogeologic Data Analysis*. John Wiley & Sons.
- Béjar-Pizarro, M., Ezquerro, P., Herrera, G., Tomás, R., Guardiola-Albert, C., Ruiz Hernández, J.M., Fernández Merodo, J.A., Marchamalo, M., Martínez, R., 2017. Mapping groundwater level and aquifer storage variations from InSAR measurements in the Madrid aquifer, Central Spain. *J. Hydrol.* 547, 678–689. <https://doi.org/10.1016/j.jhydrol.2017.02.011>.
- Berg, S.J., Illman, W.A., 2011a. Capturing aquifer heterogeneity: comparison of approaches through controlled sandbox experiments. *Water Resour. Res.* 47 (9), 1–17. <https://doi.org/10.1029/2011WR010429>.
- Berg, S.J., Illman, W.A., 2011b. Three-dimensional transient hydraulic tomography in a highly heterogeneous glaciogenic aquifer-aquitard system. *Water Resour. Res.* 47 (10) <https://doi.org/10.1029/2011WR010616>.
- Bianchi, M., Zheng, C., 2016. A lithofacies approach for modeling non-Fickian solute transport in a heterogeneous alluvial aquifer. *Water Resour. Res.* 52 (1), 552–565. <https://doi.org/10.1002/2015WR018186>.
- Bohling, G.C., Butler, J.J., Zhan, X., Knoll, M.D., 2007. A field assessment of the value of steady shape hydraulic tomography for characterization of aquifer heterogeneities. *Water Resour. Res.* 43 (5), 1–23. <https://doi.org/10.1029/2006WR004932>.
- Brauchler, R., Hu, R., Dietrich, P., Sauter, M., 2011. A field assessment of high-resolution aquifer characterization based on hydraulic travel time and hydraulic attenuation tomography. *Water Resour. Res.* 47 (3), 1–12. <https://doi.org/10.1029/2010WR009635>.
- Brauchler, R., Hu, R., Hu, L., Jiménez, S., Bayer, P., Dietrich, P., Ptak, T., 2013. Rapid field application of hydraulic tomography for resolving aquifer heterogeneity in unconsolidated sediments. *Water Resour. Res.* 49 (4), 2013–2024. <https://doi.org/10.1002/wrcr.20181>.
- Cardiff, M., Barrash, W., 2011. 3-D transient hydraulic tomography in unconfined aquifers with fast drainage response. *Water Resour. Res.* 47 (12) <https://doi.org/10.1029/2010WR010367>.
- Cardiff, M., Barrash, W., Kitanidis, P.K., 2012. A field proof-of-concept of aquifer imaging using 3-D transient hydraulic tomography with modular, temporarily-emplaced equipment. *Water Resour. Res.* 48 (5) <https://doi.org/10.1029/2011WR011704>.
- Cardiff, M., Barrash, W., Kitanidis, P.K., 2013. Hydraulic conductivity imaging from 3-D transient hydraulic tomography at several pumping/observation densities: Hydraulic

- Tomography Imaging Resolution. *Water Resour. Res.* 49 (11), 7311–7326. <https://doi.org/10.1002/wrcr.20519>.
- Castagna, M., Becker, M.W., Bellin, A., 2011. Joint estimation of transmissivity and storativity in a bedrock fracture. *Water Resour. Res.* 47 (9), 1–19. <https://doi.org/10.1029/2010WR009262>.
- Clark, C.L., Winter, C.L., Corley, T., 2020. Effects of percolation on the effective conductivity of irregular composite porous media. *Adv. Water Resour.* 137, 103507. <https://doi.org/10.1016/j.advwatres.2020.103507>.
- Cooper Jr., H.H., Jacob, C.E., 1946. A generalized graphical method for evaluating formation constants and summarizing well-field history. *Trans. AGU* 27 (4), 526. <https://doi.org/10.1029/TR027i004p00526>.
- Craig, A.J., 2005. Measurement of hydraulic parameters at multiple scales in two synthetic heterogeneous aquifers constructed in the laboratory. *Dept. of Civ. and Environ. Eng., Univ. of Iowa, Iowa City.*
- Dagan, G., 1982. Analysis of flow through heterogeneous random aquifers: 2. Unsteady flow in confined formations. *Water Resour. Res.* 18 (5), 1571–1585. <https://doi.org/10.1029/WR018i005p01571>.
- Dagan, G. (Ed.), 1989. *Flow and Transport in Porous Formations*. Springer-Verlag Berlin Heidelberg.
- Dagan, G., Rubin, Y., 1988. Stochastic identification of recharge, transmissivity, and storativity in aquifer transient flow: a quasi-steady approach. *Water Resour. Res.* 24 (10), 1698–1710. <https://doi.org/10.1029/WR024i010p01698>.
- Dietrich, P., Butler Jr., J.J., Faibis, K., 2008. A rapid method for hydraulic profiling in unconsolidated formations. *Groundwater* 46 (2), 323–328. <https://doi.org/10.1111/j.1745-6584.2007.00377.x>.
- Doherty, J., 2005. *PEST: Model-Independent Parameter Estimation User Manual*, fifth ed. Watermark Numer. Comput., Brisbane, Australia.
- Domenico, P.A., Mifflin, M.D., 1965. Water from low-permeability sediments and land subsidence. *Water Resour. Res.* 1 (4), 563–576. <https://doi.org/10.1029/WR001i004p00563>.
- Don, N.C., Araki, H., Yamanishi, H., Koga, K., 2005. Simulation of groundwater flow and environmental effects resulting from pumping. *Environ. Geol.* 47 (3), 361–374. <https://doi.org/10.1007/s00254-004-1158-1>.
- Elfeki, A.M.M., Uffink, G., Lebreton, S., 2012. Influence of temporal fluctuations and spatial heterogeneity on pollution transport in porous media. *Hydrogeol. J.* 20 (2), 283–297. <https://doi.org/10.1007/s10040-011-0796-0>.
- Erban, L.E., Gorelick, S.M., Zebker, H.A., Fendorf, S., 2013. Release of arsenic to deep groundwater in the Mekong Delta, Vietnam, linked to pumping-induced land subsidence. *Proc. Natl. Acad. Sci. USA* 110 (34), 13751–13756. <https://doi.org/10.1073/pnas.130053110>.
- Erban, L.E., Gorelick, S.M., Zebker, H.A., 2014. Groundwater extraction, land subsidence, and sea-level rise in the Mekong Delta, Vietnam. *Environ. Res. Lett.* 9 (8), 084010. <https://doi.org/10.1088/1748-9326/9/8/084010>.
- Fischer, P., Jardani, A., Jourde, H., 2020. Hydraulic tomography in coupled discrete-continuum concept to image hydraulic properties of a fractured and karstified aquifer (Lez aquifer, France). *Adv. Water Resour.* 137, 103523. <https://doi.org/10.1016/j.advwatres.2020.103523>.
- Freeze, R.A., 1975. A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resour. Res.* 11 (5), 725–741. <https://doi.org/10.1029/WR011i005p00725>.
- Garhwal, L.W., 1993. *Stochastic Subsurface Hydrology*. Prentice-Hall.
- Goode, D.J., Konikow, L.F., 1990. Apparent dispersion in transient groundwater flow. *Water Resour. Res.* 26 (10), 2339–2351. <https://doi.org/10.1029/WR026i010p02339>.
- Gottlieb, J., Dietrich, P., 1995. Identification of the permeability distribution in soil by hydraulic tomography. *Inverse Problems* 11 (2), 353–360. <https://doi.org/10.1088/0266-5611/11/2/005>.
- Hendricks Franssen, H.J., Alcolea, A., Riva, M., Bakr, M., van der Wiel, N., Stauffer, F., Guadagnini, A., 2009. A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments. *Adv. Water Resour.* 32 (6), 851–872. <https://doi.org/10.1016/j.advwatres.2009.02.011>.
- Hochstetler, D.L., Barrash, W., Leven, C., Cardiff, M., Chidichimo, F., Kitanidis, P.K., 2016. Hydraulic tomography: Continuity and discontinuity of high-K and low-K zones. *Groundwater* 54 (2), 171–185. <https://doi.org/10.1111/gwat.12344>.
- Hoeksema, R.J., Kitanidis, P.K., 1985. Analysis of the spatial structure of properties of selected aquifers. *Water Resour. Res.* 21 (4), 563–572. <https://doi.org/10.1029/WR021i004p00563>.
- Hung, W.-C., Hwang, C., Liou, J.-C., Lin, Y.-S., Yang, H.-L., 2012. Modeling aquifer-system compaction and predicting land subsidence in central Taiwan. *Eng. Geol.* 147–148, 78–90. <https://doi.org/10.1016/j.enggeo.2012.07.018>.
- Illman, W.A., Tartakovsky, D.M., 2006. Asymptotic analysis of cross-hole hydraulic tests in fractured granite. *Groundwater* 44 (4), 555–563. doi:10.1111/j.1745-6584.2006.00201.x.
- Illman, W.A., Zhu, J., Craig, A.J., Yin, D., 2010. Comparison of aquifer characterization approaches through steady state groundwater model validation: a controlled laboratory sandbox study. *Water Resour. Res.* 46 (4), 1–18. <https://doi.org/10.1029/2009WR007745>.
- Illman, W.A., Berg, S.J., Zhao, Z., 2015. Should hydraulic tomography data be interpreted using geostatistical inverse modeling? A laboratory sandbox investigation: geostatistics for hydraulic tomography? *Water Resour. Res.* 51 (5), 3219–3237. <https://doi.org/10.1002/2014WR016552>.
- Inouchi, K., Kishi, Y., Kakinuma, T., 1990. The motion of coastal groundwater approximate analytical expressions. *J. Hydrol.* 115, 165–191.
- Jiménez, S., Brauchler, R., Hu, R., Hu, L., Schmidt, S., Ptak, T., Bayer, P., 2015. Prediction of solute transport in a heterogeneous aquifer utilizing hydraulic conductivity and specific storage tomograms. *Water Resour. Res.* 51 (7), 5504–5520. <https://doi.org/10.1002/2014WR016402>.
- Keller, C.K., Van Der Kamp, G., Cherry, J.A., 1989. A multiscale study of the permeability of a thick clayey till. *Water Resour. Res.* 25 (11), 2299–2317. <https://doi.org/10.1029/WR025i011p02299>.
- Klepikova, M.V., Le Borgne, T., Bour, O., de Dreuzy, J.-R., 2013. Inverse modeling of flow tomography experiments in fractured media: flow tomography in fractured media. *Water Resour. Res.* 49 (11), 7255–7265. <https://doi.org/10.1002/2013WR013722>.
- Koltermann, C.E., Gorelick, S.M., 1996. Heterogeneity in sedimentary deposits: a review of structure-imitating, process-imitating, and descriptive approaches. *Water Resour. Res.* 32 (9), 2617–2658. <https://doi.org/10.1029/96WR00025>.
- Konikow, L.F., Neuzil, C.E., 2007. A method to estimate groundwater depletion from confining layers. *Water Resour. Res.* 43 (7) <https://doi.org/10.1029/2006WR005597>.
- Konikow, L.F., 2011. Contribution of global groundwater depletion since 1900 to sea-level rise. *Geophys. Res. Lett.* 38 (17), n/a–n/a. <https://doi.org/10.1029/2011GL048604>.
- Kuang, X., Jiao, J.J., Zheng, C., Cherry, J.A., Li, H., 2020. A review of specific storage in aquifers. *J. Hydrol.* 581, 124383. <https://doi.org/10.1016/j.jhydrol.2019.124383>.
- Leven, C., Dietrich, P., 2006. What information can we get from pumping tests? comparing pumping test configurations using sensitivity coefficients. *J. Hydrol.* 319 (1–4), 199–215. <https://doi.org/10.1016/j.jhydrol.2005.06.030>.
- Li, W., Nowak, W., Cirpka, O.A., 2005. Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown. *Water Resour. Res.* 41 (8), 1–13. <https://doi.org/10.1029/2004WR003874>.
- Linde, N., Renard, P., Mukerji, T., Caers, J., 2015. Geological realism in hydrogeological and geophysical inverse modeling: a review. *Adv. Water Resour.* 86, 86–101. <https://doi.org/10.1016/j.advwatres.2015.09.019>.
- Liu, G., Butler Jr., J.J., Bohling, G.C., Reboulet, E., Knobbe, S., Hyndman, D.W., 2009. A new method for high-resolution characterization of hydraulic conductivity. *Water Resour. Res.* 45 (8) <https://doi.org/10.1029/2009WR008319>.
- Liu, S., Yeh, T.-C., Gardiner, R., 2002. Effectiveness of hydraulic tomography: sandbox experiments. *Water Resour. Res.* 38 (4), 5–1–5–9. <https://doi.org/10.1029/2001WR000338>.
- Liu, X., Illman, W.A., Craig, A.J., Zhu, J., Yeh, T.C.J., 2007. Laboratory sandbox validation of transient hydraulic tomography. *Water Resour. Res.* 43 (5), 1–13. <https://doi.org/10.1029/2006WR005144>.
- Lu, Z., Robinson, B.A., 2006. Parameter identification using the level set method. *Geophys. Res. Lett.* 33 (6), 1–4. <https://doi.org/10.1029/2005GL025541>.
- Luo, N., Zhao, Z., Illman, W.A., Berg, S.J., 2017. Comparative study of transient hydraulic tomography with varying parameterizations and zonations: laboratory sandbox investigation. *J. Hydrol.* 554, 758–779. <https://doi.org/10.1016/j.jhydrol.2017.09.045>.
- Mao, D., Yeh, T.-C., Wan, L.i., Lee, C.-H., Hsu, K.-C., Wen, J.-C., Lu, W., 2013a. Cross-correlation analysis and information content of observed heads during pumping in unconfined aquifers. *Water Resour. Res.* 49 (2), 713–731. <https://doi.org/10.1029/2006WR002066>.
- Mao, D., Yeh, T.-C., Wan, L.i., Hsu, K.-C., Lee, C.-H., Wen, J.-C., 2013b. Necessary conditions for inverse modeling of flow through variably saturated porous media. *Adv. Water Resour.* 52, 50–61. <https://doi.org/10.1016/j.advwatres.2012.08.001>.
- Mao, D., Liu, Z., Wang, W., Li, S., Gao, Y., Xu, Z., Zhang, C., 2018. An application of hydraulic tomography to a deep coal mine: Combining traditional pumping tests with water inrush incidents. *J. Hydrol.* 567, 1–11. <https://doi.org/10.1016/j.jhydrol.2018.09.058>.
- de Marsily, G.h., Delay, F., Gonçalvès, J., Renard, P.h., Teles, V., Violette, S., 2005. Dealing with spatial heterogeneity. *Hydrogeol. J.* 13 (1), 161–183. <https://doi.org/10.1007/s10040-004-0432-3>.
- Molz, F.J., Morin, R.H., Hess, A.E., Melville, J.G., Güven, O., 1989. The Impeller Meter for measuring aquifer permeability variations: evaluation and comparison with other tests. *Water Resour. Res.* 25 (7), 1677–1683. <https://doi.org/10.1029/WR025i007p01677>.
- Neuman, S.P., Witherspoon, P.A., 1972. Field determinations of the hydraulic properties of leaky multiple aquifer systems. *Water Resour. Res.* 8 (5), 1284–1298. <https://doi.org/10.1029/WR008i005p01284>.
- Ojha, C., Shirzaei, M., Werth, S., Argus, D.F., Farr, T.G., 2018. Sustained groundwater loss in California's Central Valley exacerbated by intense drought periods. *Water Resour. Res.* 54 (7), 4449–4460. <https://doi.org/10.1029/2017WR022250>.
- Panzeri, M., Riva, M., Guadagnini, A., Neuman, S.P., 2013. Data assimilation and parameter estimation via ensemble Kalman filter coupled with stochastic moment equations of transient groundwater flow. *Water Resour. Res.* 49 (3), 1334–1344. <https://doi.org/10.1029/WR022r031334>.
- Paradis, D., Gloaguen, E., Lefebvre, R., Giroux, B., 2016. A field proof-of-concept of tomographic slug tests in an anisotropic littoral aquifer. *J. Hydrol.* 536, 61–73. <https://doi.org/10.1016/j.jhydrol.2016.02.041>.
- Pool, M., Post, V.E.A., Simmons, C.T., 2014. Effects of tidal fluctuations on mixing and spreading in coastal aquifers: homogeneous case. *Water Resour. Res.* 50 (8), 6910–6926. <https://doi.org/10.1002/2014WR015534>.
- Pool, M., Dentz, M., Post, V.E.A., 2016. Transient forcing effects on mixing of two fluids for a stable stratification. *Water Resour. Res.* 52 (9), 7178–7197. <https://doi.org/10.1002/2016WR019181>.
- Quinn, P.M., Cherry, J.A., Parker, B.L., 2016. Depth-discrete specific storage in fractured sedimentary rock using steady-state and transient single-hole hydraulic tests. *J. Hydrol.* 542, 756–771. <https://doi.org/10.1016/j.jhydrol.2016.09.046>.
- Rau, G.C., Acworth, R.I., Halloran, L.J.S., Timms, W.A., Cuthbert, M.O., 2018. Quantifying compressible groundwater storage by combining cross-hole seismic

- surveys and head response to atmospheric tides. *J. Geophys. Res. Earth Surf.* 123 (8), 1910–1930. <https://doi.org/10.1029/2018JF004660>.
- Rehfeldt, K.R., Boggs, J.M., Gelhar, L.W., 1992. Field study of dispersion in a heterogeneous aquifer: 3. Geostatistical analysis of hydraulic conductivity. *Water Resour. Res.* 28 (12), 3309–3324. <https://doi.org/10.1029/92WR01758>.
- Rezaei, A., Mousavi, Z., Khorrami, F., Nankali, H., 2020. Inelastic and elastic storage properties and daily hydraulic head estimates from continuous global positioning system (GPS) measurements in northern Iran. *Hydrogeol. J.* 28 (2), 657–672. <https://doi.org/10.1007/s10040-019-02092-y>.
- Rodell, M., Chen, J., Kato, H., Famiglietti, J.S., Nigro, J., Wilson, C.R., 2007. Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeol. J.* 15 (1), 159–166. <https://doi.org/10.1007/s10040-006-0103-7>.
- Shaver, R.B., 1998. The determination of glacial till specific storage in North Dakota. *Ground Water* 36 (4), 552–557. <https://doi.org/10.1111/j.1745-6584.1998.tb02828.x>.
- Smith, L.A., van der Kamp, G., Jim Hendry, M., 2013. A new technique for obtaining high-resolution pore pressure records in thick claystone aquitards and its use to determine in situ compressibility. *Water Resour. Res.* 49 (2), 732–743. <https://doi.org/10.1029/wrcr.20084>.
- Smith, L.A., Barbour, S.L., Hendry, M.J., Novakowski, K., van der Kamp, G., 2016. A multiscale approach to determine hydraulic conductivity in thick claystone aquitards using field, laboratory, and numerical modeling methods. *Water Resour. Res.* 52 (7), 5265–5284. <https://doi.org/10.1029/2015WR018448>.
- Soueid Ahmed, A., Jardani, A., Revil, A., Dupont, J.P., 2016. Specific storage and hydraulic conductivity tomography through the joint inversion of hydraulic heads and self-potential data. *Adv. Water Resour.* 89, 80–90. <https://doi.org/10.1016/j.advwatres.2016.01.006>.
- Straface, S., Yeh, T.C.J., Zhu, J., Troisi, S., Lee, C.H., 2007. Sequential aquifer tests at a well field, Montalto Uffugo Scalo, Italy. *Water Resour. Res.* 43 (7), 1–13. <https://doi.org/10.1029/2006WR005287>.
- Sudicky, E.A., 1986. A natural gradient experiment on solute transport in a sand aquifer: spatial variability of hydraulic conductivity and its role in the dispersion process. *Water Resour. Res.* 22 (13), 2069–2082. <https://doi.org/10.1029/WR022i013p02069>.
- Sudicky, E.A., Illman, W.A., Goltz, I.K., Adams, J.J., McLaren, R.G., 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 (1) <https://doi.org/10.1029/2008WR007558>. W01508 (1–16).
- Sun, A.Y., Green, R., Rodell, M., Swenson, S., 2010. Inferring aquifer storage parameters using satellite and in situ measurements: Estimation under uncertainty. *n/a-n/a Geophys. Res. Lett.* 37 (10). <https://doi.org/10.1029/2010GL043231>.
- Sun, N.-Z., Yeh, W.-G., 1992. A stochastic inverse solution for transient groundwater flow: parameter identification and reliability analysis. *Water Resour. Res.* 28 (12), 3269–3280. <https://doi.org/10.1029/92WR00683>.
- Sun, R., Yeh, T.-C., Mao, D., Jin, M., Lu, W., Hao, Y., 2013. A temporal sampling strategy for hydraulic tomography analysis: a temporal sampling for transient hydraulic tomography. *Water Resour. Res.* 49 (7), 3881–3896. <https://doi.org/10.1002/wrcr.20337>.
- Tamayo-Mas, E., Bianchi, M., Mansour, M., 2018. Impact of model complexity and multi-scale data integration on the estimation of hydrogeological parameters in a dual-porosity aquifer. *Hydrogeol. J.* 26 (6), 1917–1933. <https://doi.org/10.1007/s10040-018-1745-y>.
- Tartakovsky M., D., Neuman P., S., 1998. Transient flow in bounded randomly heterogeneous domains: 1. Exact conditional moment equations and recursive approximations. *Water Resour. Res.* 34 (1), 1–12. <https://doi.org/10.1029/97WR02118>.
- Theis, C.V., 1935. The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using groundwater storage. *Eos Trans. Am. Geophys. Union* 16 (2), 519–524. <https://doi.org/10.1029/TR016i002p00519>.
- Tiedeman, C.R., Barrash, W., 2020. Hydraulic tomography: 3D hydraulic conductivity, fracture network, and connectivity in mudstone. *Groundwater* 58 (2), 238–257. <https://doi.org/10.1111/gwat.12915>.
- Tsai, J.-P., Yeh, T.-C., Cheng, C.-C., Zha, Y., Chang, L.-C., Hwang, C., Wang, Y.-L., Hao, Y., 2017. Fusion of time-lapse gravity survey and hydraulic tomography for estimating spatially varying hydraulic conductivity and specific yield fields. *Water Resour. Res.* 53 (10), 8554–8571. <https://doi.org/10.1002/2017WR020459>.
- Wen, J.-C., Chen, J.-L., Yeh, T.-C., Wang, Y.-L., Huang, S.-Y., Tian, Z., Yu, C.-Y., 2020. Redundant and nonredundant information for model calibration or hydraulic tomography. *Groundwater* 58 (1), 79–92. <https://doi.org/10.1111/gwat.12879>.
- Winter, C.L., Tartakovsky, D.M., 2000. Mean flow in composite porous media. *Geophys. Res. Lett.* 27 (12), 1759–1762. <https://doi.org/10.1029/1999GL011030>.
- Winter, C.L., Tartakovsky, D.M., 2002. Groundwater flow in heterogeneous composite aquifers. *Water Resour. Res.* 38 (8), 23–1–23–11. <https://doi.org/10.1029/2001WR000450>.
- Winter, C.L., Tartakovsky, D.M., Guadagnini, A., 2002. Numerical solutions of moment equations for flow in heterogeneous composite aquifers. *Water Resour. Res.* 38 (5), 13–1–13–8. <https://doi.org/10.1029/2001WR000222>.
- Wu, C.M., Yeh, T.C.J., Zhu, J., Tim, H.L., Hsu, N.S., Chen, C.H., Sancho, A.F., 2005. Traditional analysis of aquifer tests: comparing apples to oranges? *Water Resour. Res.* 41 (9), 1–12. <https://doi.org/10.1029/2004WR003717>.
- Xiang, J., Yeh, T.C.J., Lee, C.H., Hsu, K.C., Wen, J.C., 2009. A simultaneous successive linear estimator and a guide for hydraulic tomography analysis. *Water Resour. Res.* 45 (2), 1–14. <https://doi.org/10.1029/2008WR007180>.
- Yeh, T.-C., Liu, S., 2000. Hydraulic tomography: development of a new aquifer test method. *Water Resour. Res.* 36 (8), 2095–2105. <https://doi.org/10.1029/2000WR900114>.
- Yeh, T.-C., Srivastava, R., Guzman, A., Harter, T., 1993. A numerical model for water flow and chemical transport in variably saturated porous media. *Ground Water* 31 (4), 634–644. <https://doi.org/10.1111/j.1745-6584.1993.tb00597.x>.
- Yeh, T.-C.J., Mas-Pla, J., Williams, T.M., McCarthy, J.F., 1995. Observation and three-dimensional simulation of chloride plumes in a sandy aquifer under forced-gradient conditions. *Water Resour. Res.* 31 (9), 2141–2157. <https://doi.org/10.1029/95WR1947>.
- Yeh, T.-C., Mao, D.-Q., Zha, Y.-Y., Wen, J.-C., Wan, L.i., Hsu, K.-C., Lee, C.-H., 2015. Uniqueness, scale, and resolution issues in groundwater model parameter identification. *Water Sci. Eng.* 8 (3), 175–194. <https://doi.org/10.1016/j.wse.2015.08.002>.
- Younger, P.L., 1993. Simple generalized methods for estimating aquifer storage parameters. *Quart. J. Eng. Geol. Hydrogeol.* 26 (2), 127–135. <https://doi.org/10.1144/GSL.QJEG.1993.026.02.04>.
- Zha, Y., Yeh, T.-C.J., Illman, W.A., Onoe, H., Man, C., Mok, W., Wen, J.-C., Huang, S.-Y., Wang, W., 2017. Incorporating geologic information into hydraulic tomography: a general framework based on geostatistical approach. *Water Resour. Res.* 53 (4), 2850–2876. <https://doi.org/10.1002/2016WR019185>.
- Zha, Y., Yeh, T.-C., Illman, W.A., Mok, C.M.W., Tso, C.-H., Carrera, B.A., Wang, Y.-L., 2019. Exploitation of pump-and-treat remediation systems for characterization of hydraulic heterogeneity. *J. Hydrol.* 573, 324–340. <https://doi.org/10.1016/j.jhydrol.2019.03.089>.
- Zhang, D., 1999. Nonstationary stochastic analysis of transient unsaturated flow in randomly heterogeneous media. *Water Resour. Res.* 35 (4), 1127–1141. <https://doi.org/10.1029/1998WR900126>.
- Zhao, Z., Illman, W.A., 2018. Three-dimensional imaging of aquifer and aquitard heterogeneity via transient hydraulic tomography at a highly heterogeneous field site. *J. Hydrol.* 559, 392–410. <https://doi.org/10.1016/j.jhydrol.2018.02.024>.
- Zhao, Z., Illman, W.A., Berg, S.J., 2016. On the importance of geological data for hydraulic tomography analysis: laboratory sandbox study. *J. Hydrol.* 542, 156–171. <https://doi.org/10.1016/j.jhydrol.2016.08.061>.
- Zhu, J., Yeh, T.C.J., 2005. Characterization of aquifer heterogeneity using transient hydraulic tomography. *Water Resour. Res.* 41 (7), 1–10. <https://doi.org/10.1029/2004WR003790>.
- Zhuang, C., Zhou, Z., Illman, W.A., Guo, Q., Wang, J., 2017. Estimating hydraulic parameters of a heterogeneous aquitard using long-term multi-extensometer and groundwater level data. *Hydrogeol. J.* 25 (6), 1721–1732. doi: 10.1007/s10040-017-1596-y.