

Accepted Manuscript

Research papers

Comparative study of transient hydraulic tomography with varying parameterizations and zonations: Laboratory sandbox investigation

Ning Luo, Zhanfeng Zhao, Walter A. Illman, Steven J. Berg

PII: S0022-1694(17)30649-2
DOI: <https://doi.org/10.1016/j.jhydrol.2017.09.045>
Reference: HYDROL 22265

To appear in: *Journal of Hydrology*

Received Date: 23 June 2017
Revised Date: 21 September 2017
Accepted Date: 25 September 2017

Please cite this article as: Luo, N., Zhao, Z., Illman, W.A., Berg, S.J., Comparative study of transient hydraulic tomography with varying parameterizations and zonations: Laboratory sandbox investigation, *Journal of Hydrology* (2017), doi: <https://doi.org/10.1016/j.jhydrol.2017.09.045>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Comparative study of transient hydraulic tomography with varying parameterizations and zonations: Laboratory sandbox investigation

Ning Luo^a, Zhanfeng Zhao^{a,b}, Walter A. Illman^a and Steven J. Berg^{a,c}

Revised manuscript submitted to Journal of Hydrology

September 20, 2017

^a Department of Earth and Environment Sciences, University of Waterloo, Waterloo, ON, N2L3G1, Canada

^b Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

^c Aquanty Inc, Waterloo, ON, N2L5C6, Canada

Keywords: Aquifer heterogeneity; Hydraulic tomography; Transient analysis; Geological information; Inverse modeling; Model calibration and validation; Model comparison

* Corresponding author: Ning Luo, nluo1222@gmail.com

Abstract

Transient hydraulic tomography (THT) is a robust method of aquifer characterization to estimate the spatial distributions (or tomograms) of both hydraulic conductivity (K) and specific storage (S_s). However, the highly-parameterized nature of the geostatistical inversion approach renders it computationally intensive for large-scale investigations. In addition, geostatistics-based THT may produce overly smooth tomograms when head data used to constrain the inversion is limited. Therefore, alternative model conceptualizations for THT need to be examined. To investigate this, we simultaneously calibrated different groundwater models with varying parameterizations and zonations using two cases of different pumping and monitoring data densities from a laboratory sandbox. Specifically, one effective parameter model, four geology-based zonation models with varying accuracy and resolution, and five geostatistical models with different prior information are calibrated. Model performance is quantitatively assessed by examining the calibration and validation results. Our study reveals that highly parameterized geostatistical models perform the best among the models compared, while the zonation model with an excellent knowledge of stratigraphy also yields comparable results. When few pumping tests with sparse monitoring intervals are available, the incorporation of accurate or simplified geological information into geostatistical models reveals more details in heterogeneity and yields more robust validation results. However, results deteriorate when inaccurate geological information are incorporated. Finally, results reveal that transient inversions are necessary to obtain reliable K and S_s estimates for making accurate predictions of transient drawdown events.

1. INTRODUCTION

The detailed and accurate characterization of subsurface heterogeneity in hydraulic conductivity (K) and specific storage (S_s) are of great importance to groundwater resource management, its security, and remediation of contaminants. Typically, mapping of subsurface heterogeneity in K is accomplished through the geostatistical analyses of small-scale K values obtained from core samples, slug tests, flowmeter surveys, and single-hole pumping or injection tests. In contrast, heterogeneity in S_s has been ignored in many studies as its variability is considered to be much less than K . Hence, little work has been done in characterizing S_s heterogeneity.

One alternative to the geostatistical analysis of small scale data is hydraulic tomography (HT). The performance of HT has been evaluated through a number of numerical (Yeh and Liu, 2000; Bohling et al., 2002; Zhu and Yeh, 2005), laboratory (e.g., Liu et al., 2002, 2007; Illman et al., 2007, 2008, 2010, 2015; Berg and Illman, 2011a, 2012; Zhao et al., 2015, 2016), and field (e.g., Bohling et al., 2007; Straface et al., 2007; Illman et al., 2009; Cardiff et al., 2009, 2012, 2013; Berg and Illman, 2011b, 2013, 2015; Brauchler et al., 2011; Castagna et al., 2011; Paradis et al., 2016; Zha et al., 2015, 2016; Zhao and Illman, 2017) studies.

Fundamentally, HT involves the inverse modeling of hydraulic head data obtained during multiple pumping/injection tests. There are a number of inverse modeling approaches (e.g., Yeh and Liu, 2000; Bohling et al., 2002; Brauchler et al., 2003; Zhu and Yeh, 2005, 2006; Xiang et al., 2009; Cardiff and Barrash, 2011; Mao et al., 2013) to map the spatial variations of hydraulic parameters. For instance, Yeh and Liu (2000) proposed a sequential successive linear estimator (SSLE) to interpret steady state HT (SSHT) data. They evaluated this approach through the

examination of uncertainties associated with input parameters, such as mean values and correlation scales. However, the uncertainty related to the assumption of boundary conditions was not addressed. Zhu and Yeh (2005), then developed a transient hydraulic tomography (THT) algorithm based on SSLE that allowed for the joint estimation of heterogeneity in K and S_s as well as their uncertainties.

To overcome the impact of uncertain boundary conditions on K estimation and to maintain computational efficiency, Bohling et al. (2002) proposed a steady shape analysis of transient drawdown data for HT. Later, through a field study conducted in an alluvial aquifer located in Kansas, USA, Bohling et al. (2007) concluded that the steady shape analysis of transient drawdown data yields similar performance in estimating K profiles when compared to THT, suggesting the viability of the steady shape inversion approach. Nevertheless, the steady-state and steady shape approaches do not allow for the estimation of S_s , which is critical for assessing the availability of groundwater in a basin and is of paramount importance to groundwater resource management (Wu et al., 2005).

Hu et al. (2011) proposed a combined procedure of travel time (Brauchler et al., 2003) and steady shape (Bohling et al., 2002) inversions for subsurface heterogeneity characterization. In their study, the K distribution was estimated from steady shape inversion, while S_s values were calculated based on the estimated K profile and the obtained diffusivity (D) distribution from the travel time inversion ($S_s = K / D$). Therefore, S_s values were not jointly estimated with K during the inversions. Brauchler et al. (2013) then demonstrated through a field study at the Stegemuhle site in Germany that the combination of travel time and steady shape inversions is an efficient approach to characterize aquifer heterogeneity and to estimate the spatial distributions of hydraulic parameters.

However, through a HT survey conducted in one isolated subhorizontal bedrock fracture, Castagna et al. (2011) demonstrated that accurate knowledge of the spatial structures of S_s would help to obtain consistent representations of K and S_s fields. In contrast, the inaccurate assumption of spatial S_s structures (i.e., spatially heterogeneous S_s field is assumed to be homogeneous) would lead to misrepresentations of the S_s fields or poor representations of both K and S_s fields. Furthermore, based on the pilot point inversion results of HT data collected at the Stegemuhle site, Germany, Jimenez et al. (2015) pointed out that the inclusion of S_s in addition to K in inversions could help to minimize model misfit to field data. Contrasting results were obtained by Cardiff and Barrash (2011) who found that estimation of K is slightly degraded if storage parameters are jointly estimated. Therefore, further research is necessary in assessing the results from the simultaneous estimation of K and S_s from transient head data during HT surveys.

Another issue that deserves significant attention is what level of model complexity is required for HT analysis? To help answer this question, Illman et al. (2015) compared HT with different model complexities through the analysis of laboratory sandbox data of Illman et al. (2010). In particular, they compared the performance of: (1) isotropic and anisotropic effective parameter models, (2) a geological model with constant K value in each layer, and (3) a geostatistical model with a spatially variable K field. Only steady state head data were utilized for calibration and validation purposes. Results revealed that the geological model with perfect knowledge of stratigraphy performed nearly as well as the geostatistical model, especially when the number of pumping test data utilized for model calibration was reduced. Schöniger et al. (2015) also examined the issue of groundwater model complexity and experimental effort through a Bayesian model selection analysis using the steady state head data utilized by Illman et

al. (2015). They concluded that the geological zonation approach was most robust, but only if the zonation is accurate.

Illman et al. (2015) also concluded that the resulting resolution and accuracy of aquifer heterogeneity from the geostatistical interpretation of steady state head data depended on the amount of information included for model calibration affirming the conclusions by Yeh and Liu (2000) and Cardiff et al. (2013). Results from the study by Illman et al. (2015) revealed that details of aquifer heterogeneity were lost when the number of hydraulic head data was reduced for geostatistical inverse modeling, especially at or near locations where observation data were lacking. In particular, the estimated K tomograms from the sequential or simultaneous geostatistical inversions of head data were able to recover the major layers of high and low K values, but distinct layer boundaries were not recovered. These relatively smooth K fields were adequate in predicting the distributions of drawdowns from independent pumping tests not used in the calibration effort. However, the recovery of a finer scale resolution tomogram including layer boundaries is likely needed for improved predictions of solute and contaminant transport.

Parallel to the findings by Illman et al. (2015), Ahmed et al. (2015) demonstrated that K tomograms obtained from the geostatistical interpretation of steady state head data might still suffer from the issue of smoothness due to the inherent estimation of conditional means implied in most geostatistical inversion approaches, such as the quasi-linear geostatistical approach (Kitanidis, 1995) as well as the SSLE (Yeh and Liu, 2000) or the Simultaneous Successive Linear Estimator (SimSLE) (Xiang et al., 2009).

The issue of smooth distributions of estimated hydraulic parameters has also been discussed by Hu et al. (2011) and Jimenez et al. (2013, 2015). In particular, Jimenez et al. (2015) applied the travel time inversion of hydraulic head response data to obtain the information of

domain structural features, which in turn was used to guide the pilot point inversion of head data to estimate K and S_s tomograms. Zhou et al. (2014) also incorporated geological information in their inverse modeling of geophysical data. They proposed an image-guide inversion approach, in which, structural information was extracted from known geology and introduced to regularize the inversion process. More recently, Zhao et al. (2016) examined the value of integrating geological information on a HT survey through the SSHT analyses of multiple pumping test data from a laboratory sandbox (Illman et al., 2010). They found that utilizing an accurate geological model as a prior estimate for geostatistical inversions was beneficial in improving the K tomograms, layer boundaries and their connectivity.

Most recently, Zhao and Illman (2017) investigated the value of geological information on SSHT analysis of multiple pumping tests at the North Campus Research Site located on the University of Waterloo campus in Waterloo, Canada. Both the laboratory (Zhao et al., 2016) and field-based (Zhao and Illman, 2017) studies suggested the importance of including accurate geological information to improve the results of SSHT analyses of pumping test data. However, whether this conclusion translates to THT analysis in which both K and S_s are jointly estimated remains unknown.

The main objectives of this study are: 1) to extend the work of Illman et al. (2015) to the transient case to compare HT inversions of varying model complexities; and 2) to extend the work of Zhao et al. (2016) to evaluate the utility of geological information for THT analysis through the analyses of laboratory sandbox data collected by Illman et al. (2010). Since the investigation is performed in a controlled sandbox with perfect knowledge of geological structures, this study will be helpful in identifying conditions in which geological models can be useful for future field HT studies.

2. EXPERIMENTAL SETUP

2.1 Sandbox Description and Collected Data

A two-dimensional synthetic heterogeneous aquifer constructed in a laboratory sandbox is characterized using inverse models of various parameterization and zonations. The length, height, and width of the sandbox are 192.0 cm, 82.6 cm, and 10.2 cm, respectively. The sandbox is capable of maintaining three constant head boundaries simultaneously by setting two constant head reservoirs at each end and ponding water at the top. The remaining three boundaries (front, back, and bottom) are no-flow boundaries.

To create a realistic heterogeneity pattern, Illman et al. (2010) built this synthetic aquifer through the cyclic deposition of sediments under varying water flow and sediment feed rates, mimicking an interfingering natural fluvial deposit. In particular, a uniform flow rate and a specific sand type were chosen for a given layer deposition, producing small-scale heterogeneities within each layer and larger-scale heterogeneities for different layers. Through a sediment transport process, 18 layers of varying size sands were deposited, as shown in Fig. 1. Such a synthetic heterogeneous aquifer with exactly known stratigraphy is necessary to investigate the effect of geological information on groundwater flow modeling. Upon completing the deposition of the layers, 48 ports were constructed throughout the aquifer along six columns with eight ports each (see Fig. 1). Each port has a diameter of 1.3 cm and fully penetrates the aquifer width. These ports can be utilized for pumping and injection of water, as well as monitoring head levels with a pressure transducer instrumented at each port.

The synthetic aquifer was then characterized with different techniques. Core samples extracted from these ports were subjected to grain size analyses and permeameter tests to

estimate local K values. In addition, single-hole pumping tests were performed at each port to obtain small-scale estimates of K and S_s . The median value of the particle size distribution (d_{50}), as well as K and S_s estimates of the different layers are summarized in Table 1. When multiple ports are available in a given layer, the geometric mean of their estimates are provided.

Twenty-four cross-hole pumping tests were also conducted in the synthetic aquifer with constant pumping rates ranging from 2.50 to 3.17 mL/s. These tests were conducted at 16 ports along columns 2 (ports 2, 8, 14, 20, 26, 32, 38, and 44) and 5 (ports 5, 11, 17, 23, 29, 35, 41, and 47), as well as at eight additional ports (ports 13, 15, 16, 18, 37, 39, 40, and 42). Prior to each pumping test, all pressure transducers were calibrated to ensure accurate data collection and head levels in all ports were monitored over several minutes to establish a static, initial condition. During each pumping test, hydraulic head responses in all 48 ports were recorded until the aquifer reached a steady state condition which was determined by observing the stabilization of all pressure head measurements within the aquifer. The pump was then turned off to allow for the full recovery of hydraulic heads.

2.2 Data Used for Modeling

In this study, transient head data obtained from eight cross-hole pumping tests are utilized for inverse modeling and the remaining 16 tests are reserved for validation purposes. Prior to extracting data points from the drawdown records, the presence of drift in transducers is accounted for using the scheme discussed by Illman et al. (2007). Furthermore, head records from pumped ports are excluded from the analysis because these data are found to be excessively noisy due to the use of a peristaltic pump and skin effects. Previous research (Illman et al., 2007; Xiang et al., 2009) has shown that inclusion of data from the pumped port would lead to biased

results and poor hydraulic parameter estimates. In order to smooth the data, pressure head data from observation ports are then fit with a fifth- or sixth-order polynomial curve (Liu et al., 2007), and five data points that represent the early, intermediate, and late times of aquifer responses are extracted from each curve. For ports where the pressure head curve could not be properly fit with a polynomial, five data points are extracted manually to represent the overall behavior of pressure heads. In total, 235 data points are extracted from each pumping test.

Two cases, depending on the number of pumping tests and the density of observation ports, are chosen for this study to assess the performance of different models. For Case 1, eight pumping tests (ports 2, 5, 14, 17, 32, 35, 44, and 47) and data points from 47 observation ports are utilized for model calibration, while the remaining 16 independent pumping tests (ports 8, 11, 13, 15, 16, 18, 20, 23, 26, 29, 37, 38, 39, 40, 41, and 42) are utilized for model validation. For Case 2, only four pumping tests (ports 26, 29, 44, and 47) and 15 observation ports along the second and the fifth well columns from the left boundary of the sandbox are utilized for model calibration. We select 16 ports for Case 2 to represent the presence of only two wells with multiple screens at various depths, and this case mimics an actual field scenario where wells are sparse and the amount of pumping test data is limited for site investigation. To be consistent with Case 1, 16 independent pumping tests are utilized for model validation in Case 2.

3. GROUNDWATER FLOW MODELING APPROACHES

3.1 Model Setup

The synthetic aquifer is discretized into 741 elements and 1,600 nodes with element dimensions of 4.1 cm \times 4.1 cm \times 10.2 cm for all forward and inverse groundwater flow models considered. A finer mesh was also tested in a previous study (Illman et al., 2012), but the results

did not show significant changes in comparison to the coarser one. Therefore, for consistency with previous studies (Illman et al., 2010; Berg and Illman, 2011a; Zhao et al., 2016) who used this discretization, this coarse grid is utilized here.

Transient groundwater flow can be described by the following equation:

$$\nabla \cdot [K(\mathbf{x})\nabla h] + Q(\mathbf{x}_p) = S_s(\mathbf{x}) \frac{\partial h}{\partial t} \quad (1)$$

subject to initial and boundary conditions:

$$h|_{t=0} = h_0, h|_{\Gamma_1} = h_1, \text{ and } [K(\mathbf{x})\nabla h] \cdot \mathbf{n}|_{\Gamma_2} = q \quad (2)$$

where, in equation (1), ∇ is the gradient operator, $K(\mathbf{x})$ is hydraulic conductivity (L/T), h is hydraulic head (L), $Q(\mathbf{x}_p)$ is the pumping rate (L³/T) at location \mathbf{x}_p , and $S_s(\mathbf{x})$ is specific storage (1/T). In equation (2), h_0 represents the initial hydraulic head, h_1 is a constant head (L) at boundary Γ_1 , q is the specific discharge (L/T) at Neumann boundary Γ_2 , and \mathbf{n} is a unit vector normal to Γ_2 . In this study, the transient flow equation is solved by a 3-D finite element model MMOC3 (Yeh et al., 1993) for all cases described next.

3.2 Effective Parameter Model

The synthetic aquifer is first characterized as a homogeneous, isotropic medium to estimate the effective K and S_s values by coupling the groundwater flow model MMOC3 (Yeh et al., 1993) with the parameter estimation code PEST (Doherty, 2005). We did not consider the case in which the hydraulic parameters are treated to be anisotropic because previous research by Illman et al. (2015) showed little difference in the isotropic and anisotropic results.

The effective parameter model provides zero-resolution on aquifer heterogeneity. However, it may still be able to describe the overall behavior of the aquifer, particularly when multiple pumping tests are included for the calibration effort. Furthermore, the estimated effective K and S_s values can be used as the initial guesses of hydraulic parameters to guide the calibration of more sophisticated (i.e., highly parameterized) groundwater flow models.

For each case, all pumping test data are included for the estimation of effective parameters. In total, 1,880 data points are used for Case 1, while 300 data points are utilized for Case 2. The forward model is then automatically calibrated to obtain an optimal set of K and S_s by simultaneously matching all data points. The initial values of K and S_s input into PEST are 0.06 cm/s and 6.1×10^{-4} /cm, respectively, which are the geometric means from the kriged K and S_s fields based on the estimates from single-hole pumping tests (Berg and Illman, 2011a). In PEST, the minimum and maximum bounds are set as 1×10^{-4} and 10 cm/s for K , and 1×10^{-8} and 1.0 /cm for S_s .

3.3 Geology-based Zonation Modeling Approach

The synthetic aquifer is then characterized using various geology-based zonation models. In this approach, the synthetic aquifer is divided into different zones based on available geological information. In each zone, the porous medium is treated to be homogeneous as well as isotropic, and a uniform set of K and S_s is estimated and assigned to describe its hydraulic properties. To assess the impact of accuracy of geological information on groundwater flow modeling, four geology-based zonation models with various accuracy and resolution (GOOD, POOR1, POOR2, and POOR3, as shown in Fig. 2) are considered, following the work of Zhao et al. (2016).

These geological models are constructed using Leapfrog Hydro (ARANZ Geo Ltd.) through the inclusion of borehole information obtained along the six columns of ports. In comparison to directly mapping the stratification through the sandbox glass (e.g., the “perfect” geological model utilized in Berg and Illman (2011a) and Illman et al. (2015)), interpolation of borehole logs is more consistent in constructing geological models from field data. By interpolating incorrect stratigraphy information and/or introducing random errors to the layer thickness records, poor geological models (Figs. 2b and 2d for POOR1 and POOR3, respectively) are constructed to mimic cases of inaccurate identification of borehole information from the field. Detailed description of these geological models can be found in Zhao et al. (2016).

Different from Zhao et al. (2016), the geological model POOR2 (Fig. 2c) used in this study is constructed based on the GOOD model (Fig. 2a) by merging some layers with similar material types. This simplified geological model with only five zones is constructed to represent the scenario with a simplified description of the stratigraphy, but with well identified layer boundaries.

The four geological models are then discretized using the grid described above to construct geology-based zonation models for aquifer characterization. In a similar fashion to the effective parameter model, all geology-based zonation models are calibrated using PEST coupled with MMOC3 by simultaneously matching all data points. For each case study, the K and S_s values obtained from the effective model are used as the initial guesses of hydraulic parameters for model calibration, while the bounds of K and S_s are set to be the same as those in the effective parameter model. In total, 36 parameters are estimated for geology-based zonation models GOOD, POOR1, and POOR3, while only 10 parameters are estimated for the POOR2 model.

3.4 Geostatistical Inverse Modeling Approach

All geostatistical inversions are conducted using the Simultaneous Successive Linear Estimator (SimSLE), developed by Xiang et al. (2009). This inversion approach provides an efficient way to include all data points from multiple pumping tests simultaneously for hydraulic parameter estimation. In comparison to the Sequential Successive Linear Estimator (SSLE) developed for THT analysis (Yeh and Liu, 2000), SimSLE provides more constraints to the inverse problem, resulting in faster convergence (Xiang et al., 2009). Additionally, SimSLE avoids the computation of varying final estimates when HT data are analyzed in different sequences with SSLE (Illman et al., 2008).

Geostatistical inversion using SimSLE assumes a transient groundwater flow field, and the natural logarithm of K and S_s are both treated as multi-Gaussian, second-order stationary, stochastic processes. With given unconditional means, variances, and correlation scales of hydraulic parameters, the Successive Linear Estimator (SLE) implemented in SimSLE first creates $\ln K$ and $\ln S_s$ fields by cokriging their initial estimates and head data from all pumping tests. These parameter fields are then used to solve the governing flow equation to obtain simulated head data. Based on the differences between observed and simulated head data, SLE improves the estimates of K and S_s . The iteration process continues until convergence is achieved by: (1) the stabilization of the spatial variances of estimated hydraulic parameters; and/or (2) the difference of simulated heads between successive iterations being smaller than a prescribed tolerance.

Based on the differences in initial K and S_s fields for geostatistical inversions, two cases are investigated. First, homogeneous initial K and S_s fields are used for model calibration, which is consistent with previous HT studies (e.g., Liu et al., 2002; Illman et al., 2008; Berg and Illman, 2011a). In this case, the effective K and S_s obtained from the homogeneous model provided in

the next section are used as initial guesses and assigned to the entire simulation domain. Second, the initial K and S_s fields used for geostatistical inversions of HT data are treated to be heterogeneous and obtained from the calibrated geology-based zonation models.

For all cases, the variances of K and S_s ($\sigma_{\ln K}^2$, $\sigma_{\ln S_s}^2$) are initially set as 3.0, while the correlation scales are set as $\lambda_x = 30.0$ cm, $\lambda_y = 10.2$ cm, and $\lambda_z = 10.0$ cm for both K and S_s based on the statistical properties of kriged K and S_s fields from single-hole estimates (see Tables 3 and 4 in Berg and Illman, (2011a)). These values have been found to have negligible effects on the results due to the availability of large number of head measurements during a HT survey (Yeh and Liu, 2000).

4. RESULTS AND DISCUSSION

4.1 Effective Parameter Model

Treating the entire synthetic aquifer as a homogeneous/isotropic medium, two sets of effective K and S_s are estimated through the inclusion of a different number of head data for inversions. For Case 1, in which data from eight pumping tests and 47 observation ports are utilized for calibration, the effective K and S_s as well as their 95% confidence intervals are estimated as $K = 9.57 \times 10^{-2} \pm 2.15 \times 10^{-3}$ cm/s and $S_s = 6.32 \times 10^{-5} \pm 4.30 \times 10^{-6}$ /cm. For Case 2, the effective K and S_s with 95% confidence intervals are $K = 9.58 \times 10^{-2} \pm 4.81 \times 10^{-3}$ cm/s and $S_s = 7.25 \times 10^{-5} \pm 1.11 \times 10^{-5}$ /cm. The 95% confidence intervals are calculated with PEST based on the implied linearity assumption used to derive the equation for parameter improvement.

The estimated effective K and S_s from the two cases are close to each other, while the 95% confidence intervals for both K and S_s increase after reducing the number of data points used for

inverse modeling. This result implies that more pumping tests with dense observation locations may still be required to obtain reliable estimates of effective parameters.

Previously, Berg and Illman (2011a) determined the effective parameters of the synthetic aquifer by taking the 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). These values, however, were found to be poor in predicting drawdowns from independent pumping tests, suggesting that the effective parameters obtained from local estimates may not be representative of the aquifer. Illman et al. (2015) reached the same conclusion based on the effective K that they estimated for the same synthetic aquifer. In particular, they found that the effective K estimated by simultaneously analyzing eight pumping tests provided improved results in terms of model calibration and validation in comparison to the work of (Illman et al., 2010) in which local estimates (e.g., core and single-hole test results) were utilized to generate the effective K value of the aquifer.

Consequently, the effective K and S_s estimates obtained in this study by simultaneously analyzing multiple pumping tests are considered to be more representative of the aquifer in comparison to those from Berg and Illman (2011a). These values are then utilized as initial guesses of hydraulic parameters for geology-based zonation and geostatistical models, as discussed below.

4.2 Geology-Based Zonation Models

Four different geology-based zonation models are calibrated for each dataset case. Figs. 3 and 4 show the estimated K and S_s tomograms from different zonation models for Cases 1 and 2, respectively. Examination of Figs. 3 and 4 reveals that the estimated locations of high and low K zones vary from one zonation model to another, when the same number of head data are included

for model calibration. This is because these models are calibrated with fixed zones of parameters. With a fixed geological model, PEST focuses on the estimation of parameter values of each zone to fit the simulated to observed data as close as possible. These results imply that attention should be paid when constructing zonation models for aquifer heterogeneity characterization, since the inaccurate identification of structural features could lead to unrealistic parameter estimates. In addition, the comparison of results from Cases 1 and 2 (Figs. 3 and 4) when the same zonation model is calibrated reveals that the identification of high and low K zones varies when different numbers of head data are included for model calibration.

The estimated S_s tomograms do not show distinct structural features, except for the simplified zonation model POOR2 (Figs. 3f and 4f), revealing that the heterogeneity of S_s in this synthetic aquifer is milder in comparison to that of the K . Moreover, the estimated S_s values decrease from the top to the bottom for most zonation models. Such a decreasing trend of S_s was explained by Berg and Illman (2011a) that the upper sands were less compressed compared to the deeper sand bodies. Although the spatial variance of S_s is estimated to be relatively small, differences in estimated S_s tomograms can still be observed.

Within each zone, a uniform set of K and S_s is estimated to describe its hydraulic properties. Fig. 5 shows the estimated K values as well as their corresponding 95% confidence intervals from all zonation models for both cases, while Fig. 6 shows the same, but for S_s estimates. The estimated K and S_s values as well as their 95% confidence intervals are provided in the Supplementary Material section as Tables S1 to S4 for the GOOD, POOR1, POOR2, and POOR3 models. Fig. 5 reveals that when calibrating geology-based zonation models with eight pumping tests (Case 1), all K estimates have narrow confidence intervals, suggesting the high confidence of these estimates. However, when the number of head data is reduced, noticeable

increases in the confidence intervals of K estimates are observed in some zones; in particular, Zone 6 for the POOR1 model and Zones 8, 13, and 18 for the POOR3 model. The main reason for this is that no observation data are available in these zones when observation ports are reduced from 47 (Case 1) to 15 (Case 2).

It is interesting to note that all S_s estimates from the calibrated geology-based zonation models result in narrow confidence intervals (Fig. 6), except for the case in which the simplified geology-based zonation model (POOR2) is calibrated using fewer data (as shown in green areas of Fig. 6c). With given structural features, the obtained S_s estimates for different zones are close to each other. This is also the case when comparing S_s estimates from different zonation models. These results suggest that the estimation of S_s for this synthetic aquifer is less likely to be affected by structural errors.

4.3 Geostatistical Inverse Model with Homogeneous Initial K and S_s Fields

Without providing additional prior information, the geostatistical inversion of THT data using SimSLE starts with homogeneous initial K and S_s fields. In SimSLE, the L_2 norm between the simulated and observed head is computed for each iteration. Fig. S2 in the Supplementary Material section shows how the L_2 norm evolves with the iteration number. As suggested by Xiang et al. (2009), we select the inversion results when the L_2 norm stabilizes. Here, stabilization in the L_2 norm is meant when the variation of L_2 from one iteration to the next becomes smaller than 3×10^{-4} (see Fig. S2).

Fig. 7 shows the estimated K and S_s tomograms as well as the corresponding $\ln K$ and $\ln S_s$ variance maps for Case 1, while Fig. 8 shows the same, but for Case 2. Black lines in K tomograms indicate the exact layer boundaries of the synthetic aquifer, which are delineated

based on the photograph of the deposits (see Fig. 1). Such stratigraphic information is not included for the estimated S_s tomograms because the spatial variation of S_s of this synthetic aquifer does not reveal distinct structural features. Different from the zonation modeling approach, the geostatistical inversion of THT data estimates hydraulic parameters for each finite element, resulting in relatively smooth distribution of K and S_s estimates. Through the simultaneous inversion of transient head data from eight pumping tests with 47 observation ports (Case 1), the estimated K tomogram (Fig. 7a) reveals considerable details to aquifer heterogeneity. The estimated high and low K zones show significant agreement to most visible layers of the synthetic aquifer in terms of their positions. When fewer pumping tests with fewer observation ports are utilized for the geostatistical inversion (Case 2), the estimated K tomogram (Fig. 8a) shows a similar pattern of high and low K zones, but with great loss of detail in heterogeneity, particularly on both sides of the aquifer where observation data are removed. Although the estimated S_s of this synthetic aquifer is much less variable when compared to K , the loss of detail in heterogeneity can still be observed in the estimated S_s tomograms (Figs. 7c and 8c for Cases 1 and 2, respectively). The comparison of Cases 1 and 2 reveals that a large number of pumping tests with dense observation intervals is required to capture most heterogeneity features, which is in line with the conclusion of previous studies (e.g., Cardiff et al., 2013; Illman et al., 2015). This study shows that the inclusion of additional hydraulic head data through transient analysis does not negate the necessity of a large number of monitoring points to accurately depict the structural features and their boundaries.

The corresponding $\ln K$ and $\ln S_s$ variances computed by SimSLE indicate the uncertainty of parameter estimates, with larger variance values indicating higher uncertainty. For each case, small $\ln K$ and $\ln S_s$ variances are obtained around pumping ports, while variances become larger

when moving away from the ports. In general, for both cases, the $\ln S_s$ variances are larger than those of $\ln K$, revealing that it is more difficult to estimate S_s . Comparing variance maps from Cases 1 and 2, a significant increase in values are observed for both $\ln K$ and $\ln S_s$ variances when fewer head data are utilized for geostatistical inversions. These results indicate that more accurate K and S_s tomograms will be obtained in areas where there are available head data in comparison to areas where head data are lacking, which again emphasizes the importance of the availability of head data for aquifer heterogeneity characterization using geostatistical models.

The geostatistical inverse modeling of transient head data using SimSLE is demonstrated to reveal great details of aquifer heterogeneity; however, the estimated major zones fail to capture the precise shapes of stratigraphic features by using hydraulic head data only. Without providing the layer information prior to inverse modeling, the estimated K tomograms result in smooth transitions from one layer to the next, and the layer boundaries become ambiguous, especially when the amount of head data is limited for aquifer characterization. To improve the results in terms of preserving stratigraphic feature shapes and revealing layer boundaries, additional information is needed when conducting geostatistical inversions for aquifer heterogeneity characterization. We next utilize geological information as prior estimates for geostatistical inverse modeling.

4.4 Geostatistical Inverse Models with Heterogeneous Initial K and S_s Fields

The incorporation of geological information into the geostatistical inversion approach is achieved by constructing geology-based heterogeneous initial parameter fields for model calibration. Zhao et al. (2016) applied this method for the SSHT analysis of head data. In their study, permeability K values were assigned to geological models to construct the initial K fields.

Different from Zhao et al. (2016), this study utilizes the estimated K and S_s tomograms from the calibrated geology-based zonation models as initial guesses. The utilization of calibrated geological information avoids the uncertainty associated with small scale estimates. To provide a detailed investigation of the effect of geological information on aquifer heterogeneity characterization, four types of geological information with varying accuracy and resolution (GOOD, POOR1, POOR2, and POOR3) are incorporated for both Cases 1 and 2.

Fig. 9 illustrates the estimated K and S_s tomograms for Case 1 (see Fig. S2a for L_2 norms). In particular, Figs. 9a and 9b show the K and S_s tomograms, respectively, when the GOOD geological information is incorporated into the geostatistical inversion of THT data. Black lines that represent the accurate stratification of the synthetic aquifer are also included in the K tomogram. In comparison to the K tomogram estimated from the homogeneous initial field (Fig. 7a), Fig. 9a preserves more stratigraphic features and layer boundaries. On the other hand, the estimated S_s tomogram (Fig. 9b) remains almost the same as the one estimated without providing geological information (Fig. 7c).

The estimated K and S_s tomograms when inaccurate geological information is included as initial guesses are shown as Figs. 9c-9h. The estimated K tomograms (Figs. 9a, 9c, 9e, and 9g for GOOD, POOR1, POOR2, and POOR3, respectively) are similar in terms of the patterns of estimated high and low K zones, while the shapes and the continuity of these zones are slightly different among each other. Differences can also be observed in the estimated S_s tomograms (Figs. 9b, 9d, 9f, and 9h for GOOD, POOR1, POOR2, and POOR3, respectively) when geological information of varying resolution and accuracies are incorporated into inverse modeling. In particular, when the simplified geological information (POOR2) is included as an

initial guess, unexpected low S_s zones are estimated across the middle of the simulation domain (Fig. 9f).

Fig. 10 shows the estimated K and S_s tomograms for Case 2 (see Fig. S2b for L_2 norms) when four types of geological information are introduced separately during the calibration process. Through the incorporation of the GOOD geological model as prior information, the estimated K tomogram (Fig. 10a) reveals more details of heterogeneity in comparison to the one obtained with homogeneous initial parameter fields using the same dataset (Fig. 8a), particularly at the fringes of the aquifer where observation ports are lacking. Even with limited head data, the estimated high and low K zones still show significant agreement with most layers. However, the estimated S_s tomogram does not show significant change in comparison to the one without providing geological information (Fig. 8c).

In contrast to Case 1 results, the estimated K tomograms in Case 2 are quite different among each other when various geological information are incorporated. Similar differences in estimated S_s tomograms are also visible in terms of the pattern of high and low S_s zones as well as their shapes. This suggests that when pumping and observation densities are high, hydraulic head data will dominate the inversion process and lead to similar K and S_s tomograms rather than reflect the prior geological information. However, the effects of prior geological information on inverse modeling become more significant when fewer pumping test data are available for calibration. The accuracy of these K and S_s tomograms are examined in later sections through the investigation of their abilities in predicting drawdowns from independent pumping tests.

4.5 Model Calibration and Validation

The calibration and validation results associated with different models are first assessed qualitatively by plotting the scatterplots of simulated versus observed drawdowns, which provides visual information of the spatial distribution of errors in terms of scatter and bias. Then, quantitative evaluation of model error is performed by computing the mean absolute errors (L_1) and mean square errors (L_2) between simulated and observed drawdown values using:

$$L_1 = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (3)$$

$$L_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (4)$$

where n is the total number of drawdown data, i indicates the data number, x_i and \hat{x}_i represent i -th simulated and observed drawdown values, respectively. The L_1 norm is calculated to analyze the discrepancy between simulated and observed drawdowns, while the L_2 norm tends to magnify large discrepancies and allow one to better assess the performance of different models.

The calibration scatterplots of all investigated models are provided in the Supplementary Material section and illustrated as Figs. S3 - S4 for Case 1, and as Figs. S5 - S6 for Case 2. A linear model is fit to each scatterplot and the corresponding coefficient of determination (R^2) values are provided. These scatterplots reveal that the calibration result improves when a larger number of estimated parameters are considered in the inverse model, and the geostatistical model yields the best result. This makes sense since the highly parameterized geostatistical model allows for the adjustment of K and S_s estimates in each element to fit the observation data.

For Case 1, the validation scatterplots that compare the simulated drawdown values from different models against the observed drawdowns from 16 independent pumping tests are illustrated in Figs. 11 and 12. In each scatterplot, the linear model of the fit, as well as the coefficient of determination (R^2) are provided at the bottom. Comparing the three different modeling approaches (Fig. 11), the geostatistical model (Fig. 11f) performs the best in predicting drawdowns for the entire domain, closely followed by the GOOD geology-based zonation model (Fig. 11b), while the utilization of effective homogeneous model (Fig. 11a) yields biased predictions of drawdowns with relatively larger scatter. This is consistent with the SSHT results of Illman et al. (2015). On the other hand, it is interesting to note that the zonation models based on inaccurate geological information (Figs. 11c - 11e) yield slightly better prediction results in comparison to the effective homogeneous model. The main reason for this is that more parameter sets are estimated for zonation models (18 for POOR1 and POOR3, and 5 for POOR3), which in turn emphasizes the importance of parameterization for groundwater flow modeling. Examination of Fig. 12 reveals that the estimated K and S_s tomograms from the geostatistical inverse model with GOOD geological information (Fig. 12a) as initial parameter fields yield minor improvements in predicting drawdown values for the entire domain compared to the case with homogeneous initial parameter fields (Fig. 11f). On the other hand, the incorporation of inaccurate geological information does not significantly impact the prediction results. This makes sense because the inversion results reflect more about hydraulic head information rather than prior geological information when abundant head data are available for inverse modeling, and the resulting K and S_s tomograms perform similarly in predicting independent pumping test data.

For Case 2, the validation scatterplots are illustrated in Figs. 13 and 14. The effective homogeneous model (Fig. 13a) still performs the worst in predicting drawdowns from

independent pumping tests. However, it is surprising to find that the GOOD geology-based zonation model (Fig. 13b) provides prediction results that are indistinguishable to the geostatistical inverse model with homogeneous initial K and S_s fields (Fig. 13f). This finding suggests that when the number of head data is limited, the utilization of geological model with good knowledge of stratification yields results that are comparable to those obtained by the geostatistical model, which is in line with the conclusion provided by Illman et al. (2015), who only analyzed steady state head data. However, it should be noted that the GOOD geology-based zonation model utilized in this study is constructed based on a large amount of borehole data with accurate identification of stratifications, which is difficult to obtain in the field. On the other hand, the validation results associated with other geology-based zonation models (Figs. 13c, 13d, and 13e for POOR1, POOR2, and POOR3, respectively) are not as good as the result provided by the geostatistical inverse model.

For Case 2, some differences in validation scatterplots are evident for geostatistical inverse models with different heterogeneous initial parameter fields, as shown in Fig. 14. Through the incorporation of GOOD geological information, the estimated K and S_s tomograms provide improved prediction results (Fig. 14a) for the entire domain with higher correlation between simulated and observed drawdowns in comparison to the case with homogeneous initial parameter fields (Fig. 13f). Slight improvements in terms of bias and scatter are also observed in Fig. 14c, in which simplified geological information (POOR2) is incorporated. On the other hand, when inaccurate stratigraphy and layer thickness information are introduced during model calibration, the estimated K and S_s tomograms provide worse prediction results (Figs. 14b and 14d for POOR1 and POOR3, respectively) in comparison to the case with homogeneous initial parameter fields.

The L_1 and L_2 norms of calibration and validation results are summarized in Tables 2 for Cases 1 and 2. These values are obtained by averaging the results from different pumping tests for each model calibration and validation, while the norms associated with individual pumping tests are presented in the Supplementary Material section as Tables S5 - S8 for Case 1, while as Tables S9 - S12 for Case 2. For both cases, the highly parameterized geostatistical model performs consistently across the different pumping tests (shown as Tables S5 - S12), suggesting that the approach is more consistently reliable in characterizing aquifer heterogeneity and predicting drawdowns in comparison to the effective homogeneous and geology-based zonation modeling approaches. However, the highly parameterized geostatistical model may suffer from the issue of over-parameterization and lead to ill-posed inversion problems. In this case, a large number of dataset (e.g., transient head responses obtained from HT) is required for the geostatistical inversion model to estimate reliable spatial distributions of hydraulic parameters (Schöniger et al., 2015). After incorporating geological information into geostatistical model calibration, L_1 and L_2 norms are found to be comparable for Case 1, while significant differences are observed for Case 2, especially when the estimated K and S_s tomograms are used in predicting drawdowns from independent pumping tests.

Comparison of results from Cases 1 and 2 reveals that geological information becomes increasingly important for aquifer heterogeneity characterization, when fewer pumping tests and observation data are available. However, close attention should be paid in obtaining accurate geological data, since the incorporation of inaccurate geological information adversely impacts the accuracy of parameter estimates, which in turn leads to poor predictions of independent pumping tests.

4.6 Predictability of Transient Drawdown Curves

To further investigate the performance of different models in predicting independent pumping tests, simulated drawdown curves at 16 selected ports are plotted against actual data. Figs. 15 and 16 illustrate the simulated drawdown curves using K and S_s tomograms from different models when conducting a pumping test at port 40 for Case 1, while Figs. 17 and 18 illustrate the same, but for Case 2. Results for all other pumping tests used for model validation are provided in the Supplementary Material section as Figs. S7 - S66. In each subplot, the observation data are expressed as green dots, while the simulated drawdown curves from the various models are plotted with different colors and types.

Fig. 15 shows that when a large number of head data is used for inverse modeling, the utilization of K and S_s tomograms from the geostatistical model with homogeneous initial parameter fields is able to predict drawdowns at most of the ports, followed by the zonation model with GOOD geological information. The performance of the effective model, as well as other geology-based zonation models in predicting drawdowns, in general, are poorer and vary from one port to another. Upon incorporating geological information as initial guesses into geostatistical models, Fig. 16 illustrates that the estimated K and S_s tomograms perform similarly among each other and the predictions are excellent for most ports.

When both pumping tests and observation ports are reduced, Fig. 17 shows that the geostatistical model with homogeneous initial K and S_s fields fails to capture drawdowns at some ports, particularly for the ports located at the fringes of the aquifer (e.g., ports 13, 18, 25, 30, 37, and 42). The main reason for this is that observation data at these ports are removed for inverse modeling, and the estimated K and S_s tomograms fail to reveal the details of heterogeneity in these areas, as shown in Figs. 7a and 7c. In contrast, by providing accurate stratifications, the K

and S_s tomograms associated with geostatistical model provides better predictions of drawdown curves at these ports (Fig. 18). This result again suggests that when the number of head data is limited for aquifer characterization, a good knowledge of stratification is quite important, and it can be incorporated into geostatistical models to reveal more details in heterogeneity and provide more accurate prediction results. Overall, integration of geological information into hydraulic tomography is a good practice and should result in better results when accurate geological data are available.

5. ON THE VALUE OF TRANSIENT ANALYSIS OF HYDRAULIC TOMOGRAPHY DATA

One remaining question is whether one should preferentially conduct transient inversions instead of steady state inversions for HT analysis. While steady state HT analyses under laboratory conditions are fast and have been shown to produce reliable K tomograms (Illman et al., 2007, 2008, 2010, 2015), reaching steady state conditions in the field requires long pumping tests, assuming it is possible to reach steady state, and the effects of the boundary conditions may affect the K estimates. Transient inversions, on the other hand, can be conducted with pumping tests of shorter durations and the effects of boundary conditions may be mitigated. Moreover, Castagna et al. (2011) has shown that to obtain more reliable K estimates, the simultaneous inversion of both K and S_s are necessary.

To investigate this issue, two additional cases are run by conducting SSHT, using the same pumping and observation densities (Cases 1 and 2) utilized for THT analyses. Results from the SSHT are then compared to those from THT presented earlier. In particular, K tomograms

from SSHT with a geometric mean of 48 S_s values ($S_s = 6.1 \times 10^{-4}$ /cm) from single-hole tests are used to conduct forward simulations of 16 independent pumping tests.

Results (Fig. 19) reveal that the drawdown predictions are significantly biased for both Cases 1 (Fig. 19a) and 2 (Fig. 19b) suggesting that in order to achieve accurate predictions of transient drawdowns, transient inversions are necessary.

ACCEPTED MANUSCRIPT

6. SUMMARY AND CONCLUSIONS

In this study, a synthetic heterogeneous aquifer constructed in a laboratory sandbox (Illman et al., 2010) is characterized with transient hydraulic tomography (THT) using various parameterization and zonation models through the simultaneous inversion of transient head data from multiple pumping tests. The main objectives of this THT study are: (1) to evaluate the performances of differently parameterized models in aquifer characterization and (2) to investigate the impact of geological information for inverse modeling by directly calibrating geology-based zonation models and incorporating them as prior information for geostatistical inverse models. Two cases of different pumping and observation densities are selected to accomplish this study. For Case 1, transient head data from eight pumping tests with 47 observation ports are simultaneously included for model calibration. For Case 2, the number of pumping tests is reduced to four, and head data from 15 observation ports are utilized.

The aquifer is first characterized as a homogeneous medium to estimate the effective K and S_s . Then, four geology-based zonation models of varying accuracy and resolution are constructed to characterize the aquifer with fixed zones of parameters. After that, geostatistical inverse models are utilized to map the heterogeneity in K and S_s . In addition to the case with homogeneous initial parameter fields, the estimated K and S_s tomograms from the calibrated geology-based zonation models are utilized as heterogeneous initial parameter fields for geostatistical inversions. These models are then validated through the prediction of drawdowns from 16 independent pumping tests. This study leads to the following findings and conclusions:

1. Treating the synthetic aquifer as a homogeneous medium, the simultaneous inversion of multiple pumping tests yields effective K and S_s estimates that are more representative of

the aquifer in comparison to those generated from small scale estimates. These effective values, however, are still less reliable in predicting independent pumping tests, suggesting that the accurate mapping of aquifer heterogeneity is necessary in building more robust groundwater flow models.

2. For each dataset case, all geology-based zonation models are well calibrated. This is because the calibration process related to the zonation modeling approach forces the estimation of parameters in each zone to fit the simulated data as close as possible to observed ones, which in turn results in different patterns of estimated high and low value zones of K and S_s among different geology-based zonation models. However, only the zonation model with good geological information is found to be adequate in predicting independent pumping tests. These results indicate that when constructing zonation models for aquifer characterization, accurate information will be required to construct zones for parameter estimations to achieve robust groundwater modeling results.
3. By comparing differently parameterized models, we find that the effective homogeneous model performs the worst in terms of model calibration and validation. The geology-based zonation models provide slightly improved calibration and validation results, even when inaccurate geological information is introduced. Geostatistical inverse models with spatially variable parameter fields yield the best results in terms of model calibration and validation. The comparison result emphasizes the importance of parameterization for aquifer characterization. For a HT survey when a large number of head data are available, highly parameterized models yield more accurate representations of aquifer heterogeneity as shown conclusively through our model validation results. On the other hand, with a

given amount of head data, the discretization of models used for inversion should be considered carefully in order to avoid the issue of over-parameterization.

4. In Case 2, the calibrated geology-based zonation model (with accurate geological data) provides slightly improved prediction results in comparison to the geostatistical model with homogeneous initial parameter fields, in terms of quantitative metrics (L_1 , L_2 , and R^2). This result suggests that when the number of pumping tests and observation ports are small, the inversion of zonation model with accurate geological information is able to yield comparable results to the geostatistical inverse modeling approach. In contrast, the inversion of zonation models based on inaccurate geological information provides worse prediction results when compared to the geostatistical inverse modeling case.
5. The impact of geological information on THT analysis of transient head data is further investigated by incorporating geological information of varying resolution and quality as prior information. When accurate geological information is incorporated, more structural features consistent with the known geology are revealed through the estimated K tomograms for both cases. Through the incorporation of inaccurate geological information, similar performance of models in terms of calibration and validation results is obtained for Case 1, which reveals that the impact of geological information for HT is slight when there are a large number of pumping tests and a dense network of observation intervals. For Case 2, the validation results are improved after incorporating GOOD and POOR2 geological information, while worse validation results are obtained when incorporating inaccurate geological information (POOR1 and POOR3). These results reveal that, when the number of pumping tests and observation intervals is limited, the incorporation of accurate or simplified geological information will help to reveal more

accurate heterogeneity, which in turn results in improved prediction results. The sandbox results provide important insights into field HT surveys and their interpretation. However, careful attention should be paid in obtaining more accurate geological data for including as prior information into inverse models, since inaccurate geological models will lead to adverse impacts on THT results, which will lead to poor groundwater flow models and prediction of heads.

6. We also compared the results from steady state and transient inversions of the same pumping test data. Forward simulations of 16 pumping tests conducted with the K tomogram obtained from the steady state inversion together with an estimate of S_s obtained from single-hole tests yielded significantly biased transient drawdown predictions. Therefore, our results suggest that in order to obtain accurate predictions of transient drawdowns from independent tests, K and S_s tomograms from THT analyses are necessary.
7. While the sandbox THT study is encouraging, the sandbox consists of sands of various sizes resulting in a low degree of S_s heterogeneity. Since the sandbox does not contain very low K materials (i.e., clays) representative of aquitards, we are not certain whether HT can reliably map such zones with pumping test data alone. We are currently conducting the THT analyses of multiple pumping tests at a highly heterogeneous site and examining the importance of utilizing long term pumping tests and geological data on such inversions. Results from this field-based study should yield important insights on the abilities for THT to map both aquifers and aquitards.

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). The second author acknowledges the support of “China Scholarship Council”. We thank the Associate Editor (Niklas Linde), Peter Bayer, and an anonymous reviewer for their helpful comments in improving the manuscript.

REFERENCES:

- Ahmed, A.S., Zhou, J., Jardani, A., Revil, A., Dupont, J., 2015. Image-guided inversion in steady-state hydraulic tomography. *Advances in Water Resources*, 82: 83-97.
- Berg, S.J., Illman, W.A., 2011a. Capturing aquifer heterogeneity: Comparison of approaches through controlled sandbox experiments. *Water Resour. Res.*, 47(9). DOI:10.1029/2011WR010429
- Berg, S.J., Illman, W.A., 2011b. Three-dimensional transient hydraulic tomography in a highly heterogeneous glaciofluvial aquifer-aquitard system. *Water Resour. Res.*, 47(10). DOI:10.1029/2011WR010616
- Berg, S.J., Illman, W.A., 2012. Improved predictions of saturated and unsaturated zone drawdowns in a heterogeneous unconfined aquifer via transient hydraulic tomography: Laboratory sandbox experiments. *J. Hydrol.*, 470–471: 172-183. DOI:<http://dx.doi.org/10.1016/j.jhydrol.2012.08.044>
- Berg, S.J., Illman, W.A., 2013. Field Study of Subsurface Heterogeneity with Steady-State Hydraulic Tomography. *Ground Water*, 51(1): 29-40. DOI:10.1111/j.1745-6584.2012.00914.x
- Berg, S.J., Illman, W.A., 2015. Comparison of Hydraulic Tomography with Traditional Methods at a Highly Heterogeneous Site. *Ground Water*, 53(1): 71-89. DOI:10.1111/gwat.12159

- Bohling, G.C., Zhan, X., Butler, J.J., Zheng, L., 2002. Steady shape analysis of tomographic pumping tests for characterization of aquifer heterogeneities. *Water Resour. Res.*, 38(12). DOI:10.1029/2001wr001176
- 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). DOI:10.1029/2006wr004932
- Brauchler, R., Liedl, R., Dietrich, P., 2003. A travel time based hydraulic tomographic approach. *Water Resour. Res.*, 39(12). DOI:10.1029/2003wr002262
- 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).
- 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 Resources Research*, 49(4).
- Cardiff, M., Barrash, W., Kitanidis, P., Malama, B., Revil, A., Straface, S., Rizzo, E., 2009. A potential-based inversion of unconfined steady-state hydraulic tomography. *Ground Water*, 47(2): 259-270.
- Cardiff, M., Barrash, W., 2011. 3-D transient hydraulic tomography in unconfined aquifers with fast drainage response. *Water Resour. Res.*, 47(12).
- 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).

- Cardiff, M., Barrash, W., Kitanidis, P.K., 2013. Hydraulic conductivity imaging from 3-D transient hydraulic tomography at several pumping/observation densities. *Water Resour. Res.*, 49(11): 7311-7326.
- Castagna, M., Becker, M.W., Bellin, A., 2011. Joint estimation of transmissivity and storativity in a bedrock fracture. *Water Resour. Res.*, 47(9). DOI:10.1029/2010wr009262
- Doherty, J. (Ed.), 2005. PEST Model-Independent Parameter Estimation User Manual, fifth ed. Watermark Numer. Comput., Brisbane, Australia.
- Hu, R., Brauchler, R., Herold, M., Bayer, P., 2011. Hydraulic tomography analog outcrop study: Combining travel time and steady shape inversion. *Journal of Hydrology*, 409(1): 350-362.
- Illman, W.A., Liu, X., Craig, A., 2007. Steady-state hydraulic tomography in a laboratory aquifer with deterministic heterogeneity: Multi-method and multiscale validation of hydraulic conductivity tomograms. *J. Hydrol.*, 341(3-4): 222-234. DOI:<http://dx.doi.org/10.1016/j.jhydrol.2007.05.011>
- Illman, W.A., Craig, A.J., Liu, X., 2008. Practical issues in imaging hydraulic conductivity through hydraulic tomography. *Ground Water*, 46(1): 120-132.
- Illman, W.A., Liu, X., Takeuchi, S., Yeh, T.-C.J., Ando, K., Saegusa, H., 2009. Hydraulic tomography in fractured granite: Mizunami Underground Research site, Japan. *Water Resour. Res.*, 45(1). DOI:10.1029/2007WR006715
- 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): n/a-n/a. DOI:10.1029/2009WR007745
- Illman, W.A., Berg, S.J., Yeh, T.C.J., 2012. Comparison of approaches for predicting solute transport: Sandbox experiments. *Ground Water*, 50(3): 421-431.

- Illman, W.A., Berg, S.J., Zhao, Z., 2015. Should hydraulic tomography data be interpreted using geostatistical inverse modeling? A laboratory sandbox investigation. *Water Resour. Res.*, 51(5): 3219-3237. DOI:10.1002/2014WR016552
- Jiménez, S., Brauchler, R., Bayer, P., 2013. A new sequential procedure for hydraulic tomographic inversion. *Advances in water resources*, 62: 59-70.
- 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 Resources Research*, 51(7): 5504-5520.
- Kitanidis, P.K., 1995. Quasi-Linear Geostatistical Theory for Inversing. *Water Resour. Res.*, 31(10): 2411-2419. DOI:10.1029/95WR01945
- Liu, S., Yeh, T.C.J., Gardiner, R., 2002. Effectiveness of hydraulic tomography: Sandbox experiments. *Water Resour. Res.*, 38(4).
- 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). DOI:10.1029/2006wr005144
- Mao, D., Yeh, T.C.J., Wan, L., Wen, J.C., Lu, W., Lee, C.H., Hsu, K.C., 2013. Joint interpretation of sequential pumping tests in unconfined aquifers. *Water Resources Research*, 49(4): 1782-1796.
- 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.
- Schöniger, A., Illman, W.A., Wöhling, T., Nowak, W., 2015. Finding the right balance between groundwater model complexity and experimental effort via Bayesian model selection. *Journal of Hydrology*, 531: 96-110.
- Straface, S., Yeh, T.C., Zhu, J., Troisi, S., Lee, C., 2007. Sequential aquifer tests at a well field, Montalto Uffugo Scalo, Italy. *Water Resour. Res.*, 43(7).

- Wu, C.-M., Yeh, T.-C.J., Zhu, J., Lee, T.H., Hsu, N.-S., Chen, C.-H., Sancho, A.F., 2005. Traditional analysis of aquifer tests: Comparing apples to oranges? *Water Resour. Res.*, 41(9). DOI: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). DOI:10.1029/2008WR007180
- Yeh, T.C.J., Srivastava, R., Guzman, A., Harter, T., 1993. A numerical model for water flow and chemical transport in variably saturated porous media. *Ground Water*, 31(4): 634-644.
- Yeh, T.C.J., Liu, S., 2000. Hydraulic tomography: Development of a new aquifer test method. *Water Resour. Res.*, 36(8): 2095-2105. DOI:10.1029/2000wr900114
- Zha, Y., Yeh, T.-C.J., Illman, W.A., Tanaka, T., Bruines, P., Onoe, H., Saegusa, H., 2015. What does hydraulic tomography tell us about fractured geological media? A field study and synthetic experiments. *J. Hydrol.*, 531, Part 1: 17-30. DOI:<http://dx.doi.org/10.1016/j.jhydrol.2015.06.013>
- Zha, Y., Yeh, T.C.J., Illman, W.A., Tanaka, T., Bruines, P., Onoe, H., Saegusa, H., Mao, D., Takeuchi, S., Wen, J.C., 2016. An Application of Hydraulic Tomography to a Large-Scale Fractured Granite Site, Mizunami, Japan. *Ground Water*, 54(6): 793-804.
- Zhao, Z., Illman, W.A., Yeh, T.-C.J., Berg, S.J., Mao, D., 2015. Validation of hydraulic tomography in an unconfined aquifer: A controlled sandbox study. *Water Resour. Res.*, 51(6): 4137-4155. DOI:10.1002/2015WR016910
- Zhao, Z., Illman, W.A., Berg, S.J., 2016. On the importance of geological data for hydraulic tomography analysis: laboratory sandbox study. *J. Hydrol.* DOI:10.1016/j.jhydrol.2016.08.061
- Zhao, Z., Illman, W.A., 2017. On the Importance of Geological Data for Three-dimensional Steady-State Hydraulic Tomography Analysis at a Highly Heterogeneous Aquifer-Aquitard System. *J. Hydrol.*

Zhou, J., Revil, A., Karaoulis, M., Hale, D., Doetsch, J., Cuttler, S., 2014. Image-guided inversion of electrical resistivity data. *Geophys. J. Int.*, 197(1): 292-309.

Zhu, J., Yeh, T.-C.J., 2005. Characterization of aquifer heterogeneity using transient hydraulic tomography. *Water Resour. Res.*, 41(7). DOI:10.1029/2004WR003790

Zhu, J., Yeh, T.-C.J., 2006. Analysis of hydraulic tomography using temporal moments of drawdown recovery data. *Water Resour. Res.*, 42(2). DOI:10.1029/2005WR004309

ACCEPTED MANUSCRIPT

Figure Captions

Fig. 1: Photograph of synthetic heterogeneous aquifer showing the layer (black) and port (blue) numbers (modified after Illman et al., 2010). Red circles indicate the 48 ports installed in the aquifer.

Fig. 2: Geological models with various accuracy and resolution: (a) GOOD; (b) POOR1; (c) POOR2; (d) POOR3.

Fig. 3: K and S_s tomograms estimated from geology-based zonation models for Case 1. K tomograms: (a) GOOD; (c) POOR1; (e) POOR2; (g) POOR3. S_s tomograms: (b) GOOD; (d) POOR1; (f) POOR2; (h) POOR3.

Fig. 4: K and S_s tomograms estimated from geology-based zonation models for Case 2. K tomograms: (a) GOOD; (c) POOR1; (e) POOR2; (g) POOR3. S_s tomograms: (b) GOOD; (d) POOR1; (f) POOR2; (h) POOR3.

Fig. 5: Estimated K values and corresponding 95% confidence intervals of Cases 1 and 2 for four different geology-based zonation models: a) GOOD; b) POOR1; c) POOR2; d) POOR3.

Fig. 6: Estimated S_s values and corresponding 95% confidence intervals of Cases 1 and 2 for four different geology-based zonation models: a) GOOD; b) POOR1; c) POOR2; d) POOR3.

Fig. 7: K and S_s tomograms and their corresponding variances for Case 1 with homogeneous initial K and S_s fields. (a) K tomogram, (b) the corresponding $\ln K$ variance, (c) S_s tomogram, (d) the corresponding $\ln S_s$ variance. Black lines in the K tomogram represent the accurate stratification of the synthetic aquifer.

Fig. 8: K and S_s tomograms and their corresponding variances for Case 2 with homogeneous initial K and S_s fields. (a) K tomogram, (b) the corresponding $\ln K$ variance, (c) S_s tomogram, (d) the corresponding $\ln S_s$ variance. Black lines in the K tomogram represent the accurate stratification of the synthetic aquifer.

Fig. 9: K and S_s tomograms from geostatistical models with heterogeneous initial K and S_s fields for Case 1. K tomograms: (a) GOOD; (c) POOR1; (e) POOR2; (g) POOR3. S_s tomograms: (b) GOOD; (d) POOR1; (f) POOR2; (h) POOR3. Black lines in (a) represent the accurate stratification of the synthetic aquifer.

Fig. 10: K and S_s tomograms from geostatistical models with heterogeneous initial K and S_s fields for Case 2. K tomograms: (a) GOOD; (c) POOR1; (e) POOR2; (g) POOR3. S_s tomograms: (b) GOOD; (d) POOR1; (f) POOR2; (h) POOR3. Black lines in (a) represent the accurate stratification of the synthetic aquifer.

Fig. 11: Validation scatterplots (Case 1) of simulated versus observed drawdowns for different modeling approaches. (a) effective model, (b)-(e) four geology-based zonation models: (b) GOOD, (c) POOR1, (d) POOR2, (e) POOR3; and (f) geostatistical model with homogeneous initial parameter fields.

Fig. 12: Validation scatterplots (Case 1) of simulated versus observed drawdowns for geostatistical models incorporated with four different types of geological information. (a) GOOD, (b) POOR1, (c) POOR2, and (d) POOR3.

Fig. 13: Validation scatterplots (Case 2) of simulated versus observed drawdowns for different modeling approaches. (a) effective model, (b)-(e) four geology-based zonation models: (b)

GOOD, (c) POOR1, (d) POOR2, (e) POOR3; and (f) geostatistical model with homogeneous initial parameter fields.

Fig. 14: Validation scatterplots (Case 2) of simulated versus observed drawdowns for geostatistical models incorporated with four different types of geological information. (a) GOOD, (b) POOR1, (c) POOR2, and (d) POOR3.

Fig. 15: Prediction of drawdown curves at 16 selected ports when conducting pumping test at port 40. K and S_s tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

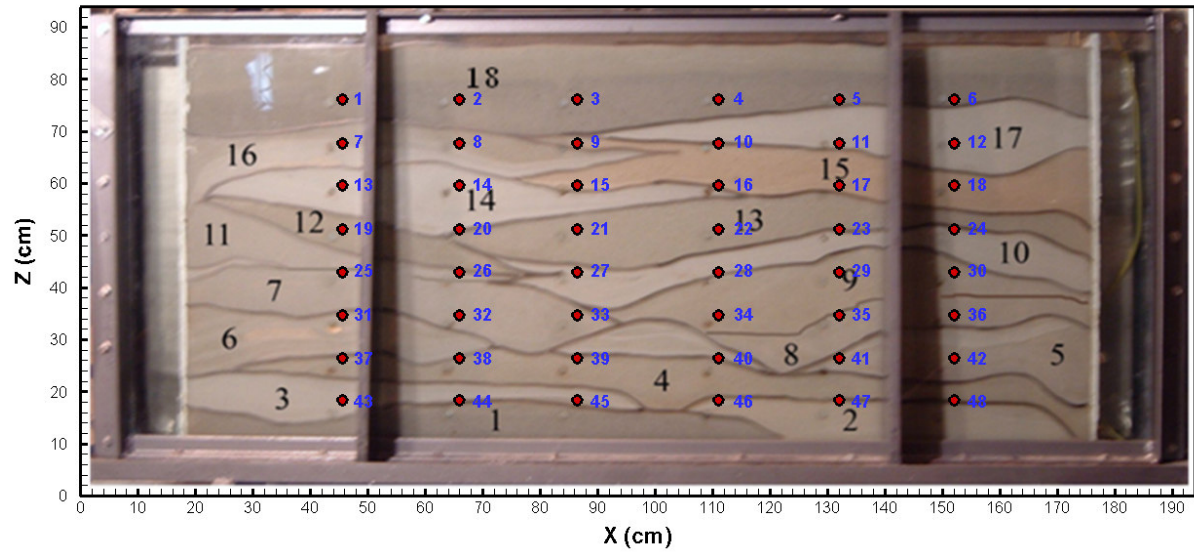
Fig. 16: Prediction of drawdown curves at 16 selected ports when conducting pumping test at port 40. K and S_s tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from eight pumping tests and 47 observation ports (Case 1).

Fig. 17: Prediction of drawdown curves at 16 selected ports when conducting pumping test at port 40. K and S_s tomograms are obtained from different modeling approaches with four pumping tests and 15 observation ports (Case 2).

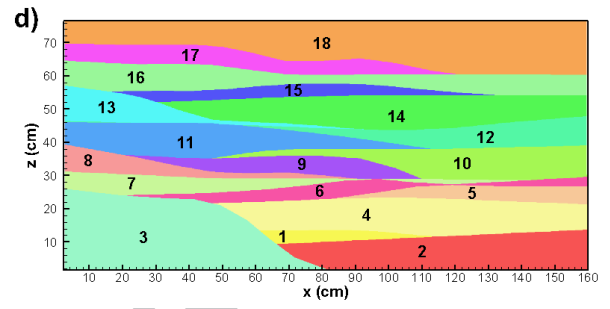
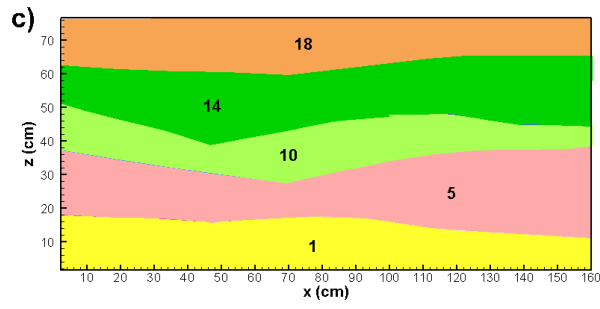
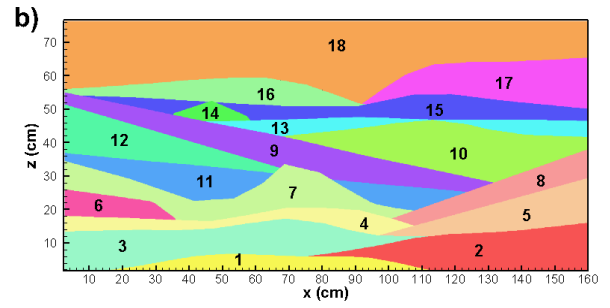
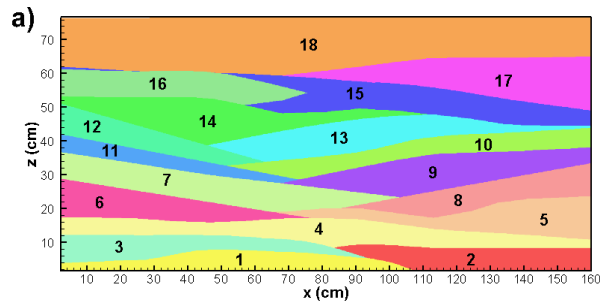
Fig. 18: Prediction of drawdown curves at 16 selected ports when conducting pumping test at port 40. K and S_s tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from four pumping tests and 15 observation ports (Case 2).

Fig. 19: Validation scatterplots of simulated versus observed transient drawdowns utilizing K tomograms obtained from SSHT coupled with the geometric mean of 48 S_y values ($S_y = 6.1 \times 10^{-4}$ /cm) obtained from single-hole tests. (a) Case 1, (b) Case 2.

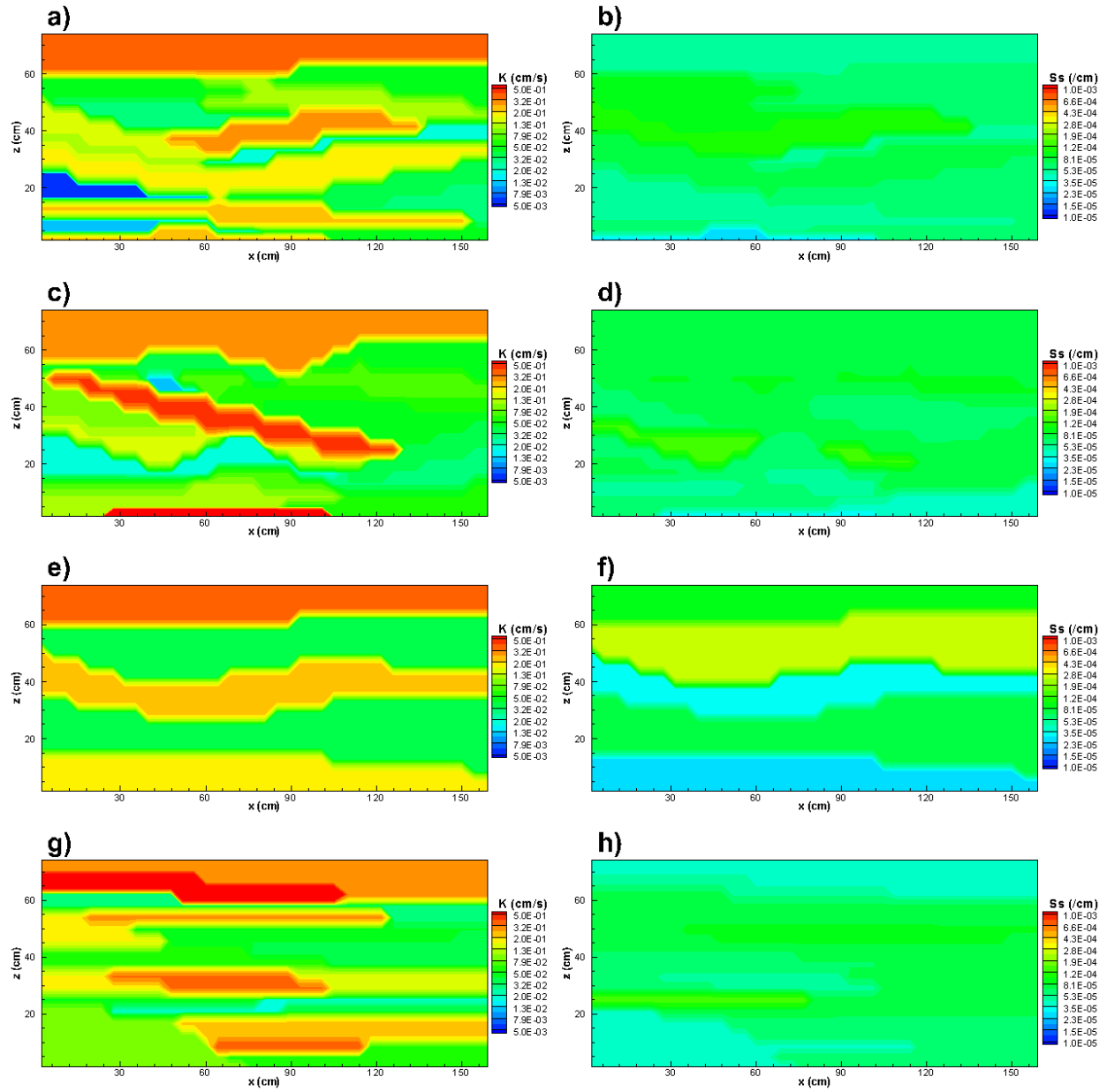
ACCEPTED MANUSCRIPT



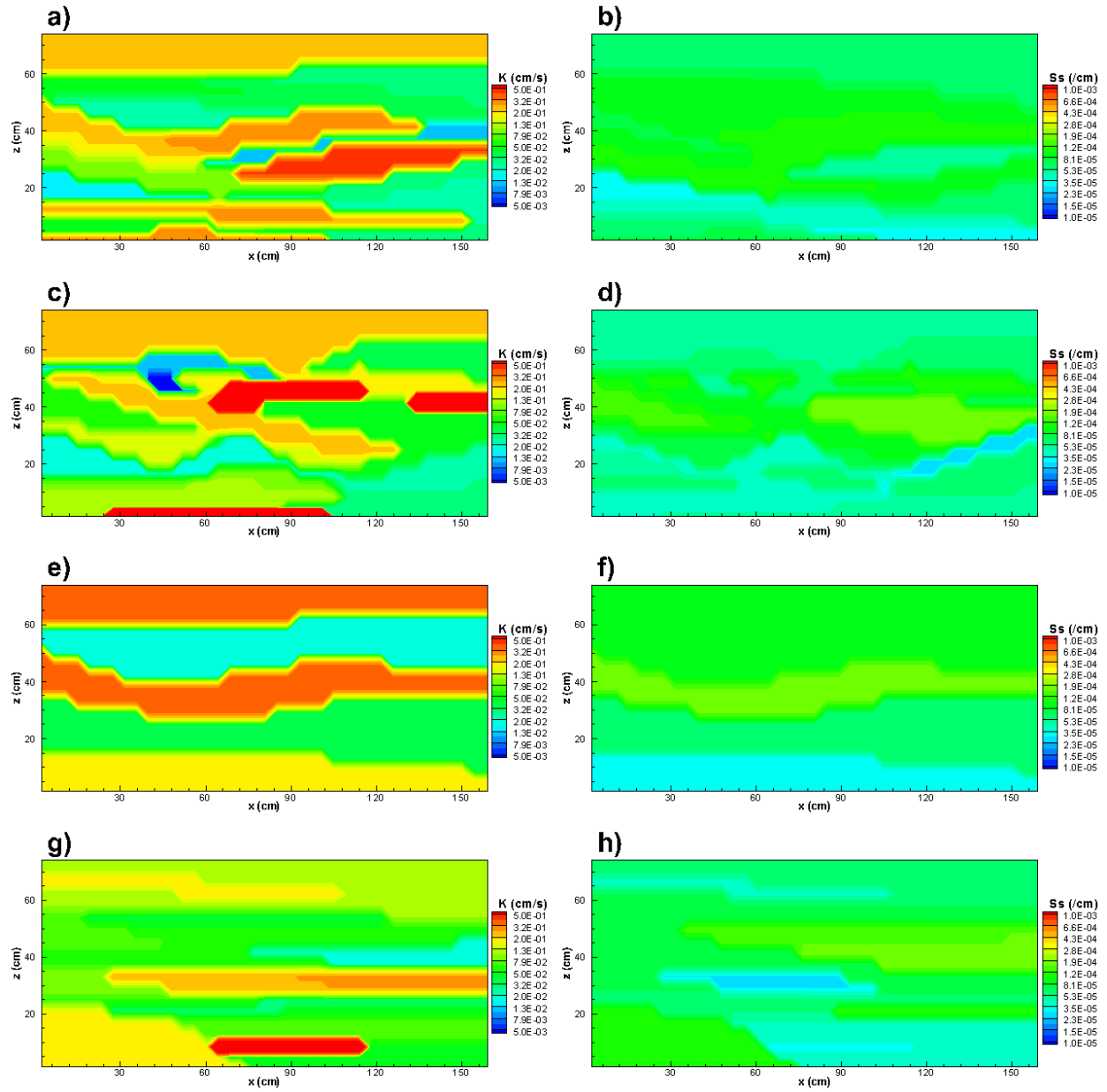
ACCEPTED MANUSCRIPT



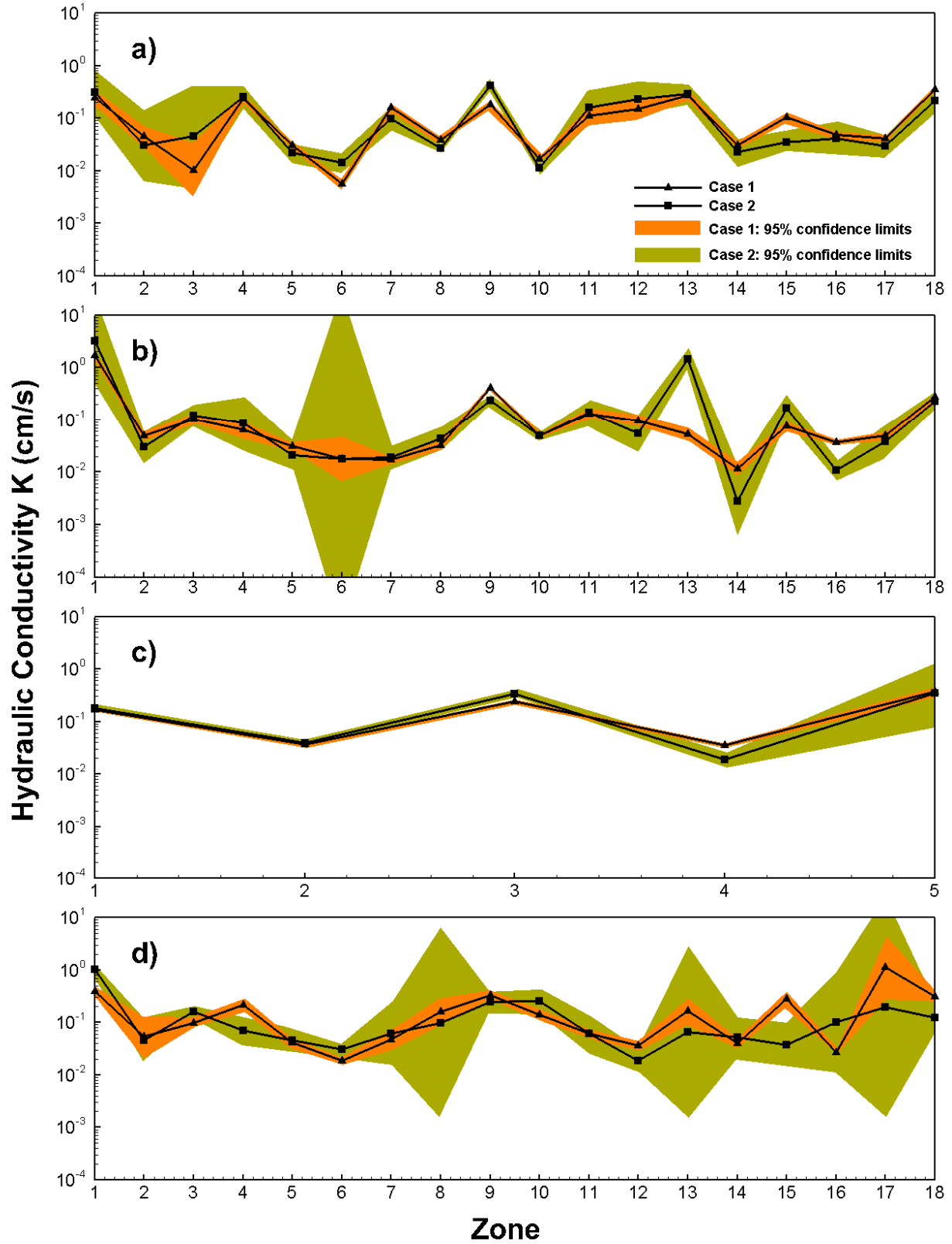
ACCEPTED MANUSCRIPT

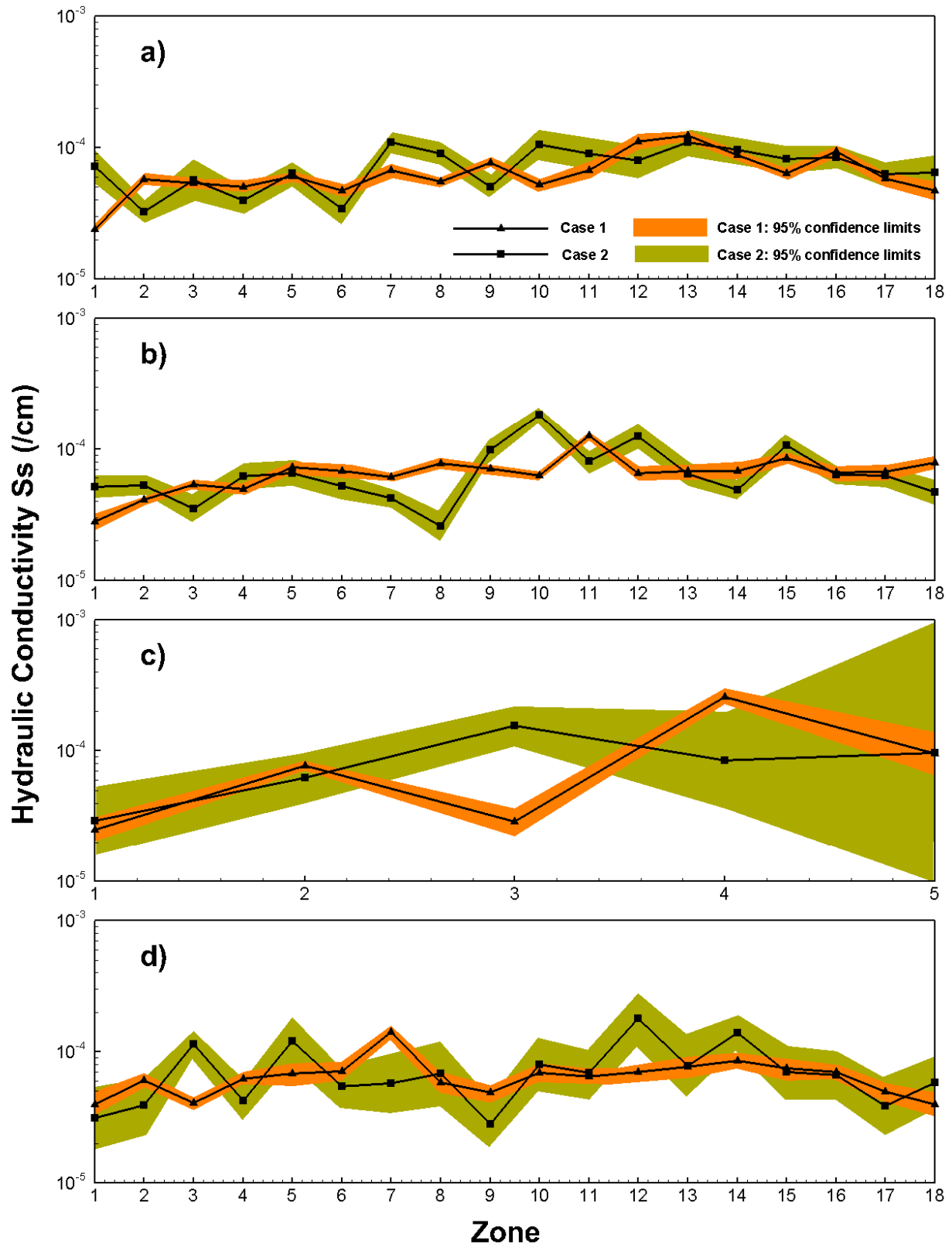


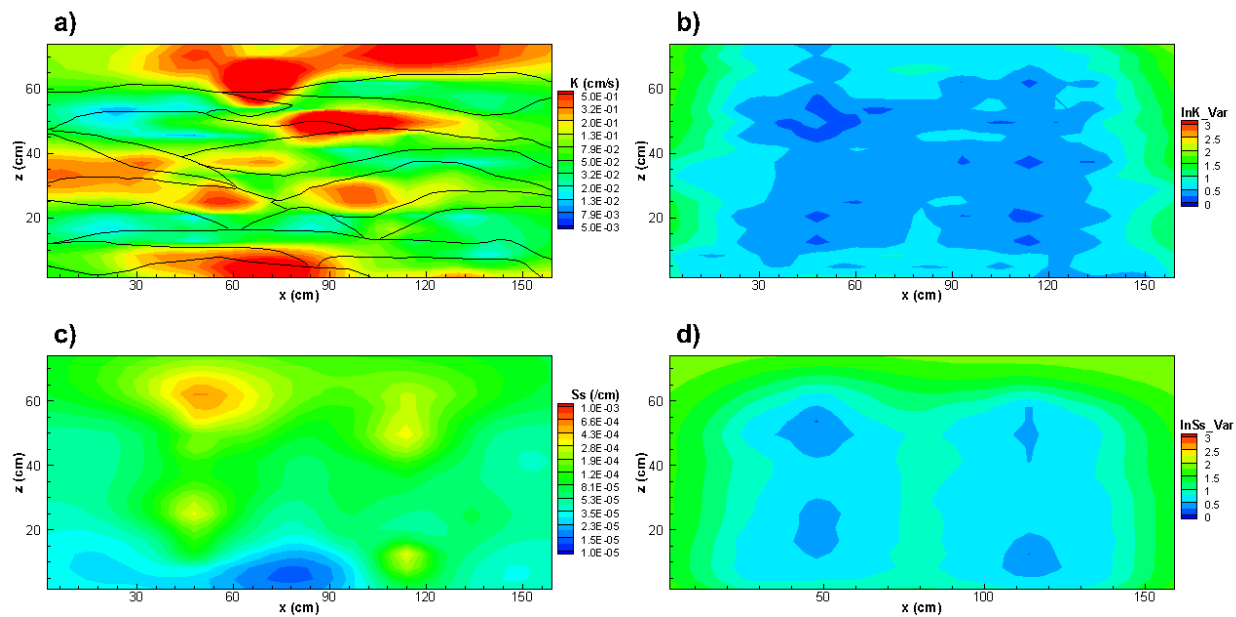
ACC



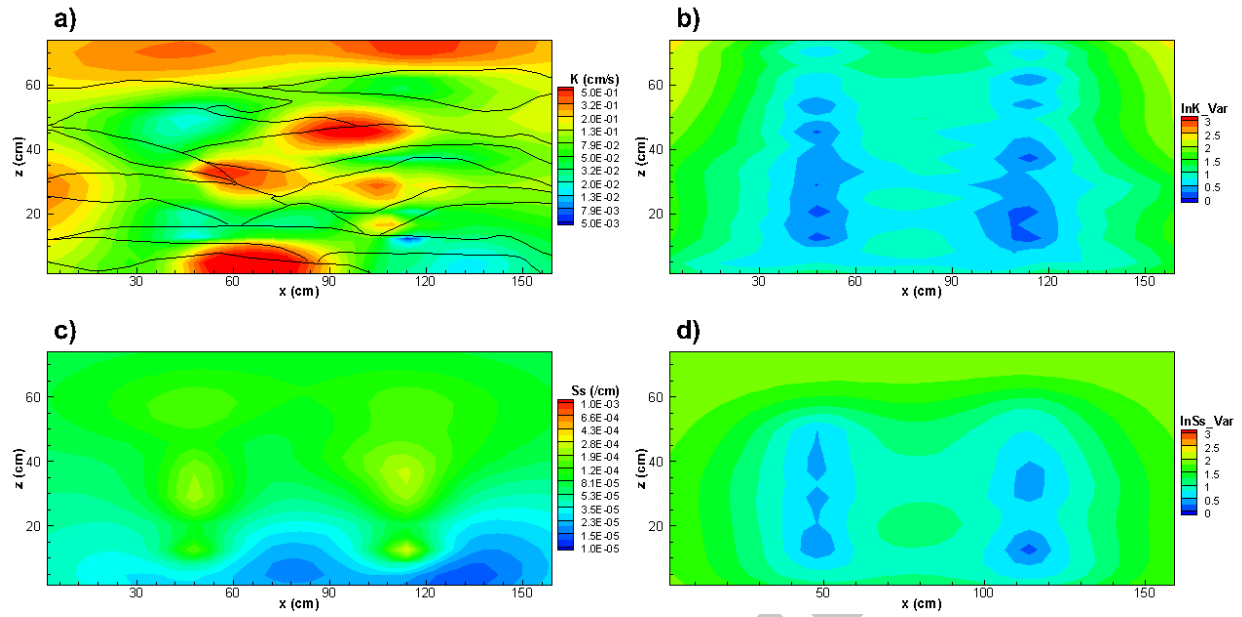
ACC



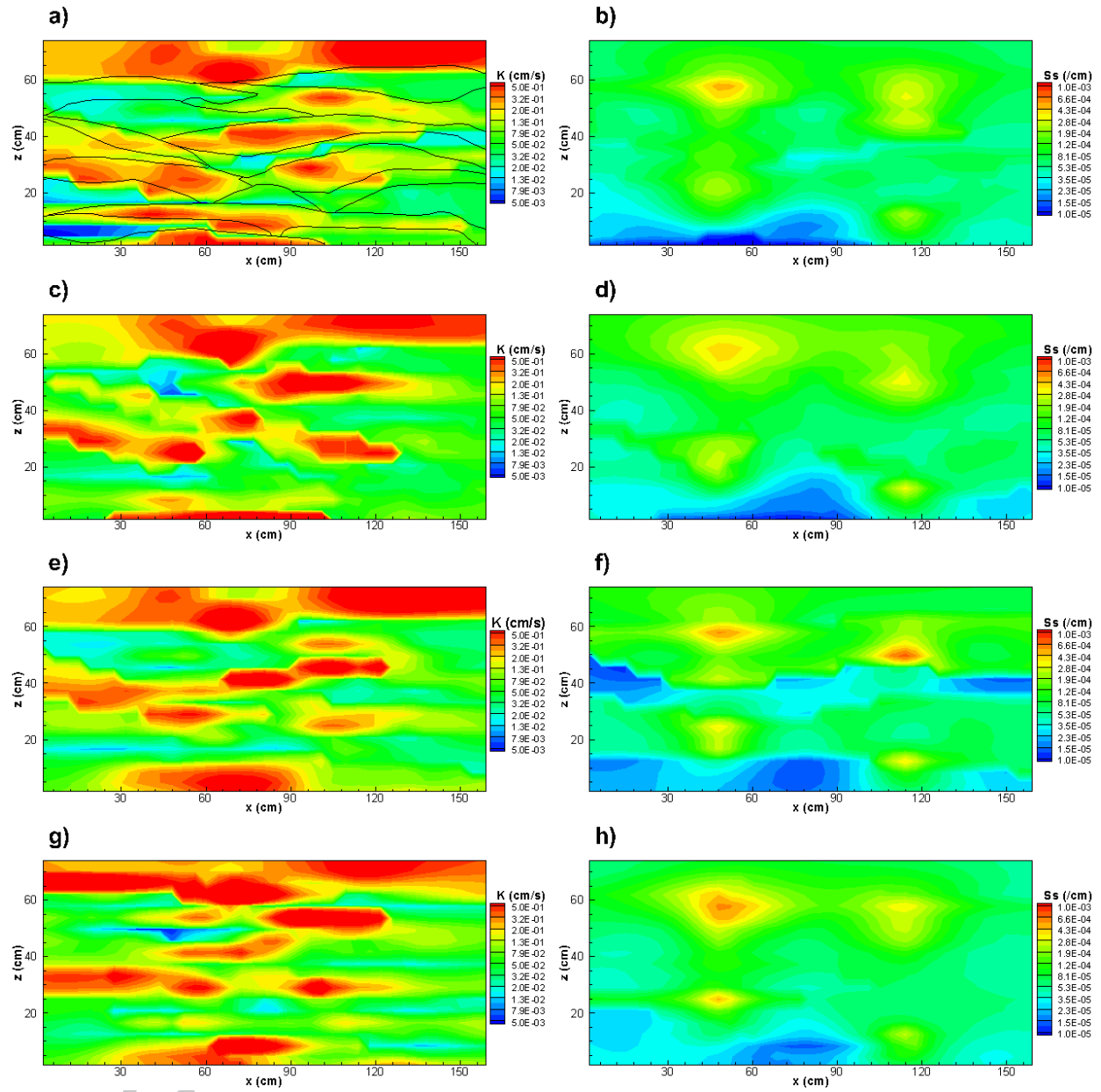




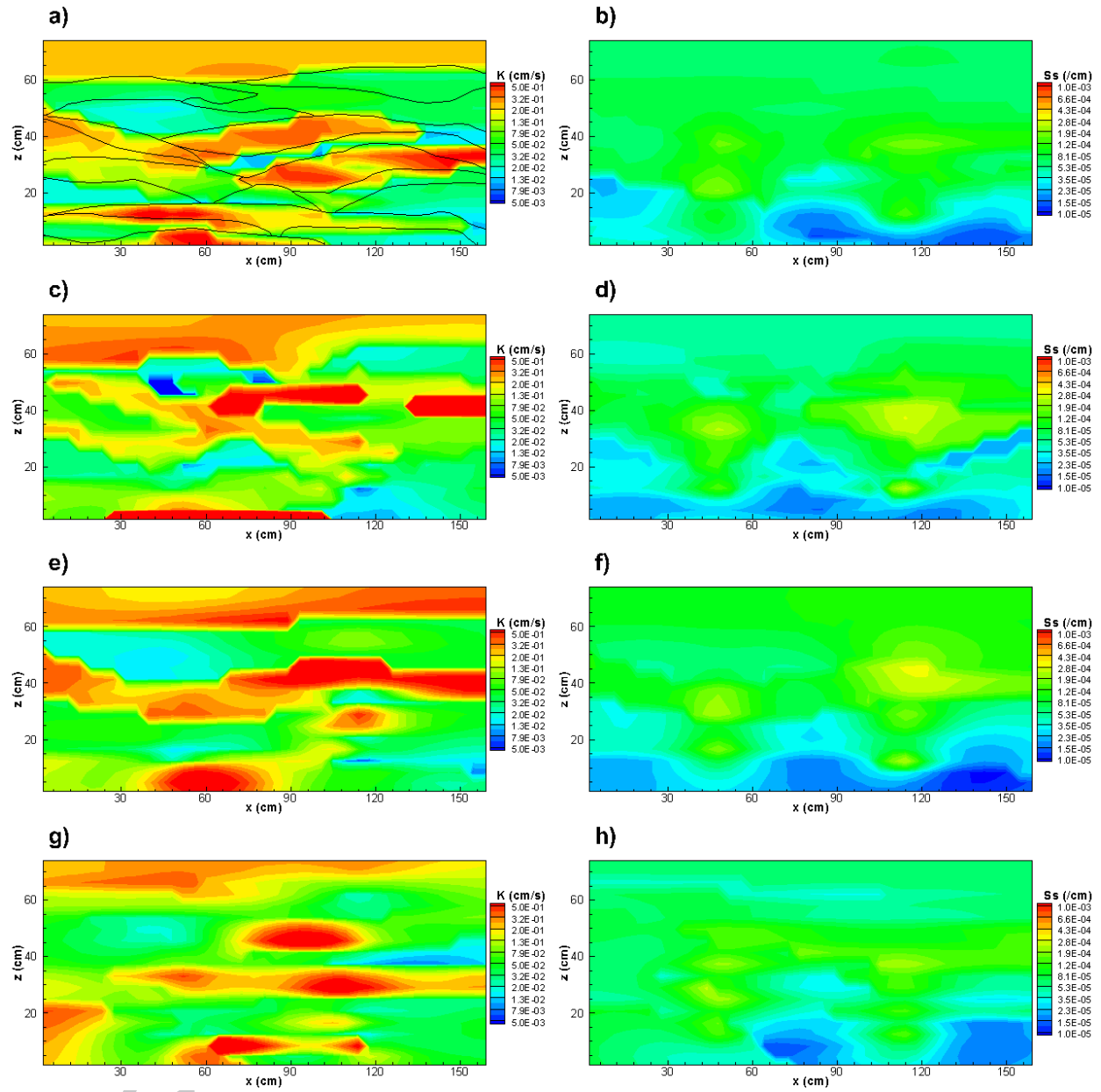
ACCEPTED MANUSCRIPT

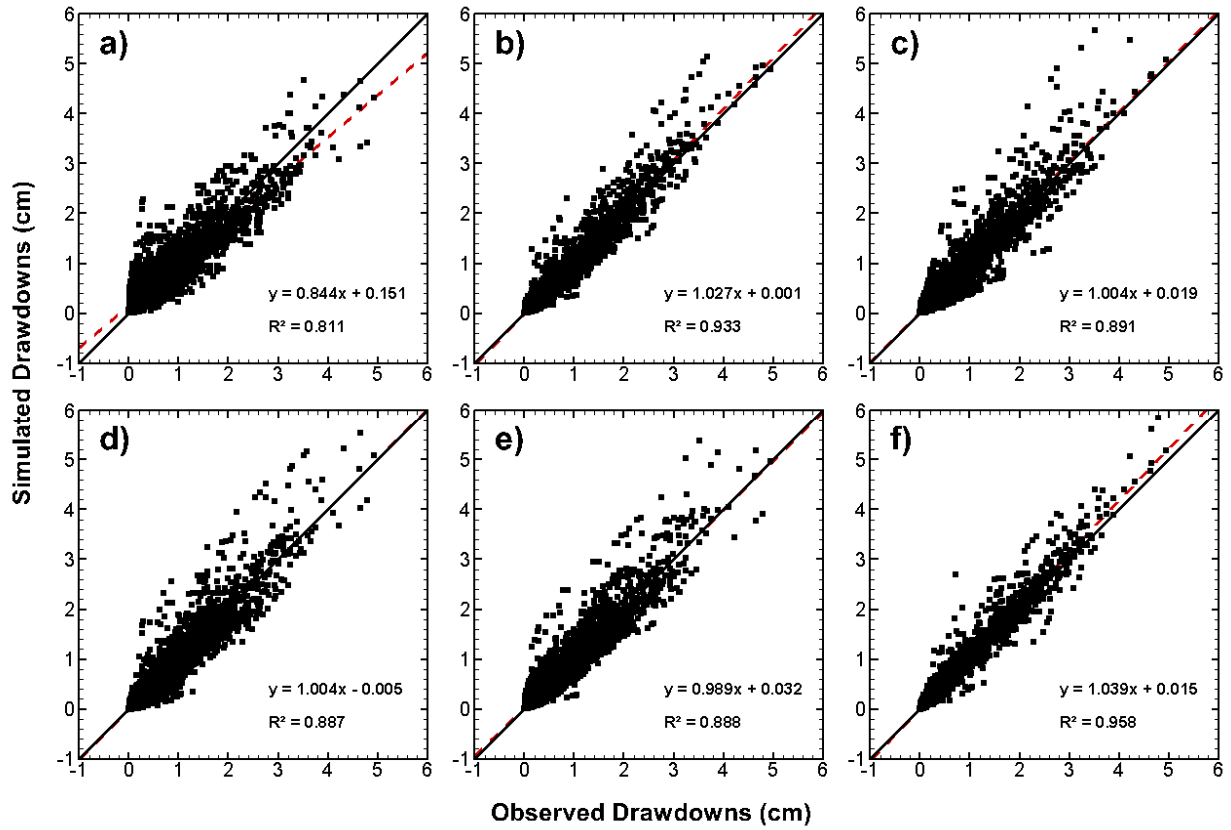


ACCEPTED MANUSCRIPT

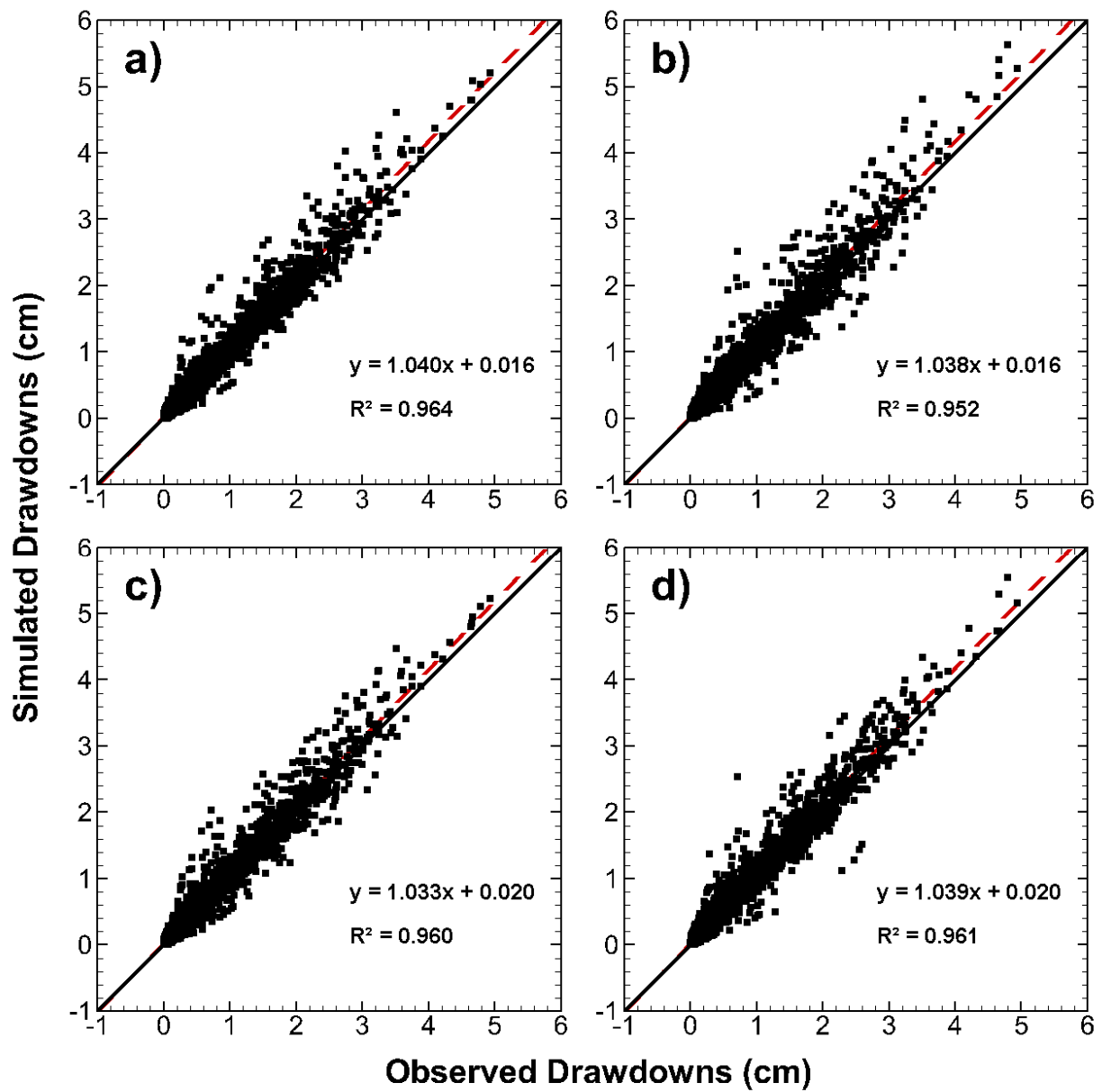


ACC

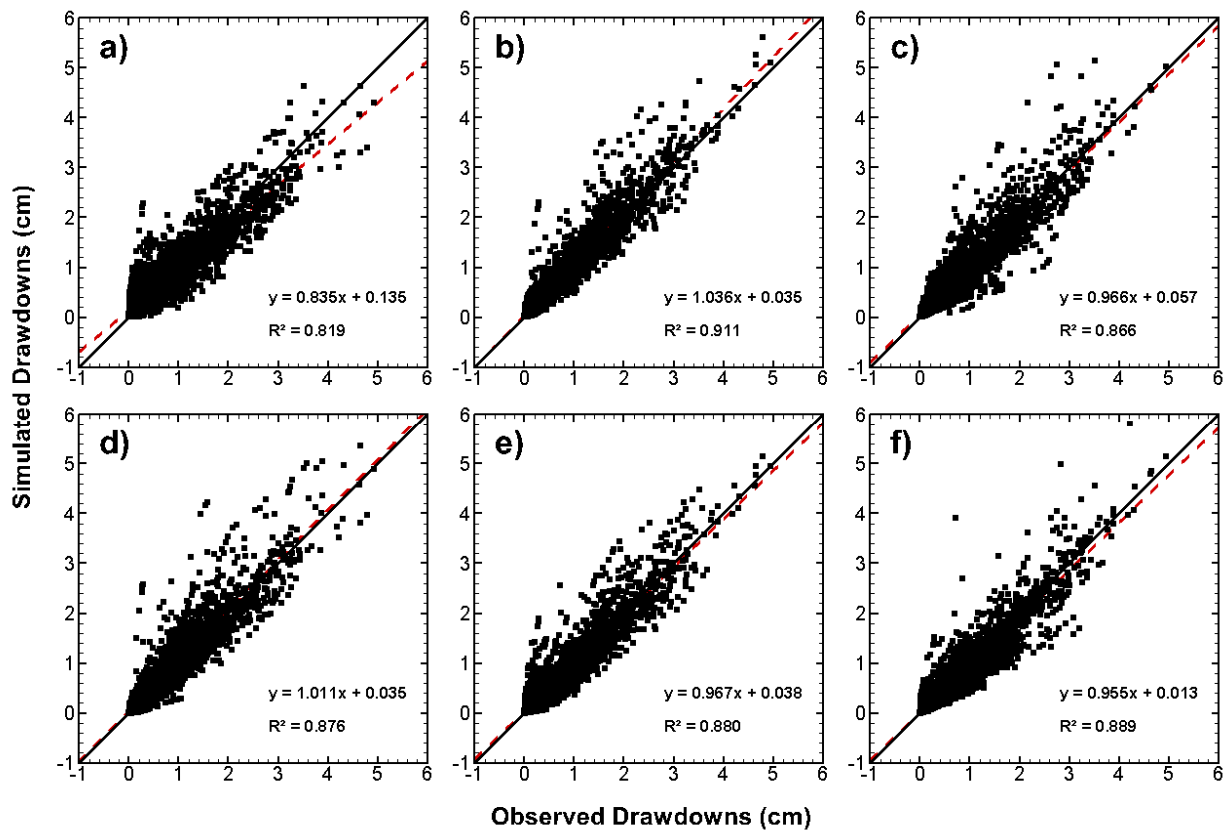




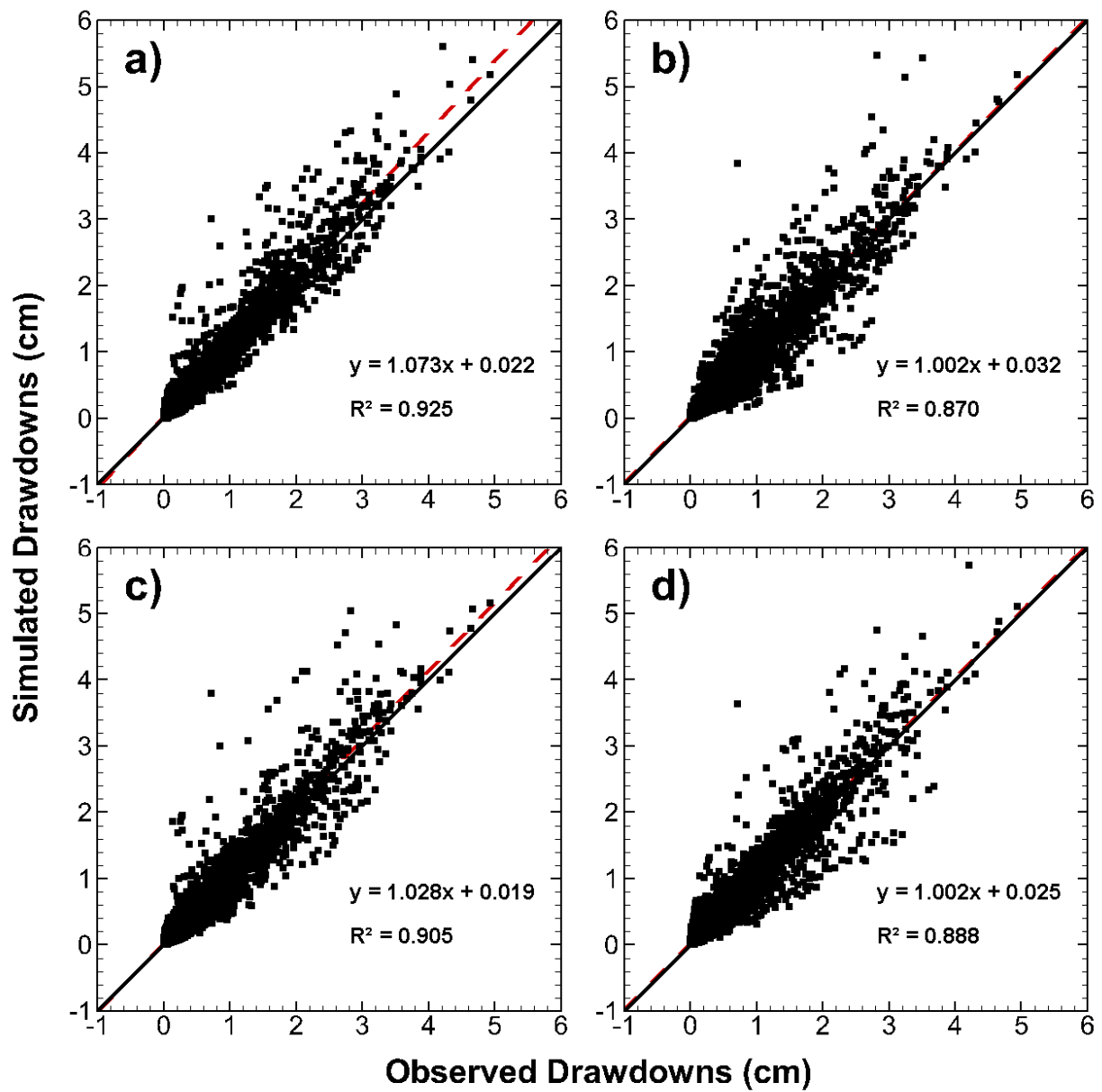
ACCEPTED



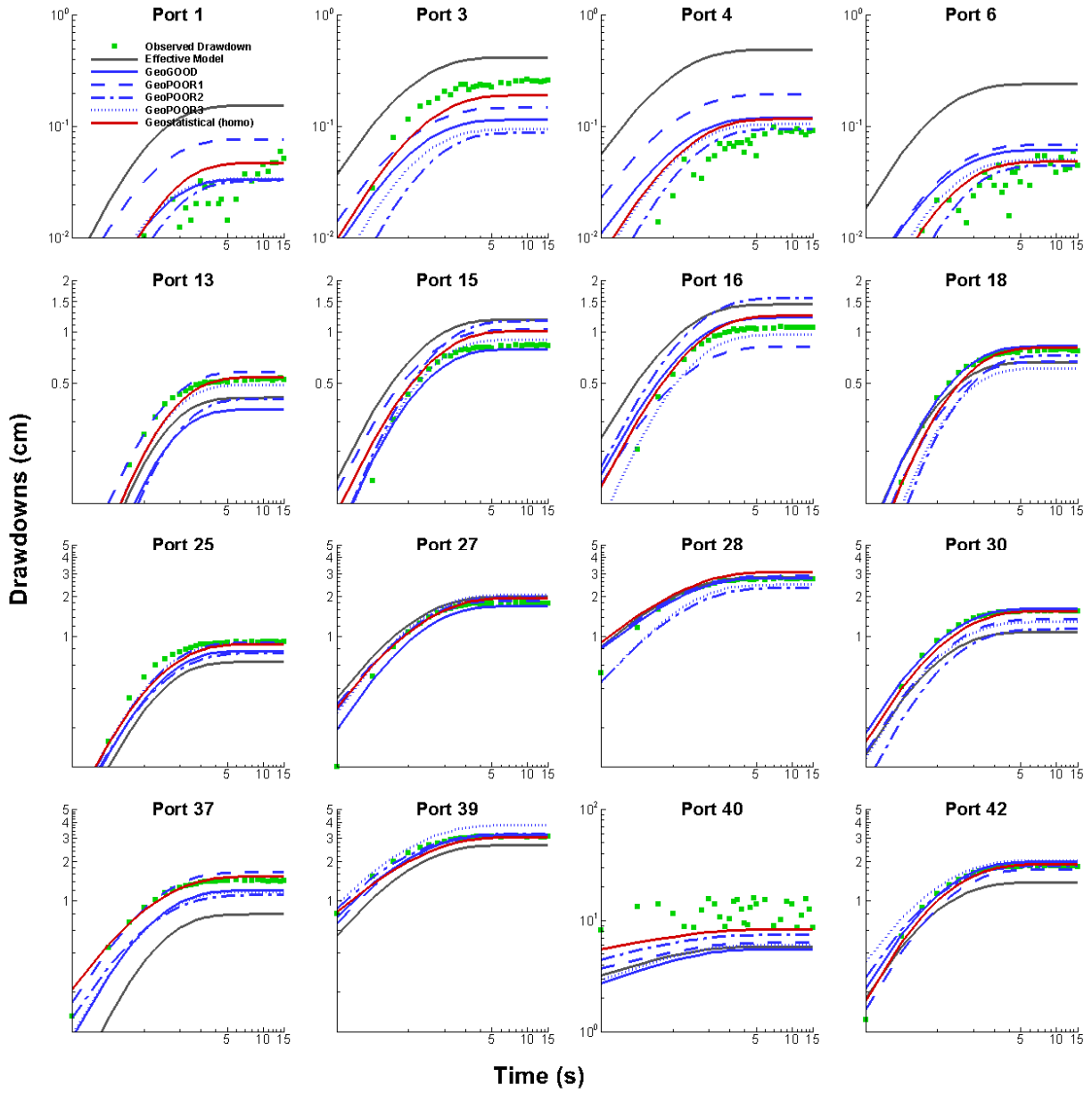
ACC



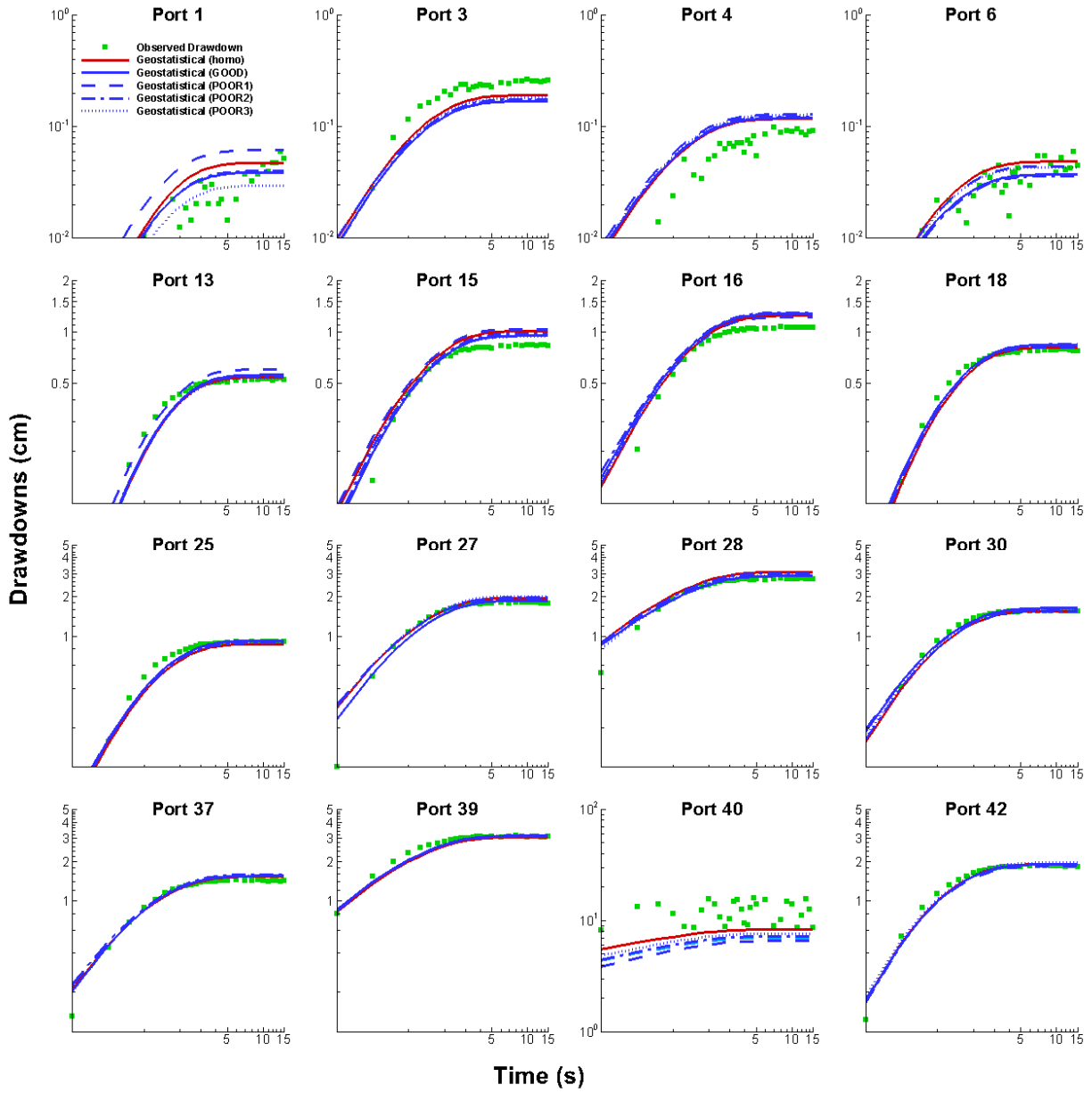
ACCEPTED



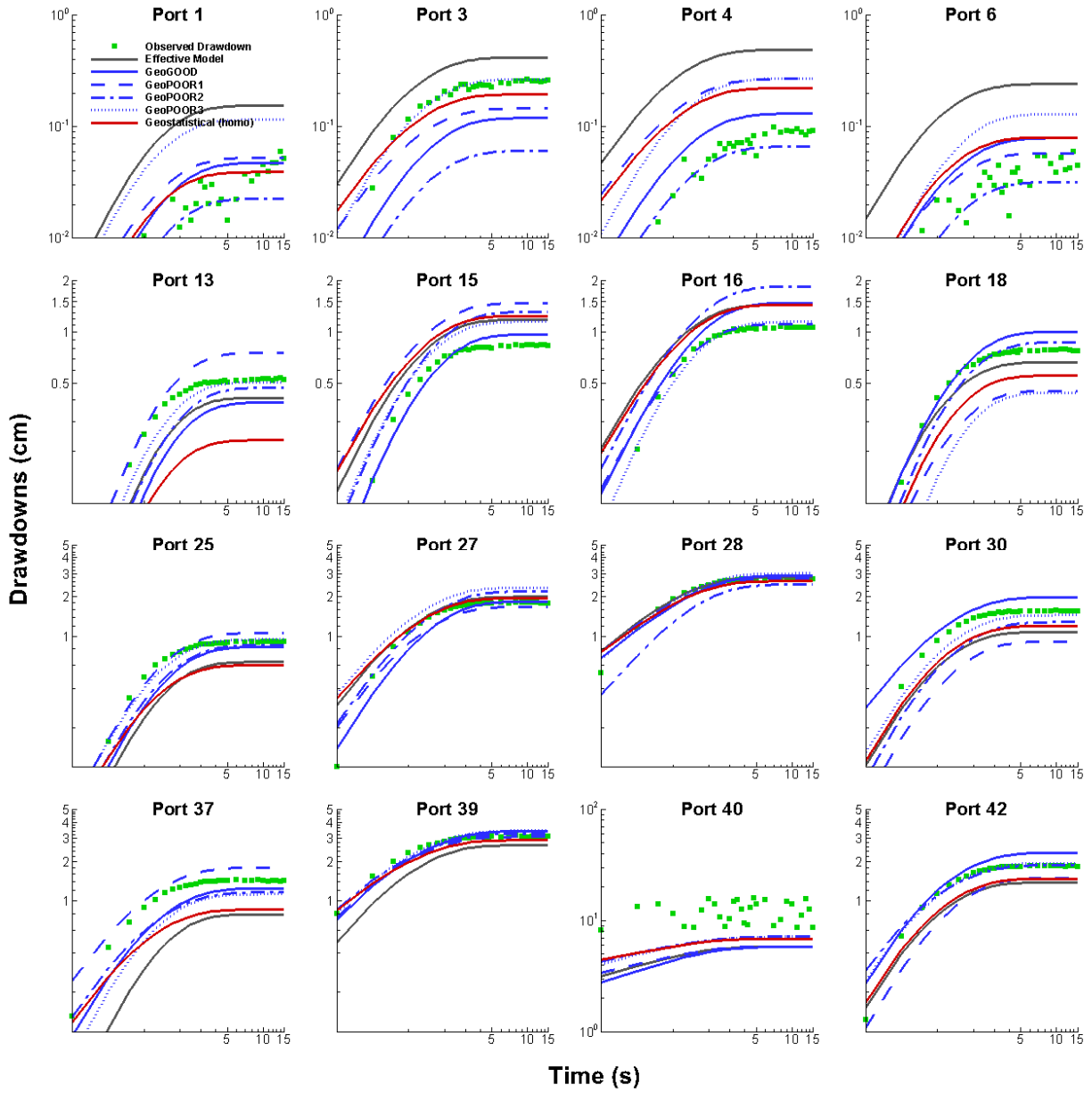
ACC



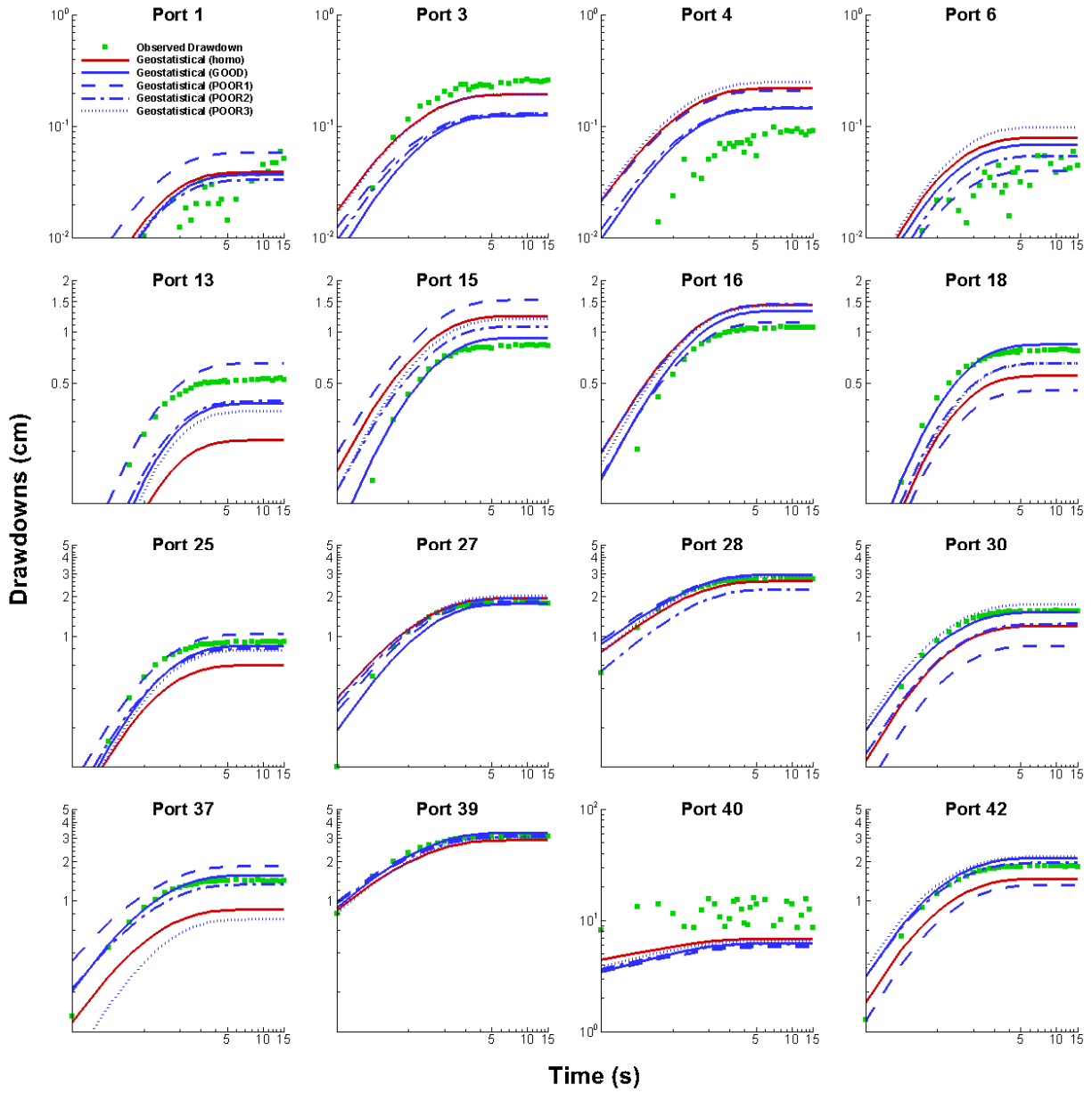
ACC



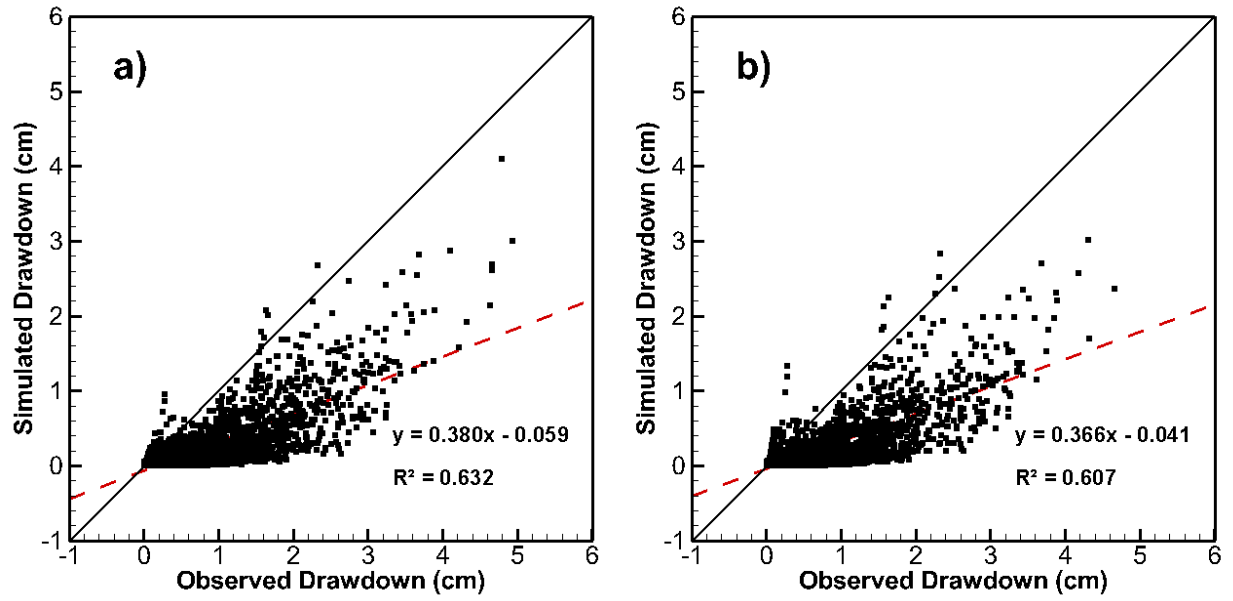
ACC



ACC



ACC



ACCEPTED MANUSCRIPT

Table 1 - Marked: Sand type, d_{50} , K and S_s estimates for each deposited layer in the synthetic heterogeneous aquifer^a.

Layer	Sand	d_{50} (mm)	K (cm/s) Shepherd	Core Permeameter K^b (cm/s)	Single-Hole K^b (cm/s)	Single-Hole S_s^b (/s)
1	20/30	0.75	1.03×10^{-1}	3.20×10^{-2}	5.32×10^{-2}	2.12×10^{-4}
2	4030	0.35	2.99×10^{-2}	5.29×10^{-2}	5.67×10^{-2}	2.60×10^{-4}
3	F-85	0.15	7.28×10^{-3}	7.14×10^{-2}	5.70×10^{-2}	5.00×10^{-4}
4	20/40	0.58	6.68×10^{-2}	5.68×10^{-2}	5.10×10^{-2}	2.22×10^{-4}
5	mix	0.46	N/A	N/A	N/A	N/A
6	mix	0.46	N/A	8.16×10^{-2}	5.00×10^{-2}	4.00×10^{-4}
7	#12	0.52	5.70×10^{-2}	1.27×10^{-1}	7.35×10^{-2}	4.20×10^{-4}
8	F32	0.5	5.33×10^{-2}	1.34×10^{-1}	4.50×10^{-2}	1.75×10^{-4}
9	20/40	0.58	6.68×10^{-2}	8.69×10^{-2}	4.60×10^{-2}	2.15×10^{-4}
10	F-65	0.2	1.20×10^{-2}	1.13×10^{-1}	8.25×10^{-2}	1.14×10^{-3}
11	#12	0.52	5.70×10^{-2}	1.37×10^{-1}	2.05×10^{-1}	2.15×10^{-4}
12	16/30	0.87	1.32×10^{-1}	3.40×10^{-2}	4.95×10^{-2}	6.32×10^{-4}
13	20/30	0.75	1.03×10^{-1}	2.60×10^{-1}	1.05×10^{-1}	9.80×10^{-4}
14	f-75	0.17	9.22×10^{-3}	9.79×10^{-2}	5.70×10^{-2}	9.80×10^{-4}
15	20/40	0.58	6.68×10^{-2}	8.58×10^{-2}	7.50×10^{-2}	2.00×10^{-3}
16	mix	0.46	N/A	4.16×10^{-2}	2.68×10^{-2}	7.11×10^{-4}
17	F-85	0.15	7.29×10^{-3}	4.51×10^{-2}	4.47×10^{-2}	1.14×10^{-3}
18	20/30	0.75	1.03×10^{-1}	1.45×10^{-1}	1.16×10^{-1}	3.38×10^{-3}

^a Data from Illman et al. (2010).

^b If multiple ports are in the same layer then the geometric mean is presented.

Table 2 - Marked: Summary of L_1 and L_2 norms of calibration and validation results for Cases 1 and 2.

		Case 1				Case 2			
		Calibration		Validation		Calibration		Validation	
		L_1	L_2	L_1	L_2	L_1	L_2	L_1	L_2
Effective Parameter Model		0.250	0.124	0.223	0.096	0.317	0.167	0.220	0.095
Geology-based Zonation Model	GOOD	0.102	0.027	0.118	0.038	0.079	0.016	0.139	0.059
	POOR1	0.112	0.029	0.152	0.061	0.096	0.020	0.168	0.077
	POOR2	0.130	0.052	0.151	0.064	0.179	0.090	0.156	0.077
	POOR3	0.135	0.046	0.159	0.061	0.159	0.043	0.175	0.067
Geostatistical Model with Homogeneous Initial Fields		0.046	0.005	0.091	0.026	0.042	0.004	0.149	0.061
Geostatistical Model with Heterogeneous Initial Fields	GOOD	0.050	0.006	0.087	0.023	0.048	0.006	0.128	0.057
	POOR1	0.051	0.005	0.096	0.030	0.050	0.006	0.159	0.080
	POOR2	0.050	0.005	0.090	0.024	0.048	0.005	0.129	0.060
	POOR3	0.050	0.005	0.091	0.025	0.047	0.005	0.149	0.067

ACCEPTED

Highlights:

1. THT is necessary in accurate estimating both K and S_y heterogeneities.
2. Geostatistics-based inverse models performs the best when there are abundant data.
3. With limited data, zonation-based models perform close to geostatistics-based ones.
4. Incorporation of accurate geological information improves THT results.

ACCEPTED MANUSCRIPT