Comparison of Hydraulic Tomography Survey with and without Cut-off Walls at the Narashino Site, Japan

by

Zeren Ning

A thesis

presented to the University of Waterloo

in the fulfillment of the

thesis requirement for the degree of

Master of Science

in

Earth Science

Waterloo, Ontario, Canada, 2021

Author’ s Declaration

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

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

Abstract

This study investigates the performance of several different conceptualizations of inverse modeling including the effective parameter, geological zonation and geostatistical inversion approaches, on characterizing the hydraulic parameter fields, such as hydraulic conductivity and specific storage, at the Narashino Site in Chiba Prefecture, Japan. The simulation of effective parameter and geological zonation approaches are achieved by coupling HydroGeoSphere (HGS) and model independent parameter estimation code PEST using head records during the step-up pumping/injection tests, while the simulation of geostatistical inversion approach is achieved by using the sequential successive linear estimator with the loop iteration scheme (SSLE_loop) to invert the hydraulic tomography (HT) test data collected during 2019 and 2020. The simulation results of the step-up pumping/injection tests also provide geological information that are used as prior information in the HT analyses.

Results obtained from different approaches suggest that HT is a robust method on characterizing the heterogeneity of hydraulic properties. Comparison of the investigated cases indicates that including more data into the inversion process is able to improve the prediction of the groundwater flow behavior. Yet, a data set that contains noise or error might result in the deterioration of estimation performance. Also, the comparison of results from the 2019 and 2020 HT surveys revealed some discrepancy in the estimated values. The discrepancy might be caused by: (1) different number of tests analyzed; (2) different number of data sets incorporated into the inverse model; (3) different order of tests included in SSLE_loop; and (4) different testing conditions. Overall, this study successfully characterizes the heterogeneity and demonstrates the effectiveness of the HT method at the Narashino Site.

Acknowledgements

This study was funded by Takenaka Corporation and the University of Waterloo. This thesis was completed under different kinds of support and encouragements from many people.

I would like to thank my supervisor Dr. Walter A. Illman for his guidance and patience throughout my project and for giving me this opportunity to study and research under his supervision. I benefit a lot from my master's courses and projects. These experiences would continually benefit and encourage me during my future study and work.

I would like to thank my committee members Dr. André Unger and Dr. Hyoun-Tae Hwang for their generous support and help which contributed to my project and thesis. Special thanks go to Dr. Steven Berg, Dr. Hyoun-Tae Hwang and Aquanty Inc. for providing technical support on HydroGeoSphere.

Thanks also go out to my college friends, including Ning Luo, Xin Tong, Xihua Wang, Huawen Zhao, Chenxi Wang and Suyuan Yang. Many parts of this study are improved based on their suggestions.

Finally, I would like to thank my family members for their endless support.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Statement

Despite decades of researches and studies, the characterization of aquifer hydraulic properties remains an ever present challenge for hydrogeologists. To understand groundwater flow and to predict contaminant transport, it is essential to accurately characterize the hydraulic parameters, such as hydraulic conductivity (K) and specific storage (S_s). Normally, these parameters are determined through empirical or analytical solutions and an assumption of homogeneity is often made, including Theis (1935) type curve matching or Cooper and Jacob (1946) straight line analysis. While the homogeneous assumption might be sufficient for some researches, many investigations, such as predicting solute plume, remediating the contaminants and designing waste disposal repositories, benefit from the accurate characterization of heterogeneity and the homogeneous assumption might lead to biased results (Wu et al., 2005).

Meanwhile, the measurement scale can also be an important factor that influences the estimation of K and S_s fields (Rovey II, 1998; Rovey II and Cherkauer, 1995). Traditional approaches of subsurface characterization rely on obtaining small scale of K measurements, such as grain size analyses and laboratory permeameter tests. Yet, these analyses are only able to obtain the K value of a point. Besides, the sampling of soil cores is highly possible to destroy the in-situ structure and condition, which leads to the biased estimation. In terms of slug tests, these tests can only obtain the aquifer characteristics of a small area near the testing well. Thus, large number of tests and analyses are required to adequately characterize subsurface heterogeneity using these

methods. However, this process is time-consuming and burdensome. Moreover, reliable S_s estimates are difficult to obtain from laboratory and small-scale field tests and are typicaly obtained through the analyses of pumping tests. There is clearly a need for an alternative method of characterizing the subsurface heterogeneity of hydraulic parameters. A higher resolution characterization of K and S_s fields should lead to improved performance in tracer transport predictions. As illustrated by Sudicky (1986), the heterogeneity of K field is able to alter the concentration profile, either enhance or delay the transport process. Therefore, the accurate delination of subsurface heterogeneity is critical in understanding and predicting groundwater flow and contaminant transport behavior, as well as in preventing and remediating groundwater contamination.

1.2 Review of Heterogeneity Modeling Methods

When modeling groundwater systems, it is necessary to acknowledge that one cannot obtain a comprehensive sample of the subsurface. In this case, inverse modeling is able to obtain more information (such as K and S_s distribution) about the subsurface using measured data. Inverse modeling consists of two parts: model identification and parameter estimation. Model identification refers to selecting the correct features of the model, such as governing equations and forcing functions (i.e., initial and boundary conditions as well as source and sink terms); and parameter estimation refers to assigning proper values to the groundwater model. In this study, two parameter estimators are used to characterize the Narashino Site: model-independent parameter estimator, PEST (Doherty, 2005) and Successive Linear Estimator, SLE (Yeh et al., 1996).

1.2.1 Model-independent Parameter Estimator (PEST)

PEST, as a model-independent parameter estimator, is able to assist in data interpretation, model calibration and predictive analysis by coupling with other programs. The algorithm embedded in PEST is the Gauss-Marquardt-Levenberg algorithm. For nonlinear problems, PEST use an iterative process to achieve parameter estimations. At the start of each iteration for all parameters, it is important to calculate the derivatives of all observations, for the purpose of linearizing the relationship between parameters and observations. Then a new parameter set is able to be obtained from the linearized problem and tested by running the model again. PEST then checks the convergence criteria (parameter change over certain successive iterations and maximum number of iteration) to determine whether to start a new iteration. PEST uses the finite difference method to calculate the derivatives of observation with respect to the parameter. PEST is able to switch between forward and central differences based on the user specified criteria to balance both computational efficiency and accuracy (Doherty, 2005).

1.2.2 Successive Linear Estimator (SLE)

Successive Linear Estimator (SLE) was developed by Yeh et al. (1996) to estimate the heterogeneity in parameters such as K and S_s and its uncertainty through geostatistical inverse modeling. SLE seeks the conditional means of the estimated parameter fields using the hydraulic head, the covariance of estimated parameters and the cross-covariance of the parameter and hydraulic head fields. The estimates are improved successively by updating the covariances and cross covariances during the iterative process. Besides, the iterative manner of SLE allows for the flow process to be considered by solving the governing flow equation (Yeh and Zhang, 1996; Zhang and Yeh, 1997) which is different from traditional kriging. A Sequential Successive Linear

Estimator (SSLE) approach was then developed as an extension of SLE. Unlike the SLE approach which can only invert head records from one pumping test data at a time, SSLE is able to include multiple data sets sequentially. During the inversion, SSLE treats the estimated *K* field covariance, conditioned by previous pumping test data, as prior information and uses that information to invert new pumping test data (Yeh and Liu, 2000). Compared to inverting all data sets simultaneously, SSLE avoids inverting larger size covariance and cross-covariance matrices and reduces the computational burden significantly by involving data sequentially into the inversion process. Then, a loop iteration scheme is introduced to the SSLE to improve the accuracy of sequential head data use. In the loop iteration scheme, a new data set is added into the SSLE only when all former added data sets meet the convergence criteria within one loop. This successfully avoided the situation that while the final estimates meet the convergence criteria of one data set, they may not meet the convergence criteria of another data set (Zhu and Yeh, 2005). Illman et al. (2008) reported that the order of pumping tests data included in the SSLE impacts the final estimation. Thus, to avoid this issue, a geostatistically-based estimator was developed to simultaneously includes all observed data sets in the inversion process for delineating the subsurface heterogeneity, named Simultaneous Successive Linear Estimator (SimSLE) (Xiang et al., 2009; Mao et al., 2013). SimSLE avoids the use of loop iterations and provided more constrains to the inverse problem. SimSLE requires more memory storage, since the size of the covariance and cross-covariance matrices are larger than the corresponding matrices in SSLE. To circumvent the issue of dealing with large covariance and cross-covariance matrices, a Reduced-Order Successive Linear Estimator (ROSLE) was developed by Zha et al. (2018). ROSLE reduces the covariance matrix of unknown parameters using the Karhunen-Loeve Expansion and uses a more efficient method to

obtain the eigenvalues and eigenvectors of the conditional covariance of SLE, which successively reduces the memory requirement compared to SimSLE.

For the purposes of this study, SSLE algorithm with the loop iteration scheme (SSLE_loop) was used to interpret hydraulic tests conducted in 2019 and 2020 to characterize the K and S_s heterogeneity of a field site in Japan. In the SLE algorithm, the following equation is used to describe the transient groundwater flow in a three-dimensional, heterogeneous porous medium:

$$\nabla \cdot [K \nabla H] + Q = S_s \frac{\partial H}{\partial t} \tag{1}$$

subject to the boundary and initial conditions:

$$H|_{\Gamma_1} = H_1, \qquad -K \nabla H|_{\Gamma_2} = q, \qquad H|_{t_0} = H_0, \tag{2}$$

where ∇ is the gradient operator $[L^{-1}]$, K is the saturated hydraulic conductivity $[LT^{-1}]$, H is the total head $[L]$, Q is the pumping or injection rate $[T^{-1}]$, S_s is the specific storage $[L^{-1}]$, t is time $[T]$, H_1 is the constant head $[L]$ at the Dirichlet boundary Γ_1, q is the specific discharge $[LT^{-1}]$ at the Neumann boundary Γ_2, and H_0 is the initial head $[L]$.

To represent the log-normal distribution of hydraulic parameters, SLE treats these parameters as a stochastic process and assumes $\ln K = \overline{K} + f$ and $\ln S_s = \overline{S_s} + s$, where $\overline{K}$ and $\overline{S_s}$ are the mean values and f and s are perturbations. Furthermore, transient hydraulic head H is also represented by a mean value $\overline{H}$ and corresponding perturbation h, which is $H = \overline{H} + h$. Substituting these stochastic variables into equation (1), taking the conditional expectation, and conditioning with observed data (heads and parameters) generates the following mean flow equation:

$$\nabla \cdot [\overline{K}_{con} \nabla \overline{H}_{con}] + Q = \overline{S}_{s_{con}} \frac{\partial \overline{H}_{con}}{\partial t} \tag{3}$$

where $\overline{K}_{con}$ is the conditional effective hydraulic conductivity $[LT^{-1}]$, $\overline{H}_{con}$ is the conditional effective hydraulic head $[L]$ and $\overline{S_s}_{con}$ is the conditional specific storage $[L^{-1}]$ (Zhu and Yeh, 2005).

Fig. 1 presents a flowchart of the SLE algorithm. The estimation process starts with a cokriging-like weighted linear combination of the directed measurements of parameters and transient head data. This procedure is able to obtain the initial estimates of the parameter fields. The weights are calculated based on the geostatistical properties (means, covariances, correlation lengths) of the parameter field, the covariances of the heads in space and time and the cross-covariances between parameters and head. The obtained initially-estimated field is then used to conduct forward simulation in the mean flow equation (3) to calculate the simulated head at certain observation locations and certain sampling times which are identical to the observed ones. The differences between the simulated and observed heads are then used to calculate a new weighted linear combination to improve the previously estimated fields. These two procedures iterate until the difference between two estimated parameter fields or the largest head difference between simulated and observed head is decreased to a prescribed convergence criteria or the iteration times reach the maximum iteration (Yeh and Zhang, 1996; Zhu and Yeh, 2005; Xiang et al., 2009).

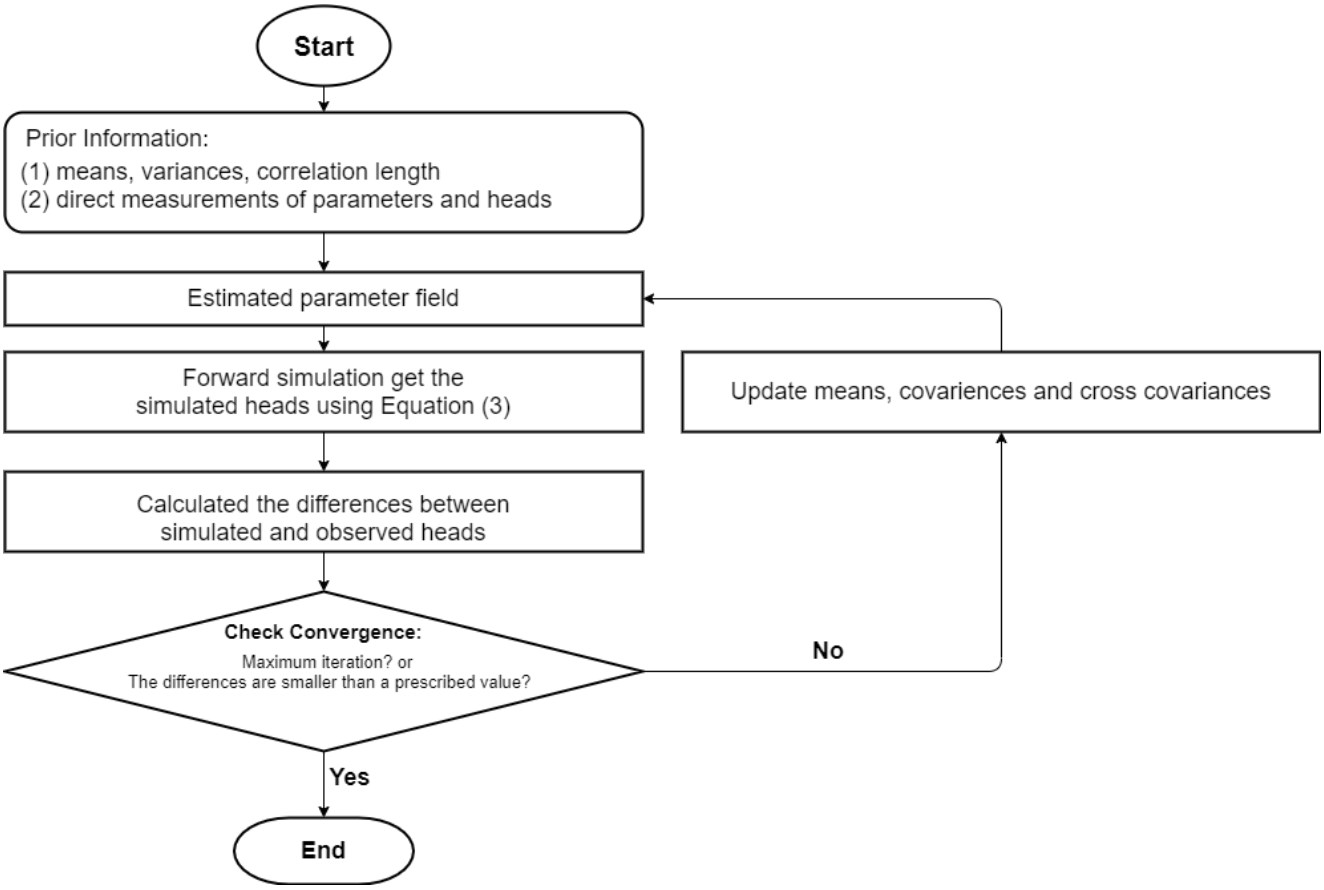

Fig. 1: *Flowchart of SLE.*

1.3 Review of Hydraulic Tomography

Hydraulic Tomography (HT), as an advanced method to delineate aquifer heterogeneity, relies on conducting a series of cross-hole pumping or injection tests in an aquifer. During each test, hydraulic heads are monitored at different observation intervals, yielding the HT data sets. To date, HT has been demonstrated to be a robust, promising method and has been applied at different scales.

This technique was first examined with synthetic numerical studies (e.g., Yeh and Liu, 2000; Zhu and Yeh, 2005; Zhu and Yeh, 2006; Xiang et al., 2009; Mao et al., 2013). Steady state hydraulic tomography (SSHT) was firstly conducted by Yeh and Liu, (2000), which showed

promising ability on estimating spatial K distribution for both two-dimensional and three-dimensional cases. Zhu and Yeh (2005) then improved HT with the interpretation of transient hydraulic data. The transient hydraulic tomography (THT) was able to successfully obtain not only the K tomogram, but the S_s tomogram as well. The tomograms are defined as an image of parameter spatial distribution obtained via HT.

In the laboratory, SSHT and THT were demonstrated to be effective in the well-controlled sandbox. The HT results also showed superiority through the comparison of several aquifer characterization approaches, such as effective parameter and geological zonation approaches (Illman et al., 2010, 2015; Berg and Illman, 2011a). Illman et al. (2008) presented a comparison of the K tomograms obtained by inversing synthetic and real pumping test data to evaluate how the signal-to-noise ratio and biases affected the HT results. They established that the quality of tomograms were related to the pumping test patterns, the order of inversing and the quality and quantity of the conditioning data. Zhao et al. (2016) and Luo et al. (2017) illustrated the importance of combining geological data in the HT survey through a sandbox study. They emphasized that geological information was able to improve the HT results, especially when the hydraulic data collected from pumping test was sparse resulting from wells being far apart. Yet, the accuracy of geological information was also found to be important, since inaccurate geological data led to poor performance in inversion results. Furthermore, HT analysis was also proven to correctly capture the heterogeneity within both confined and unconfined aquifers, under a fully controlled condition (Berg and Illman, 2012; Zhao et al., 2015). Sharmeen (2011) and Zhao (2020) also presented promising results that HT is able to delineate the fracture pattern and connectivity through the characterization of K and S_s.

In the field, several researches have been conducted during the recent years, which applied HT in both porous media and fractured media. For example, Berg and Illman (2011b) characterized a highly heterogeneous glaciofluvial aquifer and aquitard system with an HT survey at the North Campus Research Site (NCRS), located on the University of Waterloo campus. Compared to the geological information and permeameter K data, the THT analysis results showed that the overall features of the aquifer and aquitard system were able to be captured. Yet, estimation of the regions with little drawdown was not satisfactory enough. The K and S_s estimates can be smooth and erroneous when the observation densities were low. They suggested that this issue is able to be addressed by performing the pumping tests for a sufficiently long time to allow the drawdown to propagate through the porous medium especially in low K materials. Then, to characterize the interlayer and intralayer heterogeneity of highly contrasting materials, THT was conducted with existing and newly conducted long duration pumping test data by Zhao and Illman (2017, 2018). The results revealed that the inversion of long duration transient data from both aquifer and aquitard layers yielded reliable heterogeneous K and S_S distributions for even a highly heterogeneous site. Their studies suggested that it is necessary to examine the data quality, the information content of observed head from both high K and low K zones, and other types of information to avoid including redundant or erroneous data. Several studies compared the performance of different heterogeneity models applied at field studies and showed that HT analysis is best able to reproduce the hydraulic tests, in terms of smallest differences between simulated and observed drawdowns (Berg and Illman, 2013, 2015). In the study presented by Zhao and Illman (2017), the importance of geological information for three-dimensional SSHT analysis had been further applied in field studies. They concluded that the joint use of hydraulic data and

geological information can improve the accuracy of HT analysis, which is a similar conclusion to a former study obtained from the sandbox study (Zhao et al. 2016).

In fractured rocks, Illman et al. (2009) conducted two large scale cross-hole pumping tests in a fractured granite site located in Japan. They concluded that the estimated K and S_s tomograms clearly showed two fast flow pathways and low K zones at the site. Following the work of Illman et al. (2009), Zha et al. (2015, 2016) included two additional tests in the HT analysis at the fractured granite site. The results stressed that the non-redundant data sets included in the inverse process can help to improve the reliability and accuracy of the HT estimates.

Chen et al. (2019) investigated the reproducibility of K and S_s estimates from HT surveys conducted at different times. In their study, two-dimensional (2D) K and S_s tomograms obtained by inverting HT tests data collected during 2010 were compared to the K and S_s tomograms obtained by inverting data collected during 2012. Their study showed that the tomograms of the two HT analyses generally were similar, while noticeable differences were found near the boundary. This discrepancy was caused by using a 2D depth-averaged model which ignored the vertical variation of hydraulic properties. They emphasized that the reproducibility and predictability would vary with testing and interpretation conditions, such as model resolution and observation scale.

1.4 Objectives of This Study

In general, the performance of both SSHT and THT has been evaluated in various previously conducted research through numerical, laboratory and field studies. The importance of integrating geological information into HT has also been emphasized. Yet, few studies conducted comparison of HT results using data from different test periods to examine the reproducibility of HT, except a

2D case investigated by Chen et al. (2019). In this study, we focus on delineating the Narashino Site, Japan with three-dimensional (3D) models using three different parameter resolutions (effective parameter, geological zonation, and geostatistical inverse modeling). Tomograms were obtained from each groundwater flow scenarios during a two-year time period. The results of this study provide a deeper understanding of K and S_s distributions and corresponding benefits to the tracer tests conducted at the Narashino Site. Besides, this study provides understanding of the reproducibility, predictability and protentional limitations of HT. Specifically, the objectives of this study are to:

1. delineate 3D distributions of K and S_s fields of the Narashino Site, Japan operated by Takenaka corporation by conducting HT analyses using data collected in 2019 and 2020;
2. determine the test condition for HT analyses and obtain basic geological information by analyzing the step-up pumping/injection tests conducted during 2015 with effective parameter and geological zonation approaches;
3. investigate the impact of the number of data points included in the HT inversion by conducting HT analyses that include different data and evaluating tomograms of parameter fields and scatter plots that compare simulated and observed drawdowns;
4. investigate the reproducibility of HT by comparing tomograms acquired from HT analysis using data collected 2019 with those conducted in 2020.

Chapter 2

Site Description and Experimental Set-up

2.1 Site Geology

This study focuses on the Narashino Site located on the land managed by Takenaka Corporation in Narashino, Chiba Prefecture, Japan, as shown in Fig. 2. Based on the Standard Penetration Tests (SPT) results and borehole logs (shown in Appendix A), the stratigraphy that constitutes the survey area is divided into five different layers. The top layer is the fill soil layer (F) with a thickness of 2.8 m. Beneath the fill soil layer (F) is the Shimosa Group which contains four sub-layers. From top to bottom, they are Joso clay layer (Dc), the first sandy layer (Ds1), the second sandy layer (Ds2) and the third sandy layer (Ds3). The Dc layer is located between 2.8 mbgs (meters below ground surface) and 6.1 mbgs, with a thickness of 3.3 m. The Ds1 layer is the thickest and is located between 6.1 mbgs and 17 mbgs. The Ds2 and Ds3 layers have thicknesses of 5 m and 8.45 m, respectively.

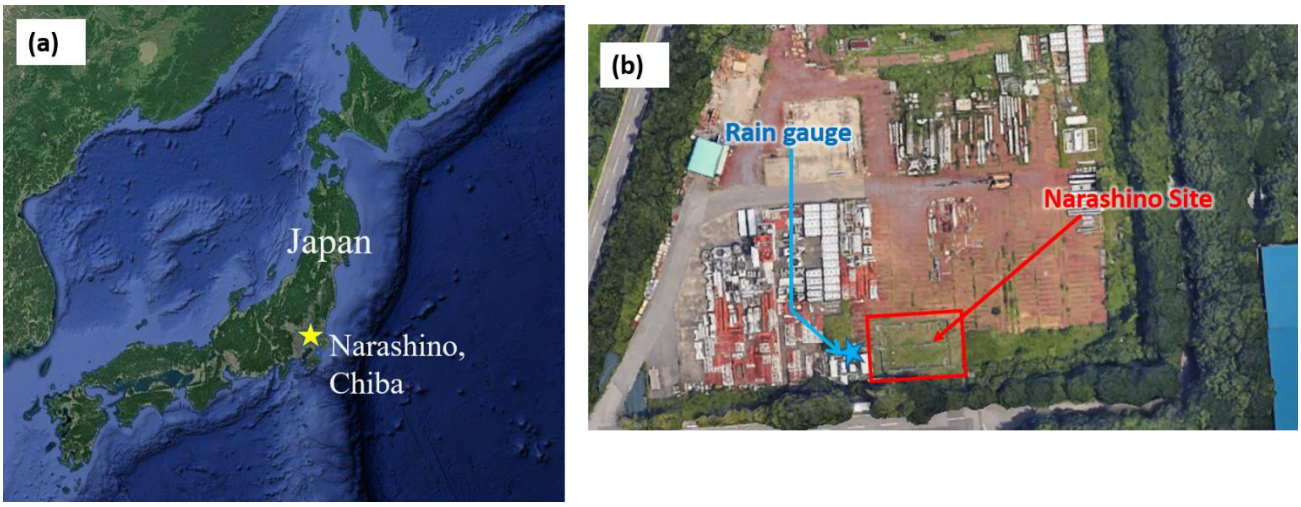

Fig. 2: *Narashino Site Map: (a) location of Narashino, Chiba, Japan (Google, n.d.-a); (b) location of Narashino Site (Google, n.d.-b). Note: these two maps are retrieved from Google Earth.*

2.2 Well Installation and Design

There were 18 wells installed at the Narashino Site, including 4 pumping/injection wells (Wells a1, a5, c1, c5) and 14 observation wells (Wells a3, b2, b3, b4, b5, c1.5, c2, c2.5, c3, c4, o1, o2, o3, o4). Among the 18 wells, Wells o1, o2, o3 and o4 are located outside the rectangular cut-off walls, and the other 14 wells are inside the cut-off walls. The cut-off walls had the dimensions of 20 m by 13m in plan view and they penetrated the subsurface to a depth of 19.2 m. Besides, the cut-off walls were removed during the 2020 HT tests. Wells a1, a5, c1, and c5 have 7.5-m long screens and were primarily used for pumping or injection purposes. All four wells were completed to a depth of 18 m with a diameter of 125 mm and each screen was located between 9.5 mbgs and 17 mbgs.

In contrast, the a-series, b-series and c-series observation wells have a diameter of 67 mm and have several screens that are 40-cm long. For Wells a3, b2, b4, b5, c2, c3 and c4, they are 17.8 m in depth. The first screen is located between 9.8 mbgs and 10.2 mbgs with the other two screens are separated 2 m apart. For Wells c1.5 and c2.5, they are 14.8 m deep and have the screen center at 10 mbgs and 13 mbgs. Well b3 is one of the deepest wells among all the wells which extend to 29.8 mbgs. It has 6 screens at 10 mbgs, 13 mbgs, 16 mbgs, 20 mbgs, 24 mbgs and 28 mbgs, respectively. Due to an issue at the site, Well b3 had been filled with bentonite from 17.8 mbgs to the bottom. Well o1 has the exact same configuration as Well b3. Wells o2 and o3 have the diameter of 67 mm, with a 11-m long screen that is located between 7 mbgs and 18 mbgs. For Well o4, its diameter is 125 mm with a screen located between 9.5 mbgs and 14.8 mbgs.

Packers were also installed in the observation wells to isolate the multiple screens. In the 2019 HT tests (Tests 1-8), the packers were installed at both 11.5 mbgs and 14.5 mbgs for the three-screen wells (Wells a3, b2, b4, b5, c2, c3, c4) and only at 11.5 mbgs for two-screen wells (Wells

c1.5 and c2.5). For Well b3, packers were installed at the depth of 11.5 mbgs and 14.5 mbgs. For Well o1, a packer was installed at the depth of 18 mbgs. Yet, in 2020 HT tests (Tests 9-12), packers in Wells a3, b2, b3 and b4 were removed during Tests 9 and 10. Then, during Tests 11 and 12, packers were reinstalled in Wells a3, b2, and b4 at the depth of 11.5 m, as shown in Table 1. Figs. 3 (a) to (d) show the different cross-sections along the Y-axis and Fig. 3 (e) and (f) show the plan view of the site and 3D well configurations, respectively.

Table 1: *Summary of packer locations installed within wells at the Narashino Site.*

Test	Packer Installation Depth	Wells										
		a3	b2	b3	b4	b5	c1.5	c2	c2.5	c3	c4	o1
1-8	11.5 mbgs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	14.5 mbgs	✓	✓	✓	✓	✓		✓		✓	✓	
	18 mbgs											✓
9, 10	11.5 mbgs					✓	✓	✓	✓	✓	✓	
	14.5 mbgs					✓		✓		✓	✓	
	18 mbgs											✓
11, 12	11.5 mbgs	✓	✓		✓	✓	✓	✓	✓	✓	✓	
	14.5 mbgs					✓		✓		✓	✓	
	18 mbgs											✓

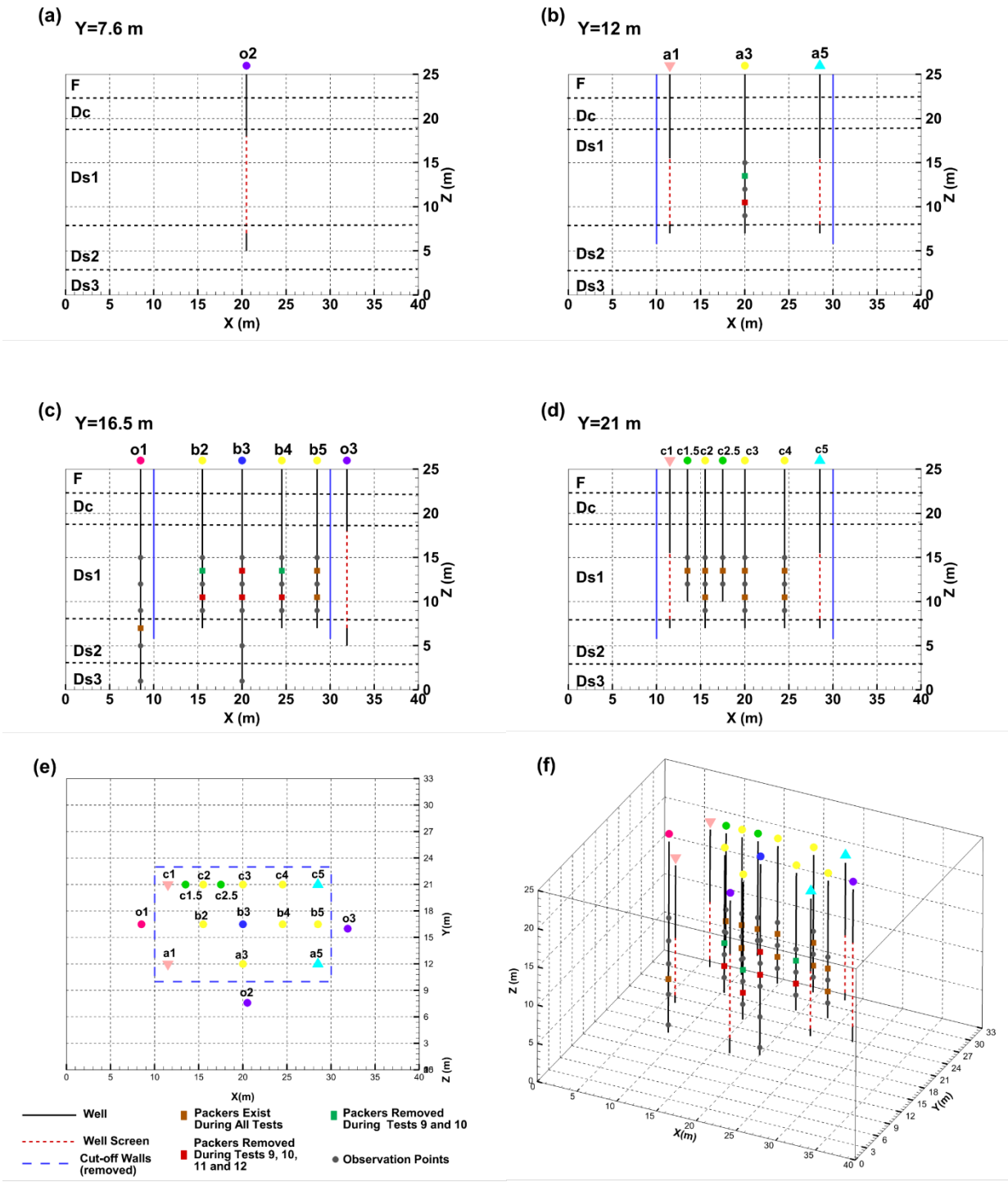

Fig. 3: *Cross sections along Y-axis: (a) Y =7.6 m; (b) Y = 12 m; (c) Y = 16.5 m; (d) Y = 21 m and (e) plan view of well locations at Narashino Site and (f) 3D well configurations without cut off walls. The black solid lines indicate the wells, the red dashed lines indicate the long well screens, the blue dashed lines indicate the cut off walls, the solid gray circles indicate the observation wells and the solid rectangles indicate the packers installed between well screens.*

2.3 Sensor and Data Collection

A network of pressure transducers was installed at the Narashino Site to collect hydraulic data during the pumping or injection tests. Depending on the particular test, up to 36 observation ports were monitored. Two types of pressure transducers were used at the site. The observation ports that were located at 10 mbgs were monitored with 0 – 20 m range pressure transducers with an accuracy of 0.1% of full scale (FS). All other observation ports were monitored with 0 – 10 m range pressure transducers with an accuracy of 0.2% FS. All transducers collected and recorded data at a frequency of $\frac{1}{60}\ H_Z$.

2.4 Tests Conducted at the Narashino Site

Various tests had been conducted at the Narashino Site. Initially, during the site set-up in 2014, SPTs were conducted. Soil samples collected during the SPTs were then used for grain size analyses. In addition, physical and thermal properties of soil samples were obtained. In-situ permeability tests were conducted at three depths to assess the horizontal and vertical permeability at the site. Next, in 2015, slug and step-up pumping/injection tests were carried out at the Narashino Site. A thermal tracer test that lasted about 75 days was conducted to further understand the thermal properties at the study site between November 2015 and February 2016. Then, in 2019, a group of HT tests as part of this project were conducted to obtain the 3D K and S_s tomograms at the site. Then, in 2020, 4 additional HT tests were performed. In this study, we focused on analyzing the step-up pumping/injection test conducted in 2015 and HT tests conducted in 2019 and 2020. Details on these tests are discussed in the following sections.

2.4.1 Step-up Pumping Tests

Eight step-up pumping and injection tests were conducted at the Narashino Site during the month of September, 2015. Among these eight tests, four of them were single well step-up pumping tests which drained the pumped water far away to avoid water recharge. The other four tests were step-up pumping and injection tests that involved two wells at the same time with one well used for pumping, while the other well was used for injection. In these four cases, the water pumped out of the subsurface was injected back at the same rate as the pumping wells to maintain mass balance. Each step-up test can be divided into four periods. In particular, each period lasted for 30 to 60 min, depending on different tests. The pumping/injection rates were kept constant during each period and increased gradually from one period to the next.

Only Wells a1, a5, c1, and c5 were included in these eight tests, and used for both pumping/injection wells and observation wells. When one or two wells were subjected to pumping or injection, the other wells were used to monitor hydraulic head change. General information on these tests is summarized in Table 2.

Table 2**:** *Summary of step-up pumping/injection tests performed at the Narashino Site in September 2015.*

Test Number	Pumping Well	Injection Well	Pumping or Injection Rate (L/min)	Pumping or Injection Duration (min)
1	a1	N/A	-11.5	40.0
			-17.6	40.0
			-24.3	40.0
			-42.0	40.0
			-50.1	45.0
2	a5	N/A	-8.8	39.0
			-14.6	40.0
			-18.5	40.0
			-33.6	40.0
			-50.6	39.0

Table 2 *(continued)*

Test Number	Pumping Well	Injection Well	Pumping or Injection Rate (L/min)	Pumping or Injection Duration (min)
3	c1	N/A	-7.5	42.0
			-12.7	39.0
			-16.1	41.0
			-29.1	45.0
			-42.4	66.0
4	c5	N/A	-14.5	42.0
			-20.4	31.0
			-31.8	30.0
			-38.0	30.0
			-49.5	35.0
5	a1	c5	-11.7/+11.7	39.0
			-17.7/+17.7	41.0
			-22.6/+22.6	40.0
			-43.0/+43.0	40.0
			-50.1/+50.1	41.0
6	c1	a5	-8.4/+8.4	44.0
			-12.2/+122	39.0
			-16.3/+16.3	40.0
			-30.0/+30.0	40.0
			-36.4/+36.4	44.0
7	c5	a1	-7.7/+7.7	31.0
			-12.4/+12.4	32.0
			-17.0/+17.0	33.0
			-29.4/+29.4	41.0
			-34.4/+34.4	45.0
8	a5	c1	-7.7+7.7	41.0
			-13.0/+13.0	44.0
			-16.7/+16.7	36.0
	c5	c1	-30.0/+30.0	60.0
			-34.2/+34.2	41.0

2.4.2 Hydraulic Tomography Tests

A total of 12 HT tests were conducted at the Narashino Site during 2019 and 2020. These tests ranged in duration from 7 to 24 hrs. Tests 1 to 8 were conducted during the months of August and September 2019 with the cut-off walls installed. Tests 9 to 12 were conducted during the months of August and September 2020. Different from Tests 1 to 8, the cut-off walls were removed during the test period of Tests 9 to 12. Furthermore, the locations of packers and pressure transducers were also slightly changed. Table 3 provides a summary of these 12 tests.

Table 3: *Summary of hydraulic tomography tests that were conducted at the Narashino Site during 2019 and 2020.*

Test Number	Pumping Well	Injection Well	Pumping or Injection Rate (L/min)	Test Start Time	Test End Time	Pumping or Injection Duration (min)	Cut-off Walls
1	a1	N/A	-20	8/8/2019 12:00	8/9/2019 12:00	1440	exist
2	c1	N/A	-20	8/6/2019 9:00	8/7/2019 9:07	1447	exist
3	a5	N/A	-20	8/11/2019 12:00	8/12/2019 12:00	1440	exist
4	c5	N/A	-20	8/14/2019 12:02	8/15/2019 12:00	1438	exist
5	a1, c1	N/A	-10	8/20/2019 9:00	8/21/2019 9:01	1441	exist
6	a5, c5	a1, c1	-10/ +9	9/1/2019 8:58	9/1/2019 15:00	422	exist
7	a1, a5, c1, c5	N/A	-5	8/27/2019 9:02	8/28/2019 9:00	1438	exist
8	a1, a5, c1, c5	N/A	-10	8/24/2019 9:00	8/25/2019 9:00	1440	exist
9	a5	N/A	-17.4	8/12/2020 9:59	8/13/2020 9:56	1440	remove
10	c5	N/A	-11.6	8/19/2020 10:00	8/20/2020 10:09	1440	remove
11	a1	N/A	-20.1	8/25/2020 10:00	8/26/2020 10:01	1440	remove
12	c1	N/A	-19.9	9/1/2020 11:00	9/2/2020 11:00	1440	remove

Note: "-" represents pumping rate; "+" represents injection rate; "N/A" represents not available

Chapter 3

Effective Parameter and Geological Zonation Approaches

3.1 Description of inverse groundwater model

Two ways of parameterization were considered to invert the step-up pumping/injection tests: (1) the effective parameter approach which treats the model to be homogeneous; and (2) the geological zonation approach which considers geological stratigraphy. The 3D groundwater model, Model 1, built for these two approaches had dimensions of 40.0 m × 33.0 m × 25.0 m. The simulation model domain was discretized into 22,644 variably sized rectangular finite elements with 25,000 nodes. The elements were refined from boundaries to the location of well screens. The maximum block size was 4.0 m × 4.0 m × 2.0 m, while the minimum block size was 0.5 m × 0.5 m × 0.5 m. For boundary conditions, the top and bottom faces were treated as no-flow boundaries, while the other four faces were treated as constant head boundaries. Fig. 4 shows the 3D model for simulating the step-up pumping tests. The unsaturated zone above the water level was considered during the simulation of step-up pumping/injection tests.

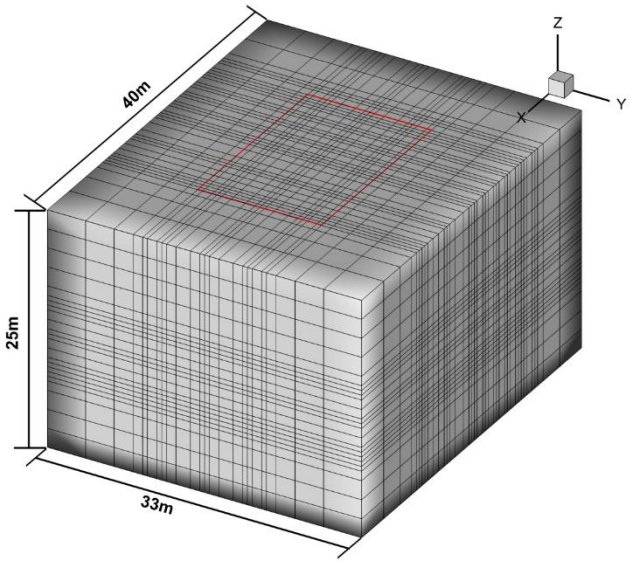

Fig. 4: *Three-dimensional model (Model 1) built for the interpretation of step-up pumping tests. The red line represents the cut-off walls.*

3.2 Data selected for analyses

A group of data from these eight tests were selected to conduct inverse modeling of the step-up pumping tests. Fig. 5 shows observed drawdowns in Wells a1, a5, c1, and c5 for Test 1. The drawdown versus time curves of the other seven tests are listed in Appendix B. To capture the different stages of the drawdown curves, data were selected every minute from the beginning of each test to the end. Data from the recovery stage were also included in the inversion processes. Number of data selected for analyses of each test is summarized in Table 4.

Table 4: *Number of data selected for analyses of each step-up pumping/injection tests.*

Step-up Pumping/Injection Tests								
Test	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8
Number of Data Points	874	834	2512	672	1080	1432	1216	1908

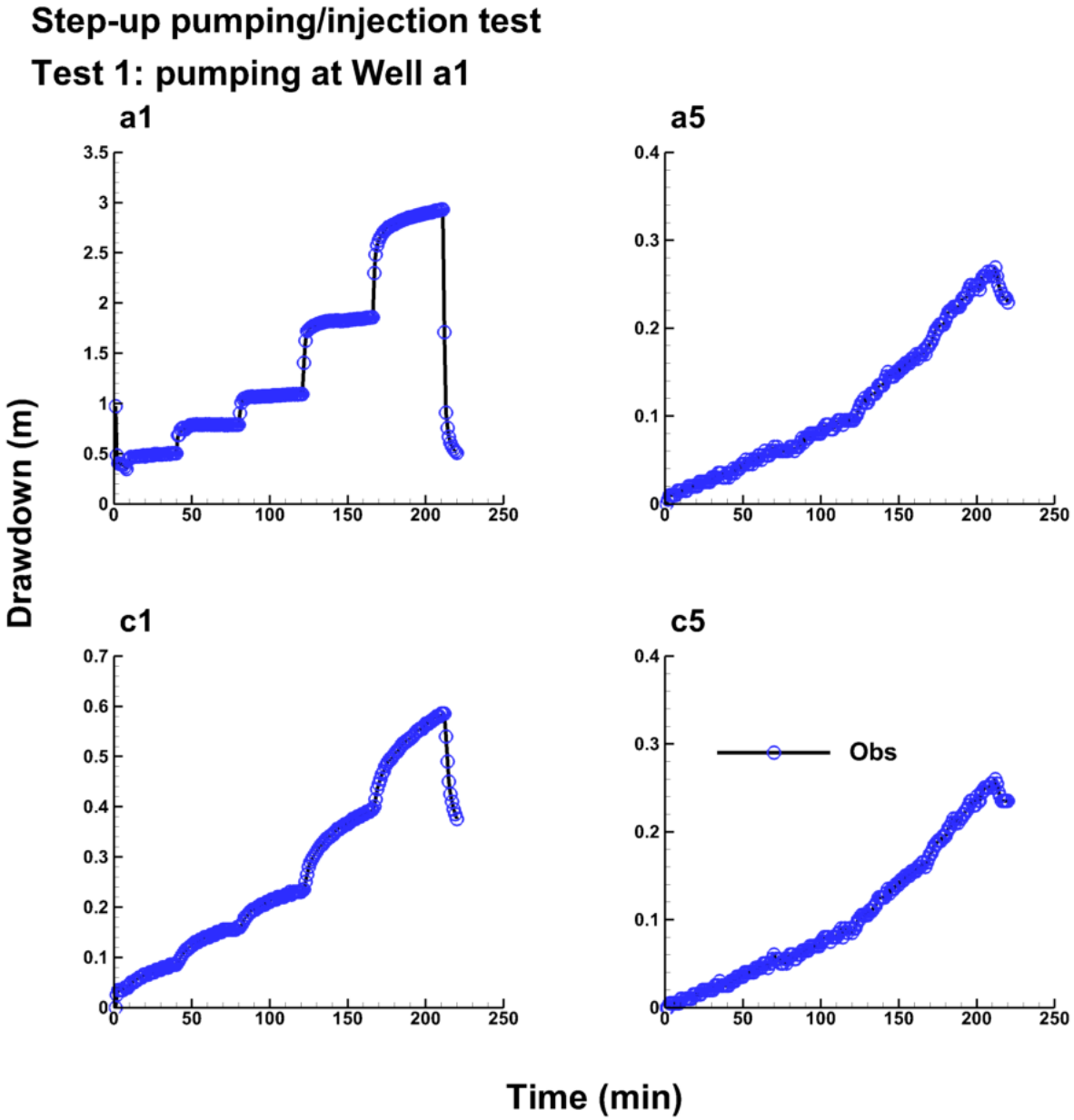

Fig. 5: *Observed drawdown curves of step-up pumping/injection Test 1 (pumping at Well a1). Open circles indicate the drawdown data used for inverse modeling.*

3.3 Effective Parameter Approach

3.3.1 Inverse Model Set-up

The effective K and S_s values were estimated by coupling the groundwater flow model HydroGeoSphere (HGS) (Therrien et al., 2010) with the model-independent parameter estimator PEST (Doherty, 2005). The entire model domain was considered to be homogeneous and isotropic. The initial K value was set to be $2.24 \times 10^{-3}\, m/min$, with a minimum bound of $1.00 \times 10^{-12}\ m/min$ and a maximum bound of $1.00 \times 10^{5}\ m/min$. The initial value of S_s was set to $9.87 \times 10^{-5}\, m^{-1}$ with a minimum bound of $1.00 \times 10^{-12}\, m^{-1}$ and maximum bound of $1.00 \times 10^{1}\, m^{-1}$. The initial value of K was the geometric mean of the K values obtained through Creager's empirical solution (Creager et al., 1945) in the report of soil survey (Tokyo Soil Research, 2015) provided by Takenaka Corporation. Equation 7 shows the Creager's empirical solution.

$$K = 0.00359 \times {D_{20}}^{2.237} \tag{7}$$

in which K is hydraulic conductivity (m/s), D_{20} indicates that 20% of the sample is smaller than that diameter (mm).

Since there is no S_s information provided, a typical value for sandy material was chosen from Batu (1998). The cut-off walls were impermeable and the K and S_s of the wall were assigned extremely small values of $1.00 \times 10^{-12}\ m/min$ and $1.00 \times 10^{-12}\, m^{-1}$, respectively. The K and S_s of cut-off walls were fixed during parameter estimation. Table 5 summarizes the initial parameter inputs for the effective parameter model.

***Table 5**: Initial parameter inputs of the effective parameter approach.*

	K (m/min)			S_s (1/m)		
	Initial value	Minimum Bound	Maximum Bound	Initial value	Minimum Bound	Maximum Bound
Model Domain	2.24E-03	1.00E-12	1.00E+05	9.87E-05	1.00E-12	1.00E+01
Cut-off Walls	1.00E-12	Fixed		1.00E-12	fixed	

3.3.2 Results of the Effective Parameter Approach

The PEST calibration of these eight step-up pumping tests using the effective parameter approach were performed on a PC with a six-core CPU and 16 GB of RAM. The calibration process took approximately 0.5 hour for each test to complete after about 50 model calls. These eight tests were inverted individually. The estimated K and S_s values and their 95% confidence intervals of each test are listed in Table 6. The estimated effective K values vary between $1.34 \times 10^{-3}\, m/min$ and $2.82 \times 10^{-3}\, m/min$ depending on the pumping and injection locations with a geometric mean of $1.97 \times 10^{-3}\, m/min$. The estimated effective S_s values also vary with pumping and injection locations and have a minimum value of $3.23 \times 10^{-4}\, m^{-1}$, a maximum value of $8.27 \times 10^{-3}\, m^{-1}$ and a geometric mean of $6.22 \times 10^{-4}\, m^{-1}$. It is worth noting that the PEST calibrated K had a narrower 95% confidence interval compared to S_s. This suggests that S_s is more difficult to estimate, and larger uncertainty exists in S_s estimation. Due to different order of magnitude of the estimates, a relative limit which is the estimated limit divided by the parameter estimate is used to evaluate the range between the upper and lower limits. The drawdown versus time curves that compares the observed and simulated drawdowns obtained from the effective parameter approach are shown in Fig. 8 and Appendix C.

Table 6**:** *Summary of PEST calibrated K and S_s values and their 95% confidence intervals for the effective parameter approach.*

Test Number	Parameter		95% Percent Confidence Interval			
			Lower Limit	Upper Limit	Relative Lower Limit	Relative Upper Limit
Test 1	*K* (m/min)	2.54E-03	2.50E-03	2.57E-03	9.84E-01	1.01E+00
	S_s (1/m)	6.68E-04	6.18E-04	7.23E-04	9.25E-01	1.08E+00
Test 2	*K* (m/min)	2.29E-03	2.25E-03	2.33E-03	9.83E-01	1.02E+00
	S_s (1/m)	7.71E-04	7.04E-04	8.44E-04	9.13E-01	1.09E+00
Test 3	*K* (m/min)	1.34E-03	1.33E-03	1.35E-03	9.93E-01	1.01E+00
	S_s (1/m)	8.27E-04	7.96E-04	8.58E-04	9.63E-01	1.04E+00
Test 4	*K* (m/min)	1.67E-03	1.66E-03	1.68E-03	9.94E-01	1.01E+00
	S_s (1/m)	7.07E-04	6.77E-04	7.39E-04	9.58E-01	1.05E+00
Test 5	*K* (m/min)	2.82E-03	2.80E-03	2.83E-03	9.93E-01	1.00E+00
	S_s (1/m)	8.13E-04	7.45E-04	8.87E-04	9.16E-01	1.09E+00
Test 6	*K* (m/min)	1.86E-03	1.84E-03	1.88E-03	9.89E-01	1.01E+00
	S_s (1/m)	7.28E-04	6.23E-04	8.51E-04	8.56E-01	1.17E+00
Test 7	*K* (m/min)	1.96E-03	1.93E-03	1.98E-03	9.85E-01	1.01E+00
	S_s (1/m)	3.90E-04	3.08E-04	4.94E-04	7.90E-01	1.27E+00
Test 8	*K* (m/min)	1.68E-03	1.67E-03	1.70E-03	9.94E-01	1.01E+00
	S_s (1/m)	3.23E-04	2.79E-04	3.73E-04	8.64E-01	1.15E+00

3.4 Geological Zonation Approach

3.4.1 Inverse Model Set-up

The inverse modeling of geological zonation model was also achieved by coupling HGS with PEST. Based on geological stratigraphy at the Narashino Site, the geological zonation model was divided into four horizontal layers. The top layer Dc merged the fill soil layer and the clay layer, which has a 6 m thickness in total. Beneath Dc is a 11 m-thick sand layer Ds1, and beneath this layer is the sand layer Ds2, which is 5 m-thick. The bottom layer is another sand layer Ds3 that is 3 m-thick. The x = 20 m cross section along the y-z direction is shown in Fig. 6.

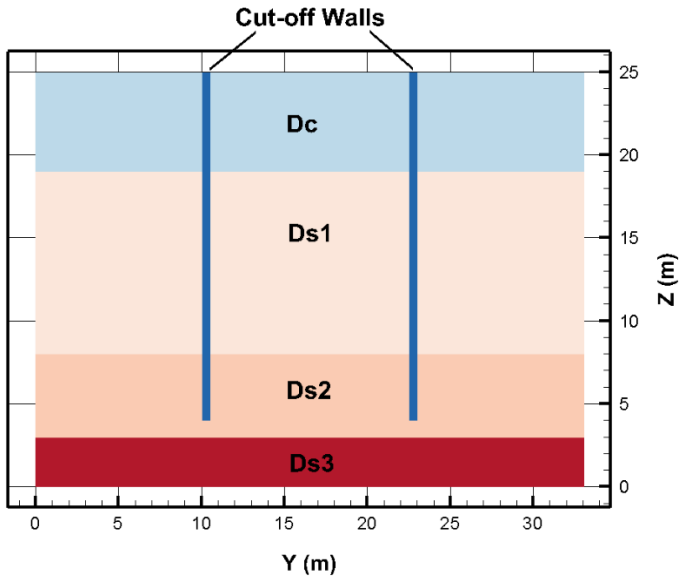

Fig. 6: *Y-Z direction cross section of the stratified model when X= 20 m. From top to the bottom, there are Dc, Ds1, Ds2, and Ds3 layers. The solid blue lines represent the cut-off walls.*

Two cases were considered to evaluate the impact of anisotropy. Case 1 treated the stratified model to be isotropic, while Case 2 treated the model to be anisotropic. For Case 1, a homogeneous and isotropic hydraulic parameters field was initially input into the geological zonation model before PEST estimates the parameter field. During the inversion process, Case 1 adjusted the K_x and assigned K_y and K_z the same value as K_x. Yet, for Case 2, different initial inputs were assigned to corresponding layers to provide geological information as prior information. Case 2 adjusted K_x and K_z during the inversion and ignored the horizontal anisotropy by assuming K_y to be equal to K_x. Besides, K and S_s of cut-off walls were fixed at $1.00 \times 10^{-12}\ m/min$ and $1.00 \times 10^{-12}\ m^{-1}$, respectively. Tables 7 and 8 summarize the initial parameter input data for Cases 1 and 2.

Table 7: *Initial parameter inputs of the effective parameter approach: Case 1 (Isotropic).*

	K (m/min)			S_s (1/m)		
	Initial value	Minimum Bound	Maximum Bound	Initial value	Minimum Bound	Maximum Bound
Dc	2.24E-03	1.00E-12	1.00E+05	9.87E-05	1.00E-12	1.00E+01
Ds1	2.24E-03	1.00E-12	1.00E+05	9.87E-05	1.00E-12	1.00E+01
Ds2	2.24E-03	1.00E-12	1.00E+05	9.87E-05	1.00E-12	1.00E+01
Ds3	2.24E-03	1.00E-12	1.00E+05	9.87E-05	1.00E-12	1.00E+01
Cut-off Walls	1.00E-12	Fixed		1.00E-12	Fixed	

Table 8: *Initial parameter inputs of the effective parameter approach: Case 2 (Anisotropic).*

	K_x/K_y (m/min)			K_z (m/min)			S_s (1/m)		
	Initial value	Minimum Bound	Maximum Bound	Initial value	Minimum Bound	Maximum Bound	Initial value	Minimum Bound	Maximum Bound
Dc	6.00E-07	1.00E-09	1.00E-05	6.00E-07	1.00E-10	1.00E-05	3.00E-03	1.00E-05	1.00E-01
Ds1	1.62E-03	1.00E-06	1.00E-01	1.62E-03	1.00E-07	1.00E-01	3.24E-04	1.00E-07	1.00E-02
Ds2	2.62E-03	1.00E-06	1.00E-01	2.62E-03	1.00E-07	1.00E-01	3.24E-04	1.00E-07	1.00E-02
Ds3	2.37E-03	1.00E-06	1.00E-01	2.37E-03	1.00E-07	1.00E-01	3.24E-04	1.00E-07	1.00E-02
Cut-off Walls	1.00E-12	Fixed		1.00E-12	fixed		1.00E-12	fixed	

3.4.2 Results of the Geological Zonation Approach

The PEST calibration of eight step-up pumping/injection tests using the geological zonation model were performed on the same PC (six-core CPU and 16 GB of RAM) used to calibrate the effective parameter models. The simulations of Case 1 were completed after about 150 model calls, while the simulations of Case 2 took about 300 model calls. The average estimated parameter values of Cases 1 and 2 for all eight tests are listed in Tables 9 and 10, respectively. The detailed estimated values and the corresponding 95% confidence interval of each test are presented in Appendix D.

Generally, PEST calibration results of Case 1 revealed a reasonable estimation for the three sand layers (Ds1, Ds2, Ds3). The averaged estimated K values vary between $1.62 \times 10^{-3}\ m/min$ and $3.60 \times 10^{-3}\ m/min$, which are close to the initial input values. However, the clay layer (Dc) which is known as a low K zone, has a K value that is similar to the sand layers. Also, the estimated S_s values of these four layers also tend to be similar to each other. This might be caused by the absence of observation points in the clay layer (Tong, 2018). Thus, no data were used directly to adjust the K and S_s of clay layer. This also led to extremely large 95% confidence intervals of the estimated K and S_s for the Dc layer.

For Case 2, when the geological information is provided as prior information to the inverse model, the values of K and S_s for all four layers are more reasonable. The estimated K value of clay is much smaller than the ones for sand. The S_s value of clay is also larger than those of sand. The ratio of estimated K_x/K_z varies from 0.01 to 73.72, with an average about 5.54. The averaged K_x/K_z ratio (5.54) is close to the K_x/K_z ratio that Takenaka Corporation provided (7.64) in the soil survey report (Tokyo Soil Research, 2015). However, for most groups, the 95% confidence intervals of Dc layer and Ds3 layer have a wider range, except for the K_x/K_y estimates of Tests 5, 7 and 8; the K_z estimates of Tests 5 and 7; and the S_s estimates of Tests 3 and 5. Since there are no observed points in Dc and Ds3 layers, one might not be able to obtain reliable parameter estimates with PEST for these layers. The drawdown versus time curves that compares the observed and simulated drawdowns obtained from the geological zonation approach are shown in Fig. 8 and Appendix C.

***Table 9**: Average of the estimated K and S_s values of 8 tests for the geological zonation approach: Case 1 (isotropic).*

Layer	Parameter	
Dc	*K* (m/min)	1.74E-03
	S_s (1/m)	9.82E-04
Ds1	*K* (m/min)	1.62E-03
	S_s (1/m)	1.02E-03
Ds2	*K* (m/min)	2.62E-03
	S_s (1/m)	3.24E-04
Ds3	*K* (m/min)	3.60E-02
	S_s (1/m)	1.79E-03

***Table 10**: Average of the estimated K and S_s value of 8 tests for the geological zonation approach: Case 2 (anisotropic).*

Layer	Parameter	
Dc	K_x/K_y (m/min)	1.13E-06
	K_z (m/min)	5.45E-07
	S_s (1/m)	9.58E-03
Ds1	K_x/K_y (m/min)	1.32E-03
	K_z (m/min)	7.44E-04
	S_s (1/m)	1.98E-04
Ds2	K_x/K_y (m/min)	1.13E-02
	K_z (m/min)	1.83E-02
	S_s (1/m)	2.35E-03
Ds3	K_x/K_y (m/min)	1.51E-02
	K_z (m/min)	5.79E-03
	S_s (1/m)	2.85E-03

3.5 Comparison of Model Calibration Performance

To evaluate model calibration performance, simulated drawdowns were compared to the observed drawdowns in a scatter plot, as shown in Fig. 7. The mean absolute error (L_1), mean square error (L_2), slope of linear model fit as well as the coefficient of determination (R^2) were

calculated to quantitatively assess model performance. The L_1, L_2 and R^2 values were calculated as:

$$L_1 = \frac{1}{n}\sum_{i=1}^{n} | x_i - \hat{x}_i| \tag{4}$$

$$L_2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{6}$$

where n is the total number of drawdown data, i is the data number, x_i represents the i-th simulated drawdown, $\hat{x}_i$ represents the i-th simulated drawdown and $\bar{x}$ is the mean value of observed drawdown.

The Creager's empirical solution (Equation 7) that Takenaka Corporation used in their soil survey report (Tokyo Soil Research, 2015) was also utilized to conduct forward simulation to obtain the simulated drawdowns and compared with the effective parameter and geological zonation approaches.

Examination of scatter plots (Fig. 7) reveals that, for the effective parameter approach, the linear model fit has a slope of 0.99 and R^2 of 0.96 between the simulated and observed drawdowns. The Creager's empirical solution yields the worst result with a slope of 0.80 and R^2 of 0.70. For geological zonation approach Case 1, the simulation results were the best, with a slope of 0.99 and R^2 of 0.97. For Case 2 which considered anisotropy, the results did not improve compared to Case 1, yielding a slope and R^2 of 0.92 and 0.96, respectively.

In terms of L_1 and L_2 norms, the numerically calibrated results performed better than the Creager's empirical solution. Also, the geological zonation approach performed better than the

effective parameter approach. Yet, the results were not improved by consideration of anisotropy. Table 11 summarizes the slope, R^2, L_1 and L_2 norms of these simulation results.

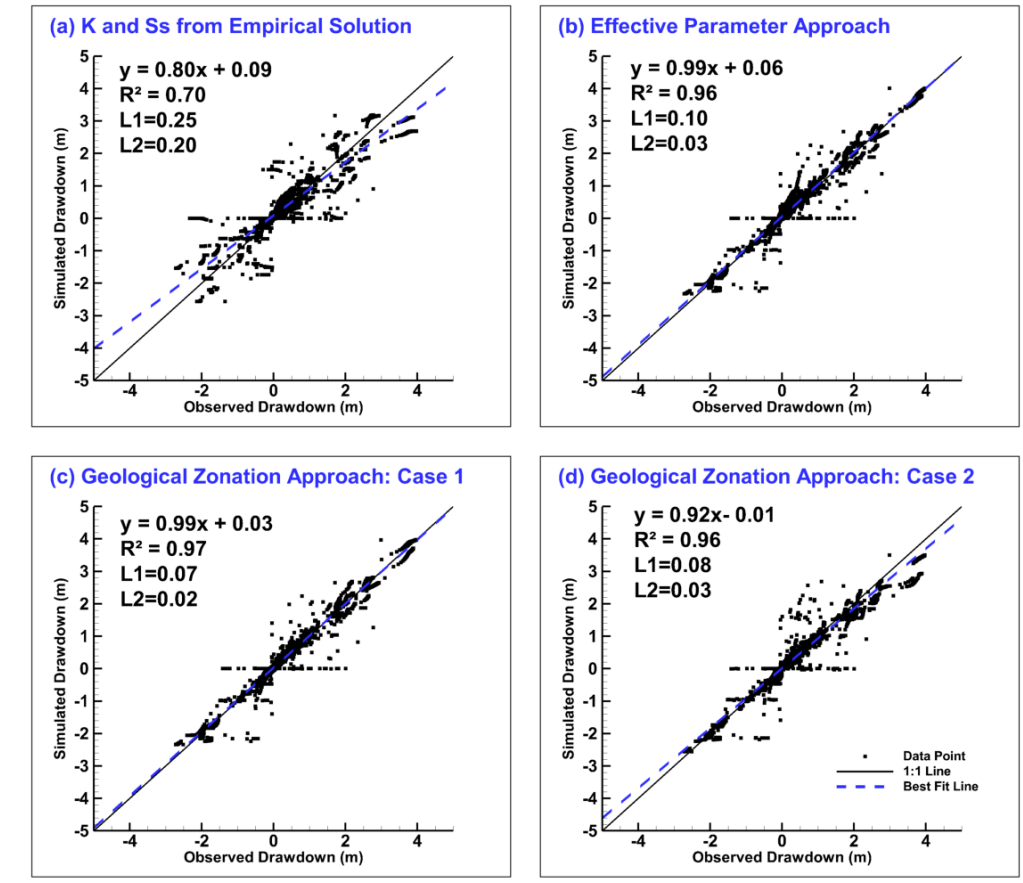

Fig. 7: *Scatter plots of observed vs. simulated drawdowns using transient data from 8 groups of step-up pumping tests: (a) isotropic effective parameter approach using K and S_s from the Creager's empirical solution; (b) isotropic effective approach using PEST estimated K and S_s; (c) geological zonation approach: Case 1 is isotropic stratified model; (d) geological zonation approach: Case 2 is anisotropic stratified model. The black solid line is a 1:1 line, which shows a perfect match, while the blue dash line is the best fit line.*

Table 11: *Results summary of the Creager's empirical solution, effective parameter, and geological zonation approaches.*

	Slope	R^2	L_1	L_2
Creager's Empirical Solution	0.80	0.70	0.25	0.20
Effective Parameter	0.99	0.96	0.10	0.03
Geological Zonation: Case 1	0.99	0.97	0.07	0.02
Geological Zonation: Case 2	0.92	0.96	0.06	0.03

Drawdown versus time curves of eight step-up pumping/injection tests were plotted to further assess the performance of these approaches. The drawdown versus time curves of Test 1, Test 4 and Test 5 (shown in Fig. 8) were discussed as examples in the paragraph. The other tests are listed in Appendix C. Overall, the Creager's empirical solution performs the worst as expected, followed by the effective parameter approach, while the geological zonation approach performs the best in terms of L_1 and L_2 norms, as both Creager's empirical solution and effective parameter approach ignore heterogeneity. However, the performance of geological zonation approach Case 1 and 2, varied depending on the pumping tests location. For example, as shown in Fig. 8, the estimated drawdowns of the geological zonation approach Case 2 (consider the anisotropy) matched better to the observed drawdowns compared with Case 1 (not consider the anisotropy) for Test 1. Yet, for Tests 4 and 5, the geological zonation approach Case 1 results performed better than Case 2. The performance difference of each case on different step-up pumping/injection test simulations suggests that simply building a four-layer model may not be able to fully represent site heterogeneity. In this study, during the simulation of step-up pumping/injection tests, homogeneous *K* and *Ss* values were assigned for each layer, and the heterogeneity within each layer was ignored. This leads to the fact that for some observation points, a better match was

achieved, while other observation points revealed more deviation. Consideration of anisotropy did improve the curve-fitting results for Tests 1 and 8, however, the geological zonation approach Case1 performed better than Case 2 for the other tests. Thus, anisotropy at this site may not be as influential as heterogeneity is. Moreover, there are some drastic fluctuations of drawdowns in the observed data (as shown in Fig. 8c in Well a5), which cannot be simulated by any of these approaches. Since these drawdown data increase rapidly and recover soon, the fluctuations may have been caused through human factors, such as pulling the tube out of or into the well or it can also be due to noise in collected data.

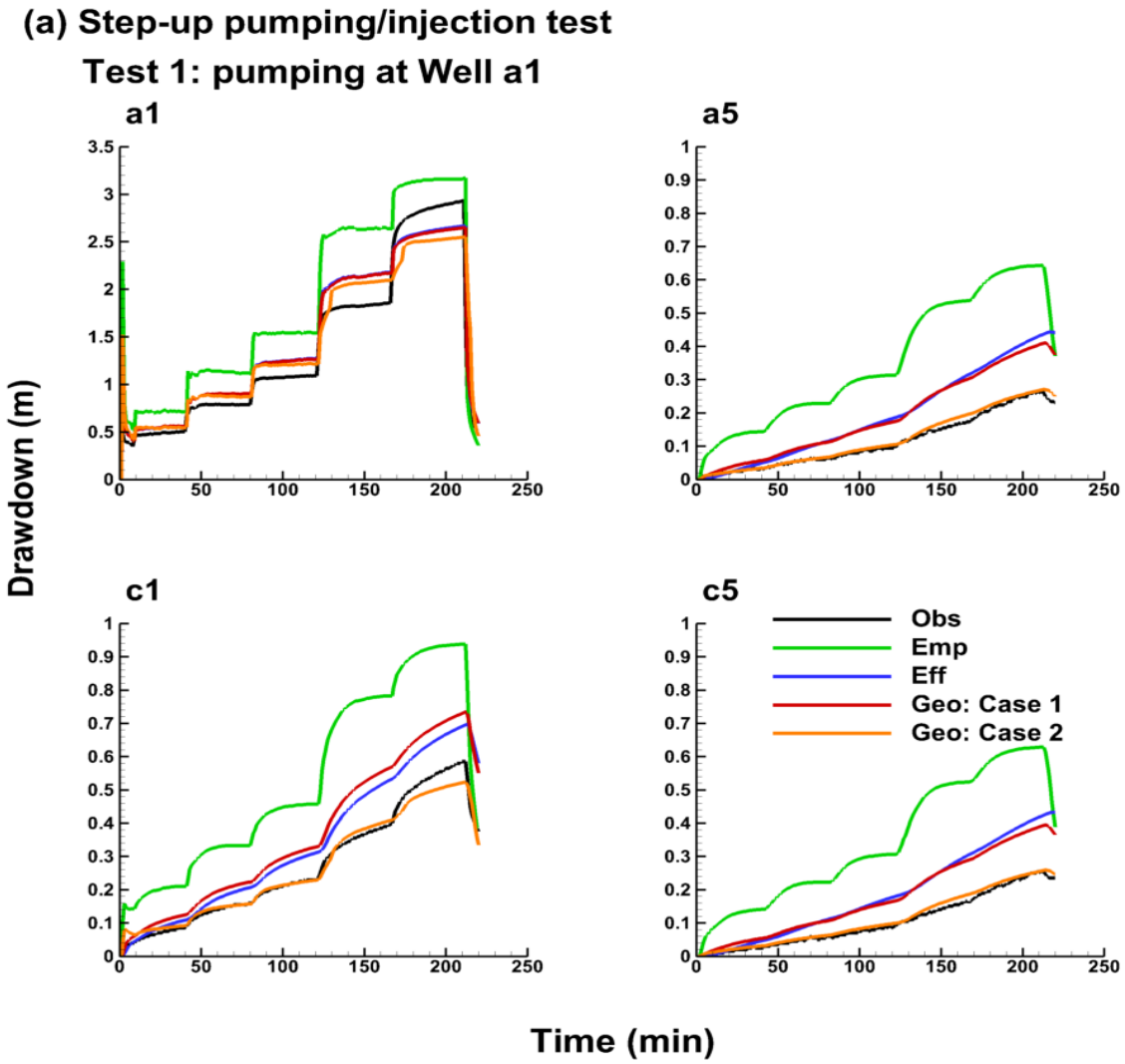

Fig. 8: *Drawdown vs. time plots for step-up pumping/injection tests: (a) Test 1: pumping at Well a1; (b) Test 4: pumping at Well c5; (c) Test 5: pumping at Well a1, injection at Well c5. The black line represents the observed drawdowns, the green line represents the simulated drawdowns of the Creager's empirical solution, the blue line represents the simulated drawdowns of effective parameter approach, the red and orange line represent the simulated drawdowns of geological zonation approach: Case 1 and Case 2, respectively.*

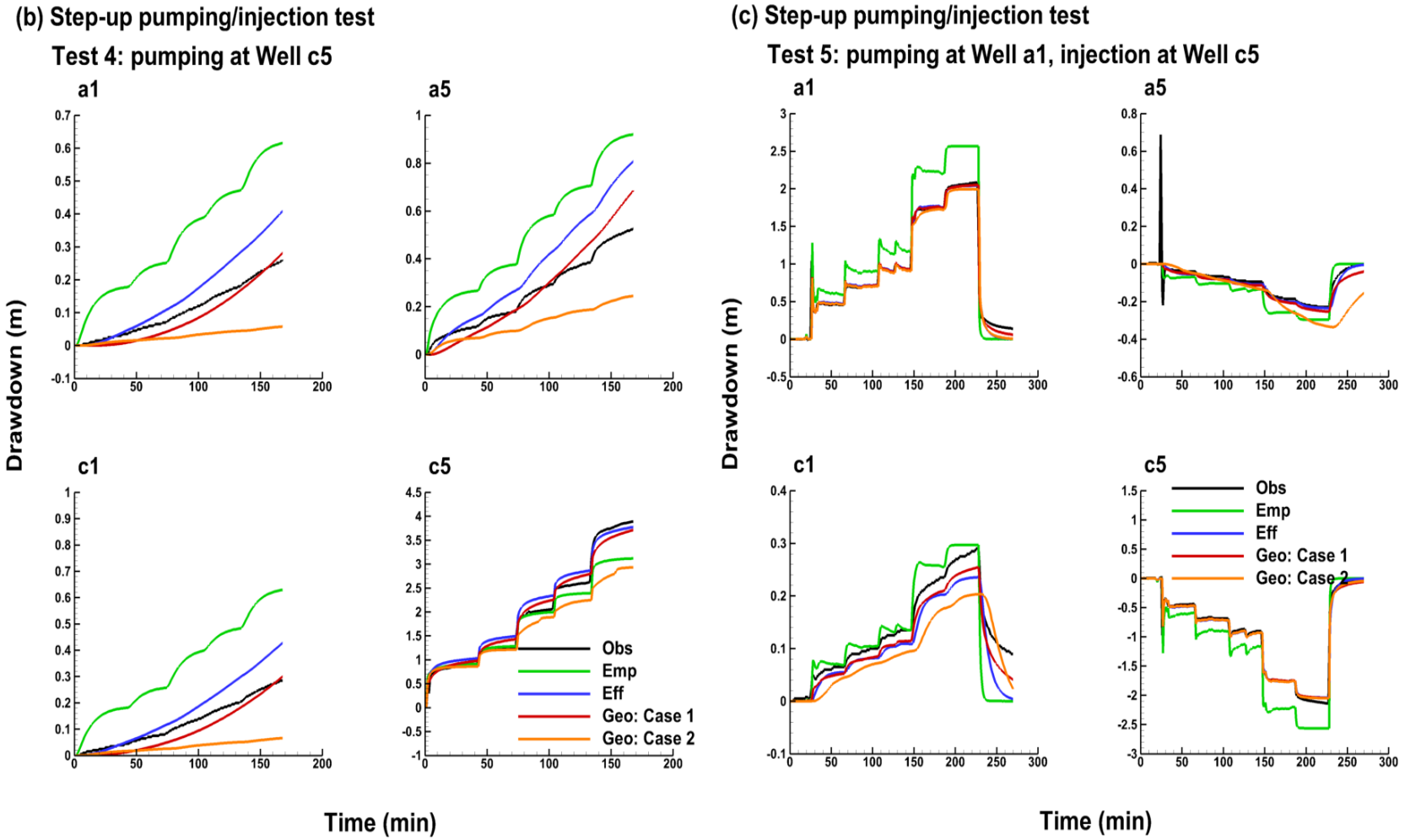

Fig. 8 *(continued)*

Chapter 4

Geostatistical Inversion Approach

4.1 Description of the geostatistical inversion model

A highly parameterized geostatistical inversion approach was considered for inverting the HT data in this study. In order to simulate the groundwater flow for both forward and inverse modeling, a smaller 3D model (Model 2), compared to Model 1, was built to conduct the inverse analyses. Model 2 (shown in Fig. 9) was extracted directly from the part that is inside the cut-off walls of Model 1. It has dimensions of 20.0 m × 13.0 m × 25.0 m and contains 11,136 elements with 12,750 nodes. The elements were gradually refined from the boundaries to the locations of well screens, with the largest ones having dimensions of 1.5 m × 1.5 m × 2 m, while the smallest ones having dimensions of 0.5 m × 0.5 m × 0.5 m. In terms of boundary conditions, the cut-off walls were treated as no-flow boundaries, while the other part of the four side boundaries were assumed to be constant head boundaries. That is, during the simulation of 2019 HT tests of which the cut-off walls existed at the well site, the four side boundaries consisted of two parts, with no-flow boundaries prescribed from 0 to 19.2 mbgs and constant head boundaries were assigned from 19.2 to 25 mbgs. In contrast, during the simulation of 2020 HT tests of which the cut-off walls were removed, the four sides were assumed to be constant head boundaries. The top and bottom faces were treated to be no-flow boundaries through the 2019 and 2020 HT tests. The unsaturated zone above the water level was not considered during the inverse modeling of HT tests.

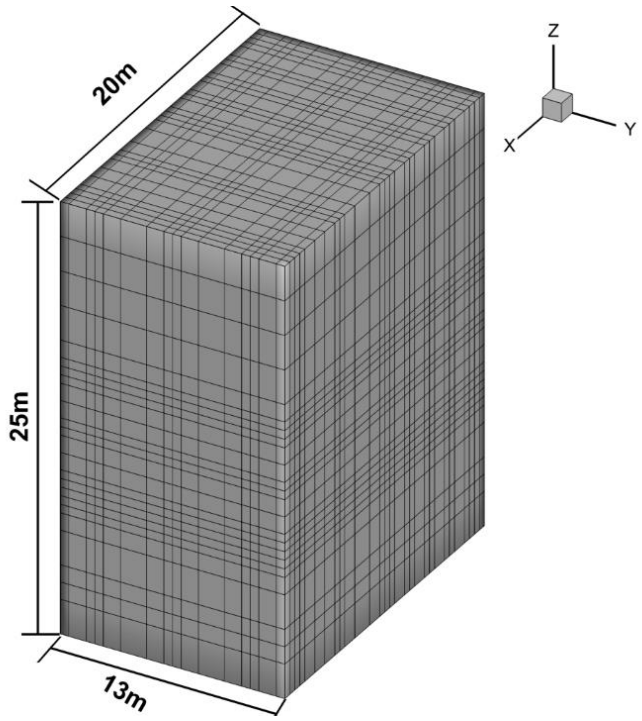

Fig. 9: *Three-dimensional model built for HT analyses (Model 2).*

4.2 Data selected for analyses

4.2.1 Data selected for 2019 HT analyses

To perform THT analyses of 2019 HT tests (Test 1 to Test 8), five data points were selected to capture the shape of the drawdown versus time curves at every observation port for each test. These transient data are collected from 60 minutes, 180 minutes, 420 minutes, 660 minutes and 900 minutes from the start of the HT tests. Up to 840 data points from 2019 HT tests were used for the inverse modeling efforts.

4.2.2 Data selected for 2020 HT analysis

Examination of water level data of HT tests collected during 2020 revealed that the water level at the Narashino Site decreased throughout the testing period of August and September, 2020,

(shown in Fig. 10), which was not observed during the step-up pumping/injection tests and 2019 HT tests. This overall decline in hydraulic head might have been caused by unknown pumping events occurring near the study area or may also be due to ambient groundwater gradient. Thus, in order to properly capture the water level changes induced through HT tests, raw data required to be adjusted to minimize the influence of the steady decrease in water levels. Data were selected from the observation ports every 30 minutes except for the pumping and recovery period to capture the linear relationship between water level decreasing and time. Data collected from the pumping wells were not included in the linear fit, since the water level in these wells were fluctuating. The linear fit has a slope of -0.03 and R^2 of 0.95, as shown in Fig. 11. The intercept -5.76 was assumed to be the initial water level. The start of 2020 HT Test 9 was treated to be Day 0.

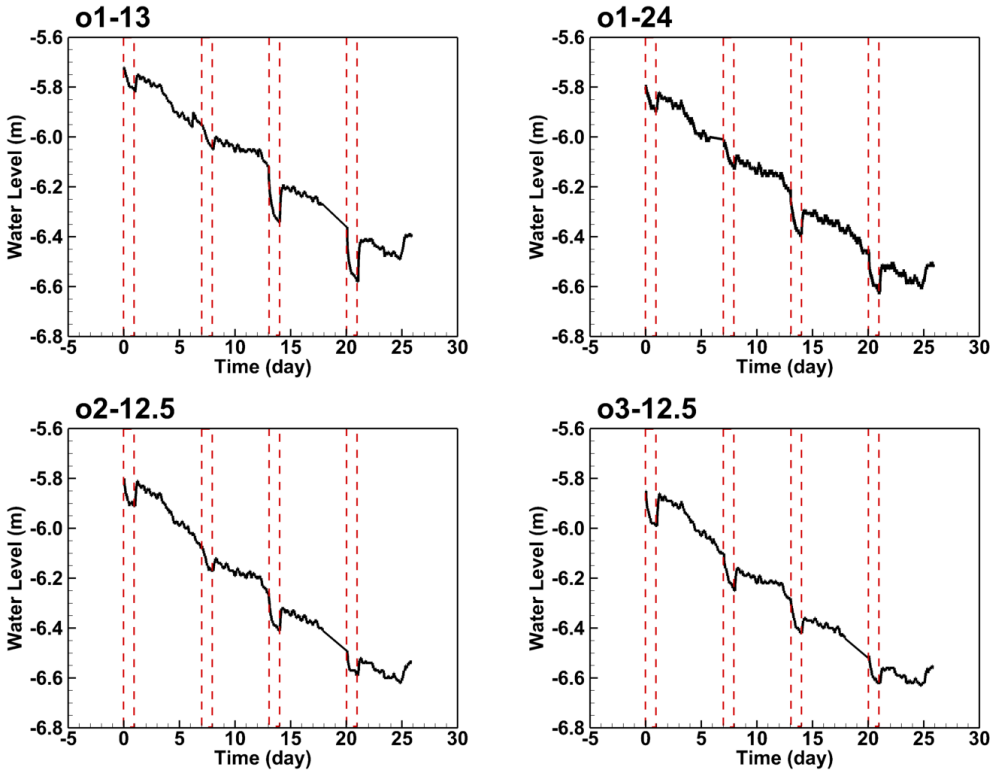

Fig. 10: *Water levels in observation wells o1, o2 and o3 during 2020 HT tests. The solid black line represents the water level and the red dashed lines represent the pumping period: Days 0 -1: Test 9; Days 7-8: Test 10; Days 13-14: Test 11; Days 20 -21: Test 12.*

Then, the water level was adjusted based on the fitted linear relationship. The adjusted water level of each observation port was calculated based on the equation:

$$WL_{adjusted} = WL_t - 0.03 \times t + (WL_{initial} - 5.76) \quad (1)$$

where $WL_{adjusted}$ is the adjusted water level $[L]$, WL_t is the original water level $[L]$ at time t, $WL_{initial}$ is the initial water level $[L]$ of the corresponding observation port and "-0.03" and "-5.76" are the slope and intercept of the best fit line.

The adjusted water level at 16 mbgs of Well c3 is shown as an example in Fig. 12.

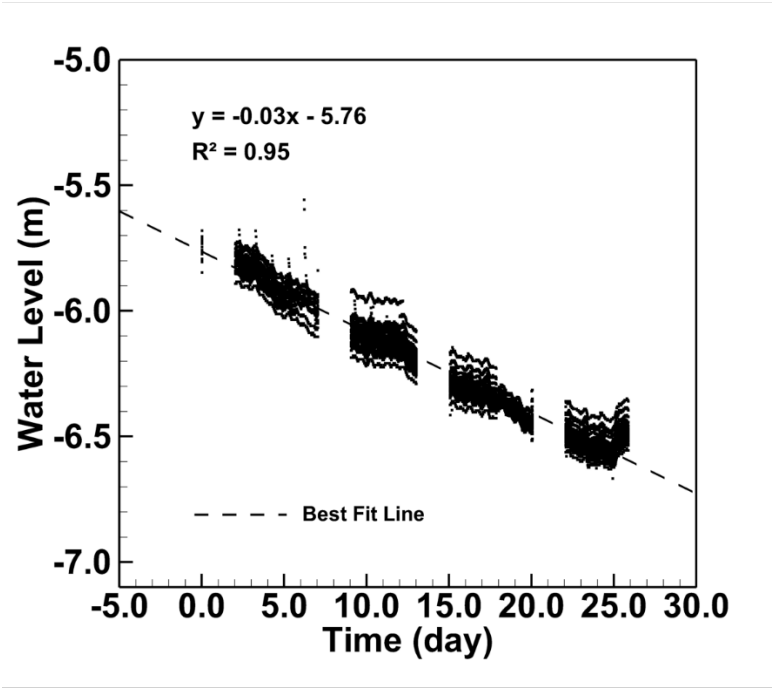

Fig. 11: *Linear fit between the water level and time. The black dashed line represents the best fit line and the solid black points represent*

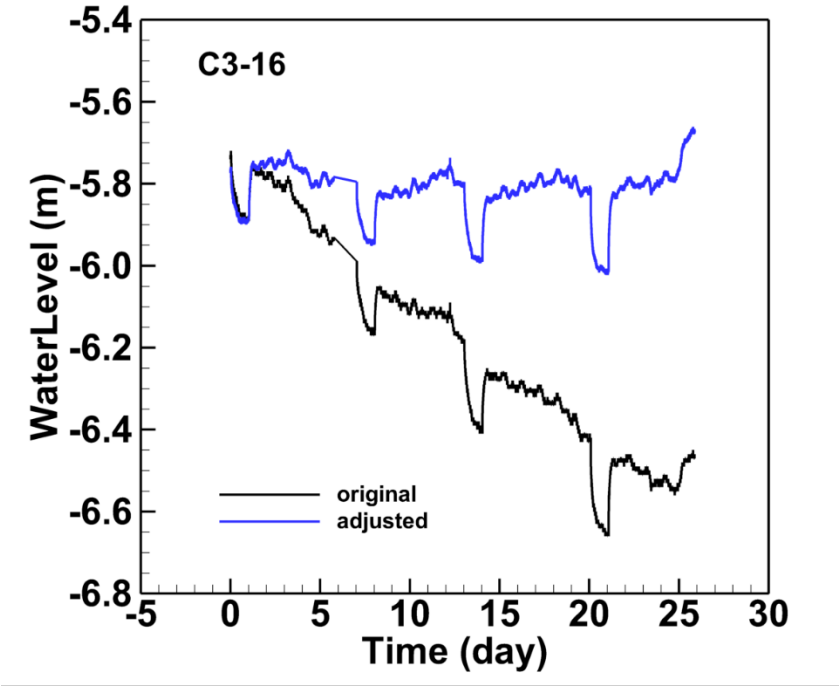

Fig. 12: *Adjusted and originally recorded water levels at 16 mbgs of Well C3. The solid black line represents the original recorded water level, while the solid blue line represents the adjusted water level.*

Based on the cross correlation analysis by Mao et al. (2013), the early time hydraulic data carry the most information about S_s, while the late time data carry the most information about the K. Thus, following the data sampling strategy suggested by Sun et al., (2013), five data points were selected from each observation ports for the model calibration of 2020 HT tests. That is, two data points were selected from the early time (5 and 20 minutes from the start of pumping), one from the intermediate time (400 minutes) and two from the late time (1000 and 1400 minutes).

4.3 Inverse Model Set-up and Inverse Cases

The field pumping tests conducted at the Narashino site during 2019 and 2020 were used to conduct THT analyses to capture the spatial distributions of K and S_s. The THT analyses were

carried out using the SSLE algorithm with the loop iteration scheme (SSLE_loop). Similar to the geological zonation model, the geostatistical model consisted of 4 horizontal layers (DC, Ds1, Ds2 and Ds3). The initial inputs for the SSLE_loop algorithm were based on the geological zonation approach Case 1 results. The corresponding values of K and S_s of different layers are listed in Table 12. Furthermore, anisotropy is neglected and not considered in the THT analyses, because (1) Case 1 of the geological zonation approach performs the best among the Creager's empirical solution, effective parameter and geological zonation approaches and (2) the SLE algorithm is not able to estimate the anisotropy of each element in K and S_s during the iterative inversion process.

Table 12: *Initial values of K and S_s input into the SSLE_loop code*.

Layers	K (m/min)	S_s (1/m)
Dc	7.58E-07	1.12E-03
Ds1	1.62E-03	3.67E-04
Ds2	2.62E-03	2.13E-04
Ds3	2.37E-03	8.60E-05

Several assumptions were made to determine initial conditions and boundary conditions of the groundwater model. The initial head is determined using the water level data collected in the a-, b- and c- series observation wells before HT tests started. Water level variation between each observation wells is disregarded, and a horizontal water level is assigned to the model at the beginning of HT tests by averaging the initial water level observed in a-, b- and c- series observation wells. In terms of constant boundary head conditions, the head along the constant head boundary was fixed during the forward and inverse simulation. It is calculated by averaging the head data collected in the o- series observations wells.

For the SSLE_loop code, the order of data included into the inversion algorithm may influence the calibration results. In particular, Illman et al. (2008) found out including larger signal to noise

(S/N) ratio data first in the HT analysis helped to obtain better results than including lower S/N ratio data. Following the guideline of Illman et al. (2008), Test 5 was included first into the SSLE_loop and Tests 2, 3, 1, 4 and 8 were included sequentially thereafter.

In total, 5 cases were analyzed to investigate a series of situations: (1) by comparing Cases 1 to 4, the impact of including more data can be investigated; and (2) by conducting Case 5, one can compare the K and S_s fields obtained after removing the cut-off walls to the ones when the cut-off walls are still present. Different numbers of data were selected for these cases. Case 1 has 135 data for model calibration and 145 data for validation of the estimated tomograms. Case 2 has 425 data for model calibration and 145 data for model validation. Cases 3 and 4 have 715 data and 695 data, respectively for model calibration. For model validation, Case 3 has 125 data and Case 4 has 145 data. In terms of Case 5, 290 data are used for model calibration and 110 data are used for model validation. The 5 THT cases and corresponding data number for both calibration and validation purposes are shown in Table 13.

Table 13: *Cases analyzed with the SSLE_loop code.*

		Calibration	Pumping Wells	Data Number	Validation	Pumping Wells	Data Number
2019	Case 1	Test 5	a1, c1	135	Test 4	c5	145
	Case 2	Test 5+2+3	a1, c1+ c1+ a5	425	Test 4	c5	145
	Case 3	Test 5+2+3+1+4	a1, c1+ c1+ a5+ a1+ c5	715	Test 8	a1, a5, c1, c5	125
	Case 4	Test 5+2+3+1+8	a1, c1+ c1+ a5+ a1+ a1, a5, c1, c5	695	Test 4	c5	145
2020	Case 5	Test 9+10+11	a5+ c5+ a1	290	Test 12	c1	110

4.4 Results from the THT Analyses

Inverse modeling of data from the HT tests was performed on the same PC with a six-core CPU and 16 GB of RAM for the effective parameter and geological zonation approaches. The model convergence time varied depending on the particular case. Case 1 with the fewest data took 42 hours to converge, while Case 3 with the most data took about 76 hours. With more data included in the SSLE_loop algorithm, the computational time increased significantly.

The estimated K tomograms of Cases 1 to 4 are shown in Figs. 13 (a), (c), (e) and (g), respectively. Fig. 13 (a) reveals that with only one test used for inversion, the K distribution is smooth, and a coarse picture of the heterogeneity pattern is mapped. With more tests included in the inversion process, more details emerge, as shown in Fig. 13 (c), (e) and (g). In particular, both the interlayer and intralayer heterogeneity are captured. In general, with geological information provided, all these four cases successfully captured the low K zones on the top of the model and the high K zones at the bottom of the model, which is a clay layer and a sand layer, respectively, based on our knowledge of site geology. Furthermore, two high K zones were found in Ds1 layer in all four cases, even though there are some differences in terms of the width and length of the sublayers.

The S_s tomogram was estimated simultaneously. In contrast to the K tomograms, the heterogeneity patterns of S_s are smoother, as shown in Fig. 13 (b), (d), (f) and (h). The difference between the clay layer (Dc) and sand layer (Ds1, Ds2 and Ds3) are significant, while little discrepancy exists among the sand layers. Different from the K tomograms, the sublayers in Ds1 were not shown in the S_s tomograms.

The 3D K and S_s tomograms obtained from Case 5 are shown in Fig. 14. For the K tomogram, the layer patterns are captured, with a low K zone being estimated at the top and relatively high K

zones beneath it. Within the Ds1 layer, there are two high K lenses located at 12 and 15 mbgs. For the S_s tomogram, the top high S_s zone is consistent with the geological information that a clay layer exists there. Yet, two high S_s zones are found in Ds1 at approximately 8 and 14 mbgs. This suggests that two clay or silt lenses might exist in the Ds1 layer. It is interesting to note that the K-tomograms obtained from Cases 1 to 4 also showed this pattern, whereas the same pattern does not show in the S_s-tomograms obtained from Cases 1 to 4. Results showing the convergence of each case are presented in Appendix E.

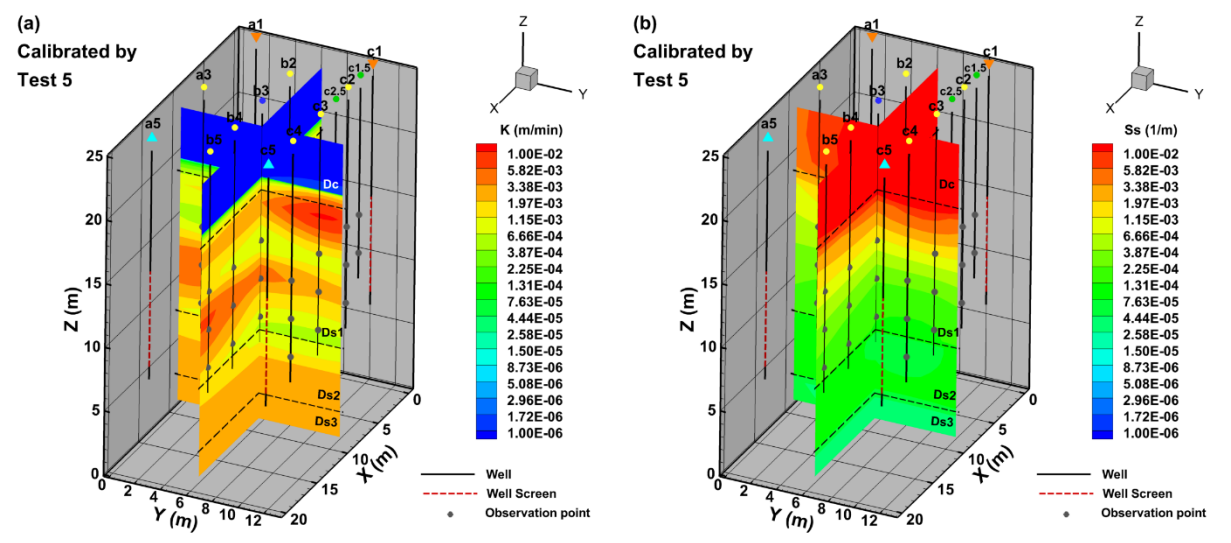

Fig. 13: *THT K-and S_s-tomograms from the 2019 HT Tests: (a) K-tomogram of Case 1 (Test 5); (b) S_s-tomogram of Case 1 (Test 5); (c) K-tomogram of Case 2 (Tests 5, 2, and 3); (d) S_s-tomogram of Case2 (Tests 5, 2, and 3); (e) K-tomogram of Case 3 (Tests 5, 2, 3, 1 and 4); (f) S_s-tomogram of Case 3 (Tests 5, 2, 3, 1 and 4); (g) K-tomogram of Case 4 (Tests 5, 2, 3, 1 and 8); (b) S_s-tomogram Case 4 (Tests 5, 2, 3, 1 and 8).*

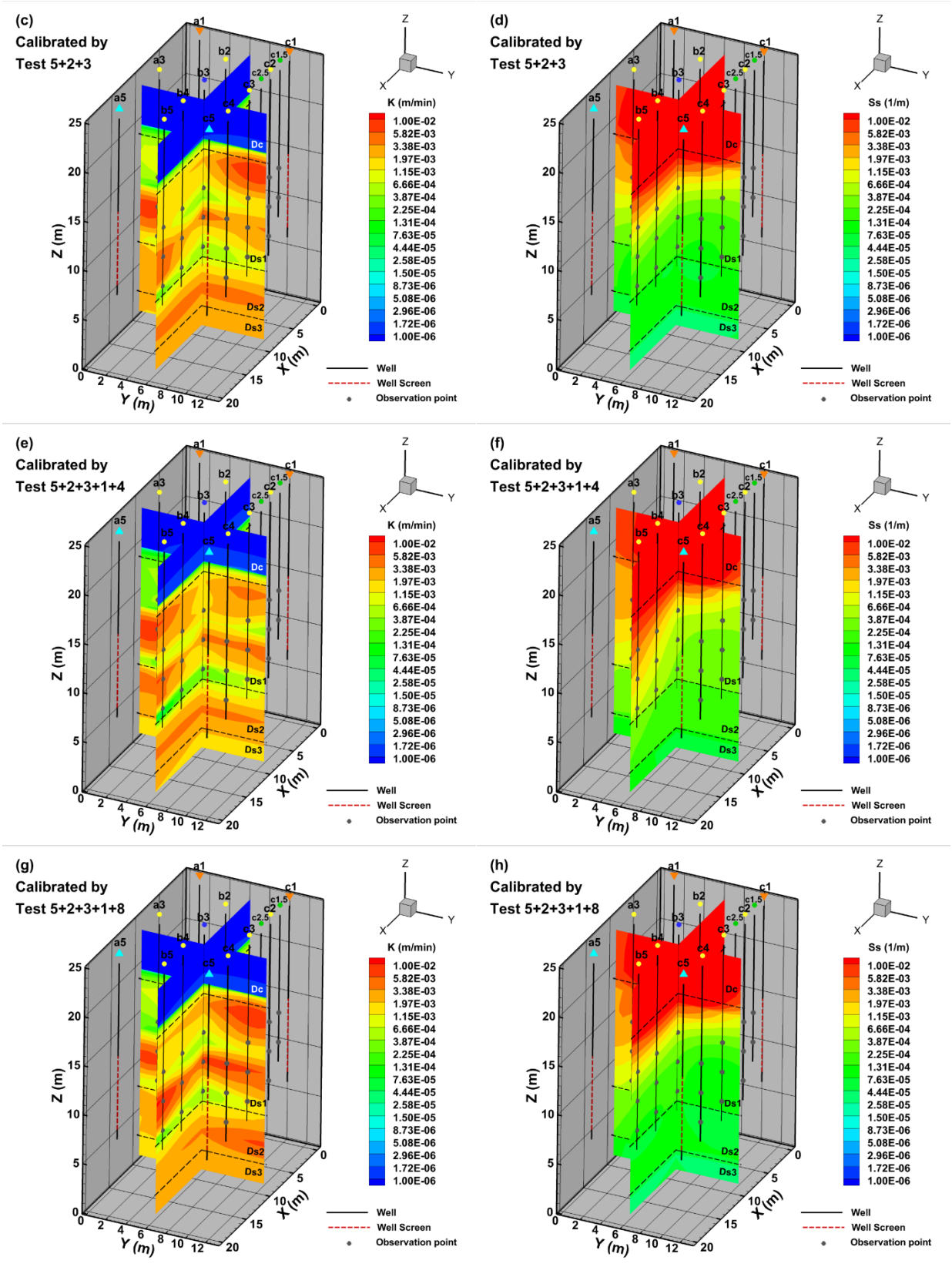

Fig. 13: *(continued)*

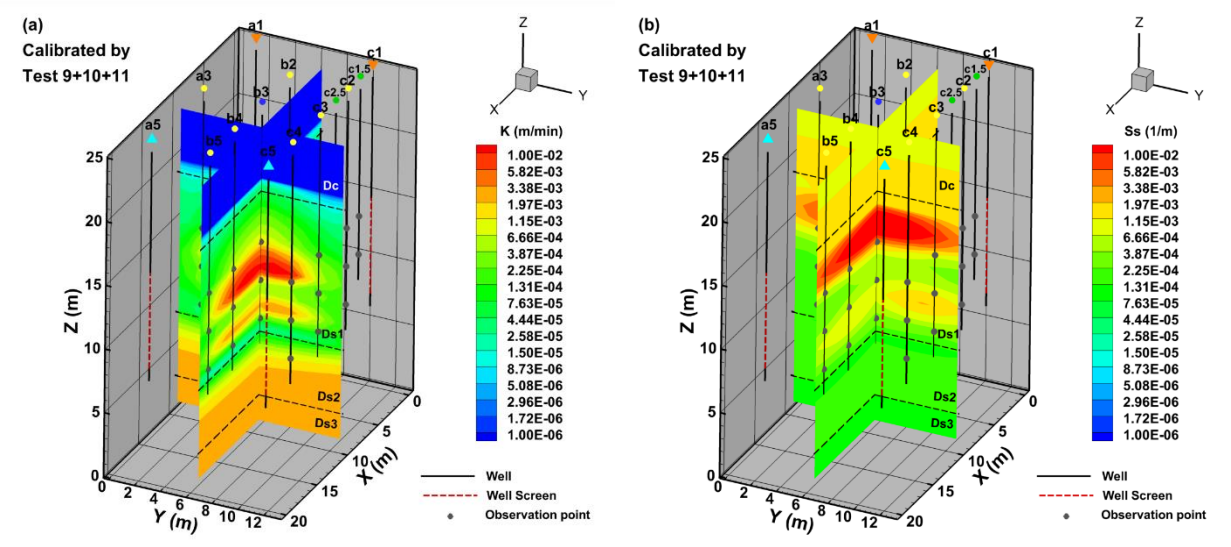

Fig. 14: *K-and S_s-tomograms obtained from Case 5 (by inverting Tests 9, 10 and 11 conducted in 2020).*

4.5 Evaluation of Model Calibration Performance

The scatter plots of simulated and observed drawdowns from model calibration are presented in Fig. 15. Examination of these scatter plots reveals that all these five cases are able to reproduce groundwater flow successfully with the geological information provided as prior information to the SSLE_loop code.

To evaluate the model calibration performance, L_1 and L_2 norms, the slope of the linear model fit, as well as the R^2 are calculated to quantitatively assess model calibration performance. Figs. 15 (a) to (e) show the scatter plots of simulated and observed drawdowns included in the calibration processes of Cases 1 to 5, respectively. For Figs. 15 (a) to (e), the linear model fit along with the L_1 and L_2 norms, as well as the R^2 are provided. In terms of the slope of linear model fit and the R^2, all cases performed similarly. The slopes vary from 0.94 to 0.98 and the R^2 values vary from 0.97 to 0.99, which all indicate that the difference between simulated and observed drawdowns are fairly small and the data points are concentrated around the best fit line.

Examination of Fig. 15 (e) reveals that the overall linear fit of the model has a slope of 0.94 and R^2 of 0.97. Comparison of the L_1 and L_2 norms reveals that, in general, the L_1 and L_2 norms vary within a small range and the values are in the same order of magnitude. Examination of Figs. 15 (a), (b) and (c) reveals that the model calibration performance does not improve significantly with more data included in the inversion in terms of L_1 and L_2 norms. Case 1 has the fewest data, yet the L_1 and L_2 norms are the smallest. Comparison of Cases 3 and 4 indicates that the calibration results of the data set with less noise tends to be better than the calibration results of the data set with more noise. The slope of linear model fit, L_1 and L_2 norms, as well as the R^2 of model calibration are summarized in Table 14.

Figs. 16 and 17 presents the observed drawdown versus simulated drawdown curves of HT Test 5 from calibration results of Cases 1 and 2, respectively. The other plots that compared observed versus simulated drawdowns of calibration process for the other HT tests and the other cases are provided in Appendix F (Figs. F1 to F15). Pressure transducer installed in Well b5 at a depth of 13 mbgs was broken during HT Test 5. Thus, no data was shown in the plot. Examinations of Figs. 16, 17 and Appendix F reveals that, although slight misfits exist at several observation wells, most calibrated drawdown fits well to the observed data for all Cases 1 to 5. The drawdown curves match better at early time than during late times (as shown in Wells b2 and b3).

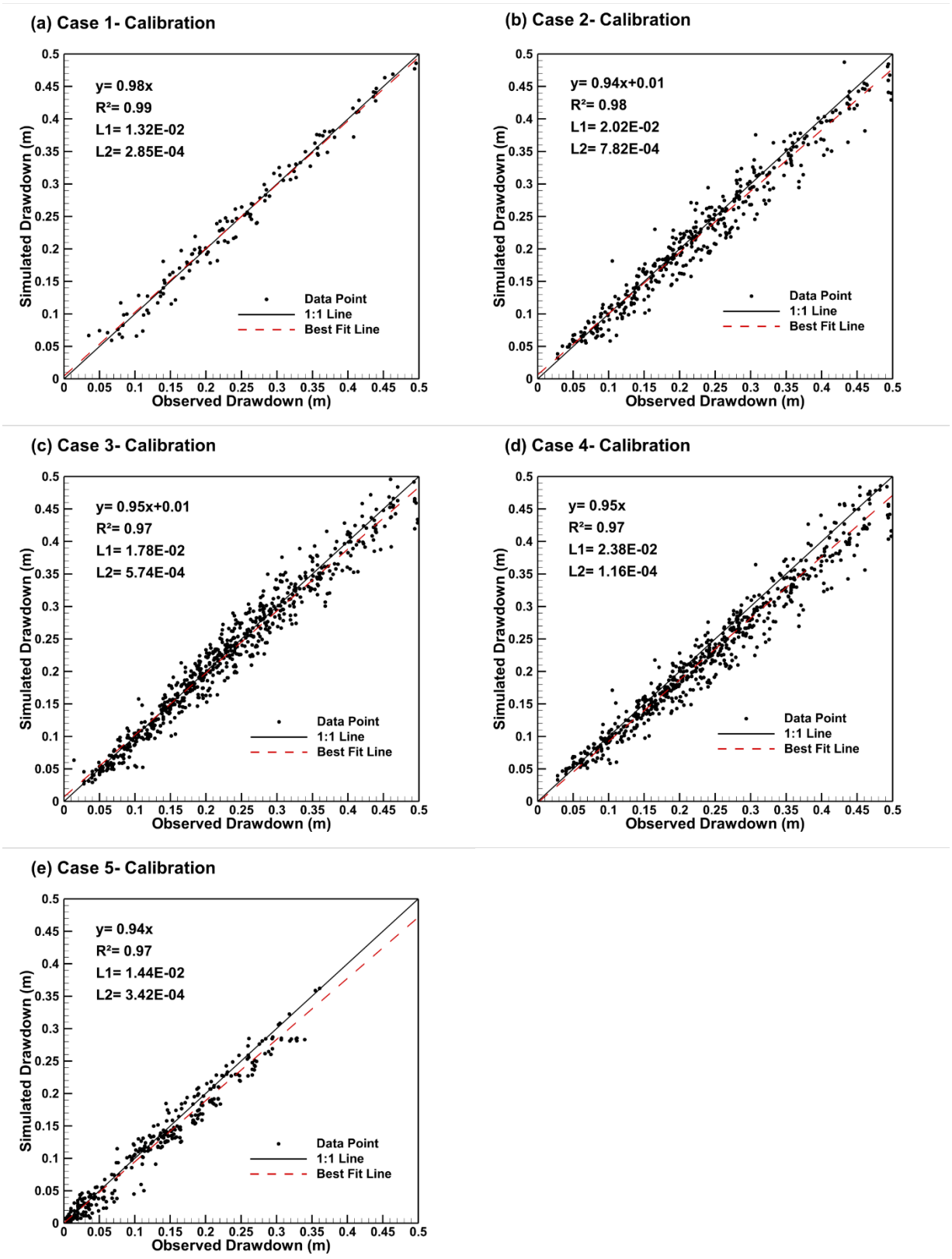

Fig. 15: *Scatter plots of the calibration results from Case 1 to Case 5: (a) Case 1; (b) Case 2; (c) Case 3; (d) Case 4; and (e) Case 5. The solid black line is the 1:1 line, while the red dashed line represents the best fit line of the data.*

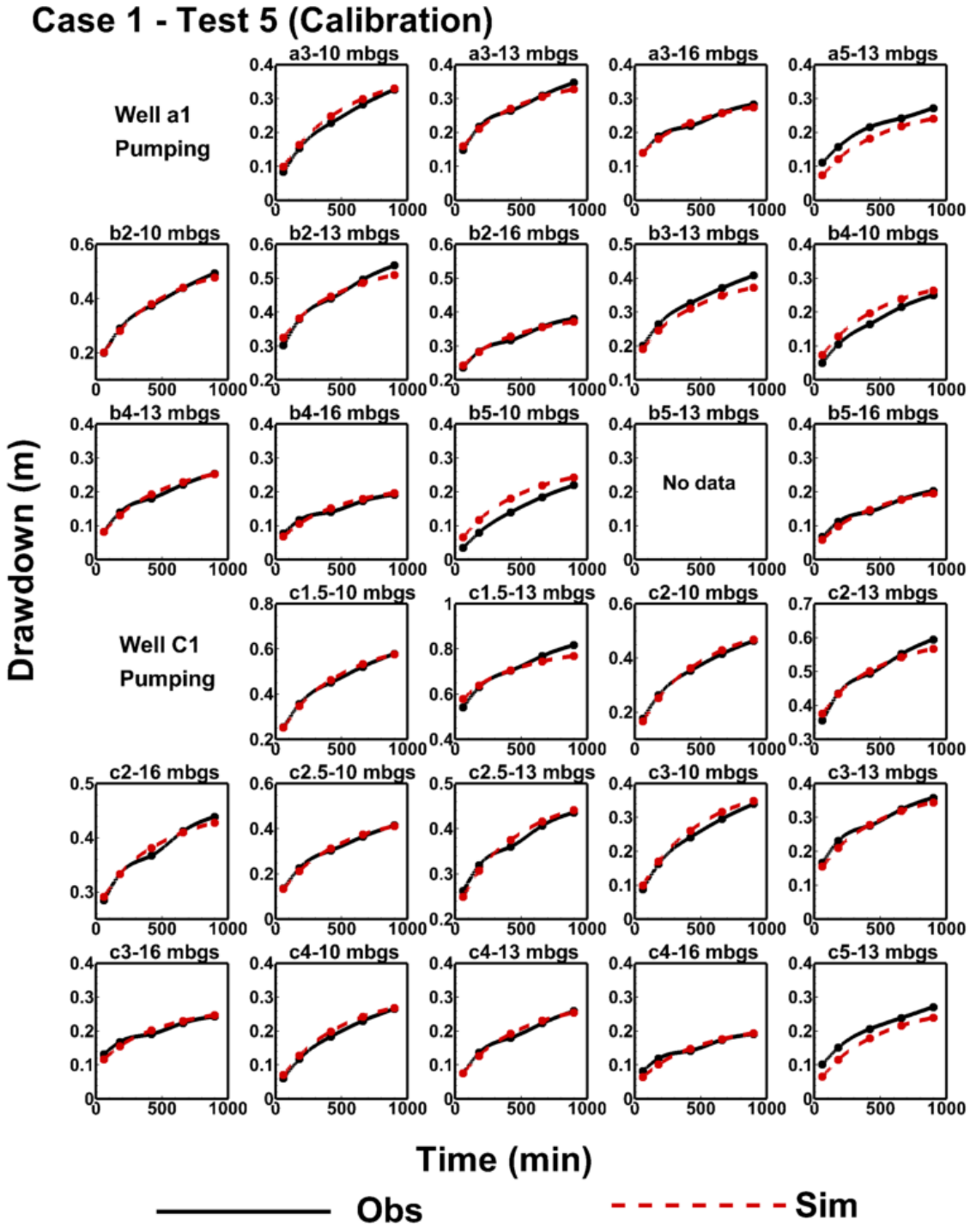

Fig. 16: *Observed and simulated drawdowns of Case 1 calibration results versus time curves at observation ports during HT Test 5 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 1. Note: "No data" indicates that data was not collected in that well during corresponding HT test. Pumping wells used in this HT test are listed in the plot as well.*

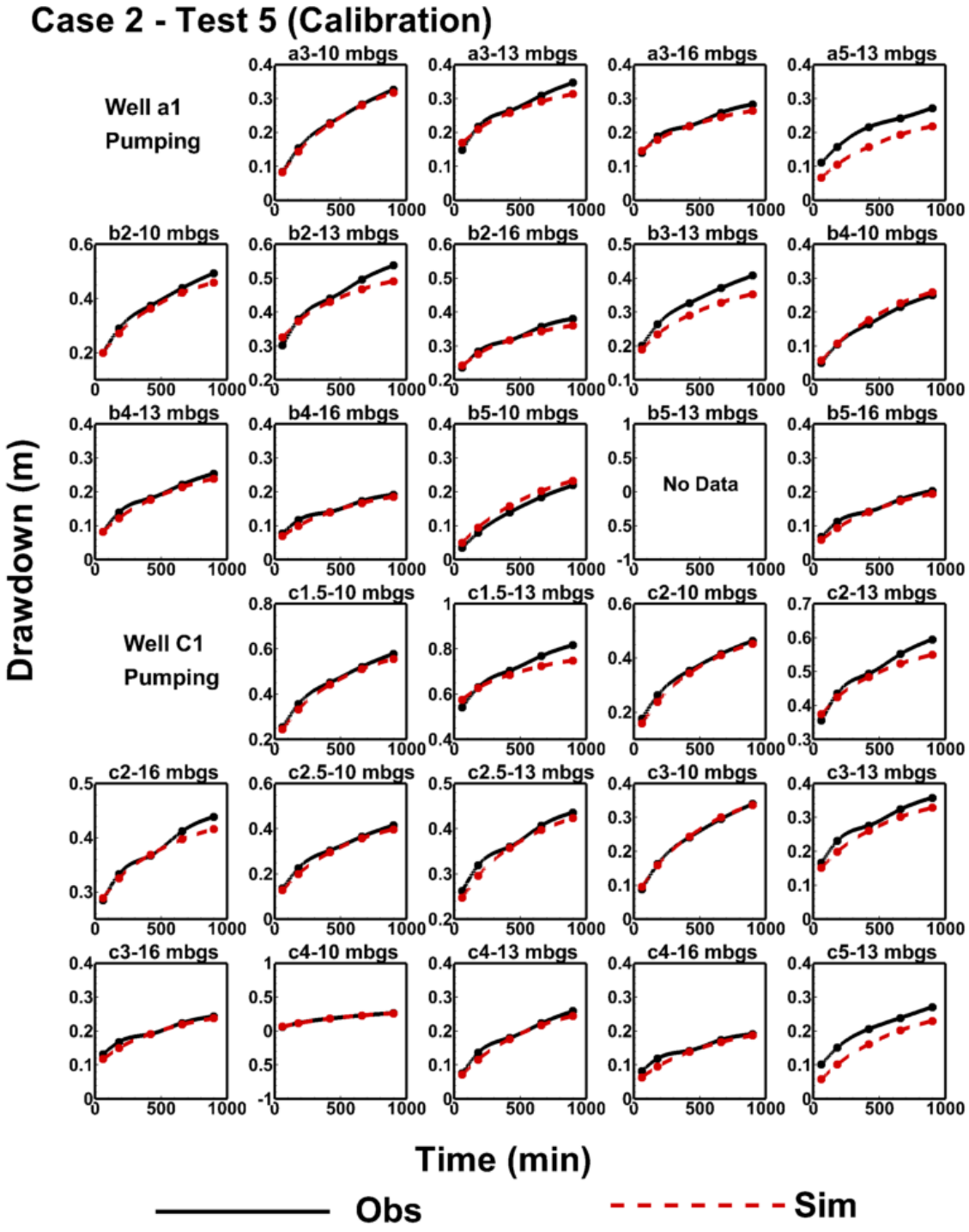

Fig. 17: *Observed and simulated drawdowns of Case 2 calibration results versus time curves at observation ports during HT Test 5 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 2. Note: "No data" indicates that data was not collected in that well during corresponding HT test. Pumping wells used in this HT test are listed in the plot as well.*

4.6 Prediction of Transient Drawdowns of an Independent Pumping Test

After the calibration of Cases 1 to 5, the obtained K and S_s fields were used for the forward simulation of an independent pumping/injection test to validate the estimated tomograms. Cases 1, 2 and 4 were validated with Test 4, while Cases 4 and 5 were validated with Tests 8 and 12, respectively. Forward simulations of the independent pumping test are conducted using the SSLE_loop algorithm with the same settings of Model 2. The validation results are presented in Fig. 18 for all five cases.

Through the visual assessment of the scatter plots for validation, the data mostly concentrate around the 1:1 line for all 5 cases, even though some outliers exist. As expected, there is a slight bias and the L_1 and L_2 norms are larger than those compared to the results from model calibration (Fig. 15). Comparing Figs. 18 (a) and (b), one is able to find out that the results improved in terms of the slope, R^2, L_1 and L_2 norms, which means the difference between observed and simulated drawdowns are decreased by including more data in the inverse model. However, when Test 8, which has a much larger influence on the boundary condition, was included into the SSLE_loop code, the validation results deteriorated significantly. This indicates that simply involving more data into the SSLE_loop might not be able to improve the validation results. The data also need to be accurate enough. Validation results of Case 3 show that the early-time fit is better than the late-time fit, and the results are not improved significantly, compared to Cases 1 and 2. This might be caused by the fact that the tests used for validation for Case 3 are different from those of Cases 1 and 2. The assumed constant head boundaries tend to cause more bias for late-time data fitting. In reality, the hydraulic head at the boundary is decreasing during the pumping test period thus this behavior should be considered more accurately for improved parameter estimates. The slope of linear model fit, L_1, L_2 norms as well as the $R^2$of model validation are summarized in Table 14.

Drawdown versus time curves that compared observed and simulated drawdowns from validation results are shown in Fig. 19 for Case 1. The validation results of the other 4 cases are listed in Appendix G. Examination of these figures reveals that, the difference between observed and simulated drawdowns are more significant, compared to the calibration results. Yet, overall, the observed groundwater flow behavior of independent tests is predicted quite well using the calibrated K and S_s tomograms. Besides, comparing the validation results of Cases 1 to 5 shows that the drawdown misfit of Case 4 is the most obvious, which is consistent with results shown in the scatter plots (Fig. 18). Since Case 4 performed worst in the calibration, it is expected that its validation results may not as accurate as the other cases.

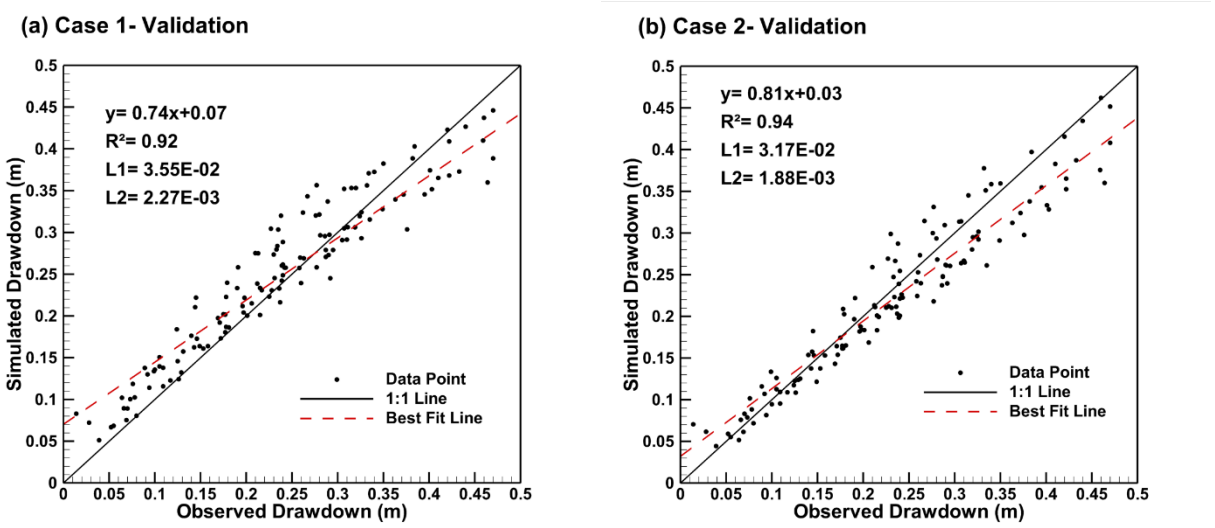

Fig. 18: *Scatter plots of the validation results from Case 1 to Case 5: (a) Case 1; (b) Case 2; (c) Case 3; (d) Case 4; and (e) Case 5. The solid black line is the 1:1 line, while the red dashed line represents the best fit line of the data.*

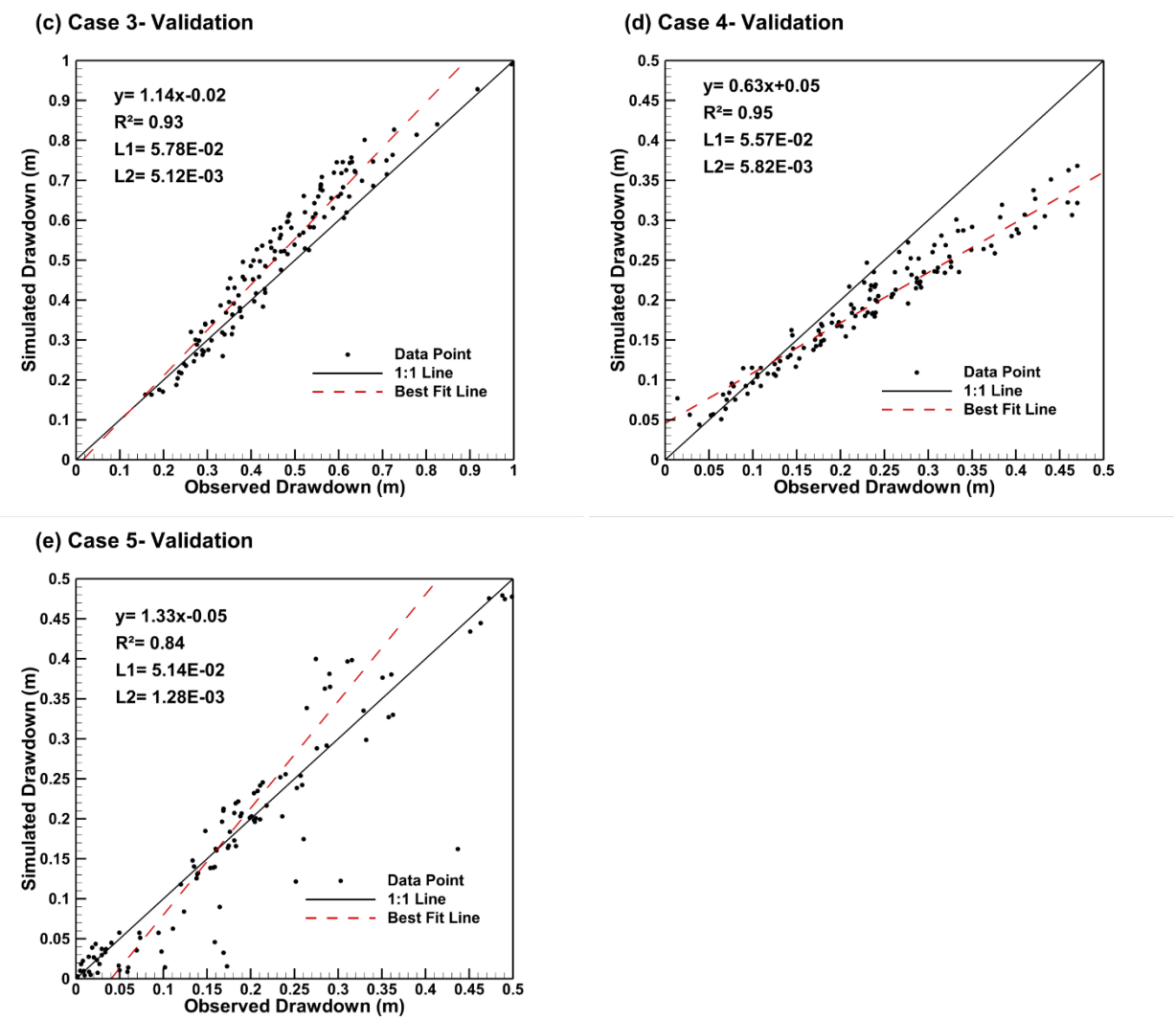

Fig. 18: *(continued)*

Table 14: *Summary of calibration and validation performance.*

	Calibration				Validation			
	Slope	R^2	L_1	L_2	*Slope*	R^2	L_1	L_2
Case 1	0.98	0.99	1.32E-02	2.85E-04	0.74	0.92	3.55E-02	2.27E-03
Case 2	0.94	0.98	2.02E-02	7.82E-04	0.81	0.94	3.17E-02	1.88E-03
Case 3	0.95	0.97	1.78E-02	5.74E-04	1.14	0.93	5.78E-02	5.12E-03
Case 4	0.95	0.97	2.38E-02	1.16E-04	0.63	0.95	5.57E-02	5.82E-03
Case 5	0.94	0.97	1.44E-02	3.42E-04	1.33	0.84	5.14E-02	1.28E-03

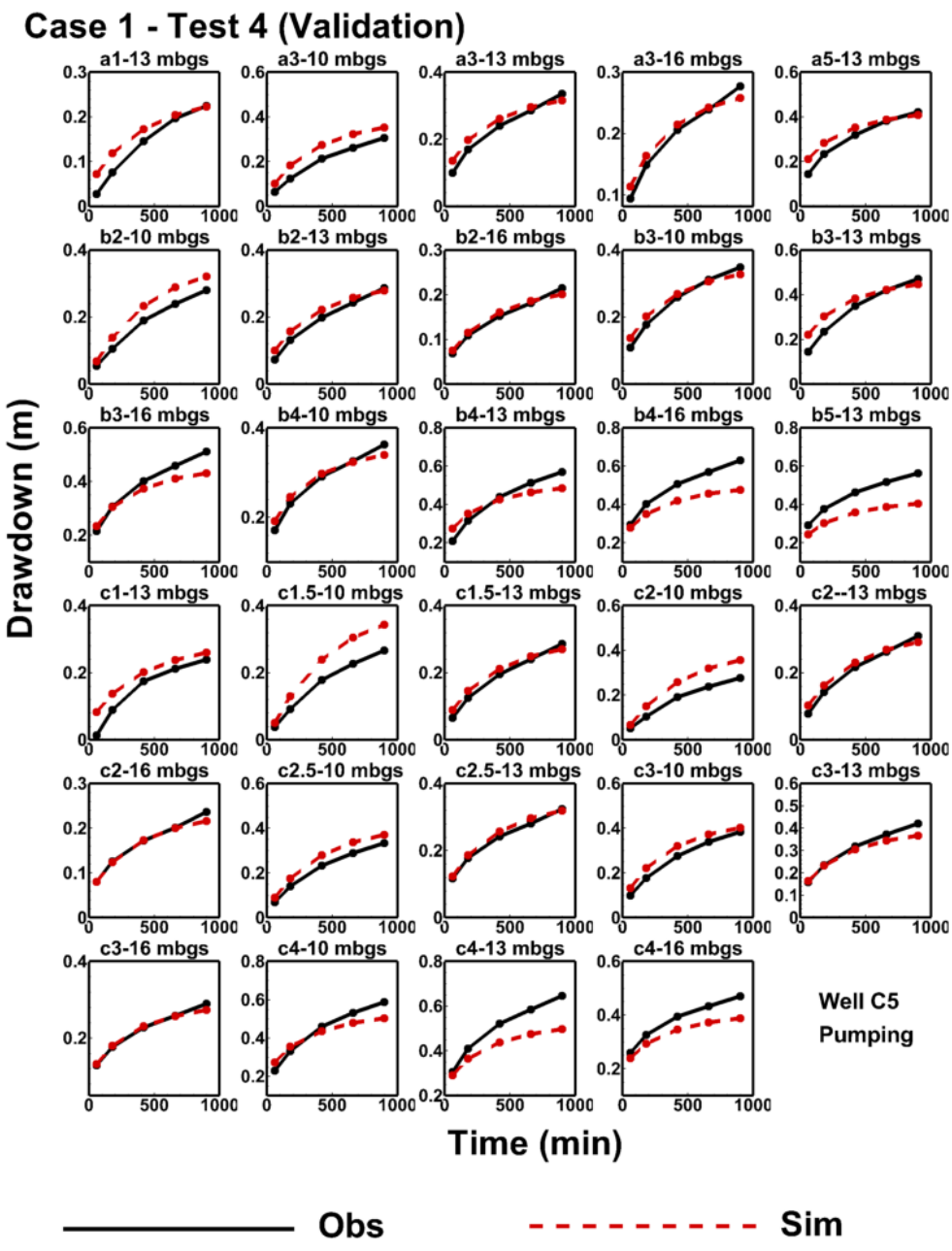

Fig. 19: *Observed and simulated drawdowns of Case 1 validation results versus time curves at observation ports during HT Test 4 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the simulated drawdowns of Case 1. Pumping wells used in this HT test are listed in the plot as well.*

4.7 Comparison of the 2019 versus 2020 HT tomograms

The tomograms estimated from 2020 HT tests, Case 5, were then compared with the best 2019 HT results, Case 3. By comparing the tomograms visually, both 2019 and 2020 HT analyses successfully captured the low *K* zone and high S_s zone on the top. The *K* values for the Dc layers of 2019 HT are close to those of 2020 HT, which are about $8 \times 10^{-7}\, m/min$. However, the S_s values of the high S_s zone delineated from the 2020 HT survey were smaller than that of mapped during the 2019 HT survey. In terms of the sandy layers beneath the clay layer, although both Cases 3 and 5 captured two high *K* zones in Ds1, the high *K* zones in Ds1 captured in the 2020 HT survey were thinner than the ones shown from the 2019 HT survey. Besides, two relatively high S_s zones were captured in Ds1 layer by 2020 HT, while this same pattern was not shown in the 2019 S_s tomogram. As for Ds2 and Ds3 layers, Cases 3 and 5 had similar results for *K*- and S_s-tomograms, respectively. Furthermore, scatter plots were used to compare the *K* and S_s fields. The low *K* zone on the top of the model with values around $8 \times 10^{-7}\, m/min$ (around -14 for ln *K*) concentrates around the bottom left of the scatter plot. Yet, the other values spread over a wide range, instead of concentrating around the 1:1 line. This suggests that some discrepancy exists between the 2019 and 2020 HT results. In terms of S_s comparison, the difference between the 2019 and 2020 HT surveys was not that significant as more data concentrate around the 1:1 line. Yet, as circled in blue on Fig. 20, the estimated S_s values of the clay layer from the 2020 HT survey are much smaller those from the 2019 HT survey.

This discrepancy might be caused by the assumptions made during the simulation. The four side boundaries were treated to be constant boundary, which is determined based on the observation wells outside the cut-off walls. The hydraulic heads during the HT tests were averaged and used for the constant head boundaries. The difference in the boundaries might cause bias in

the parameter estimation. Another reason might be that during the adjustment of the water level, the initial water level was treated to be horizontal, which means that the initial water level at each well was the same. Yet, the initial water level among the wells at Narashino Ste were different from each other. Moreover, the estimated tomograms obtained from SSLE_loop algorithm are able to be influenced by the order of tests included into the inversion, the number of data included into the inversion, the number of tests included into the inversion and the data selection strategy (Zhu and Yeh, 2005; Illman et al., 2008; Berg and Illman, 2011b; Zhang et al., 2019). These issues might potentially influence the results of this study. It is important to note that the HT results from Case 3 utilized five pumping tests and the tests were conducted with the cut-off walls at the experimental site, while Case 5 utilized three pumping tests conducted in 2020 after the removal of the cut-off walls. Besides, 2019 HT included 715 data points in total, while 2020 HT included 219 data. The order of tests included into these two cases were also different. For Case 3, the order of pumping tests is pumping at Wells a1 and c1 (Test 5); pumping at Well c1 (Test 2); pumping at Well a5 (Test 3); pumping at Well a1 (Test 1) and pumping at Well c5 (Test 4). While for Case 5, the order is pumping at Well a5 (Test 9), pumping at Well c5 (Test 10) and pumping at Well a1 (Test 11). The data were also selected from different times of the drawdown curves. Thus, these factors can also lead to the discrepancies in the estimated tomograms.

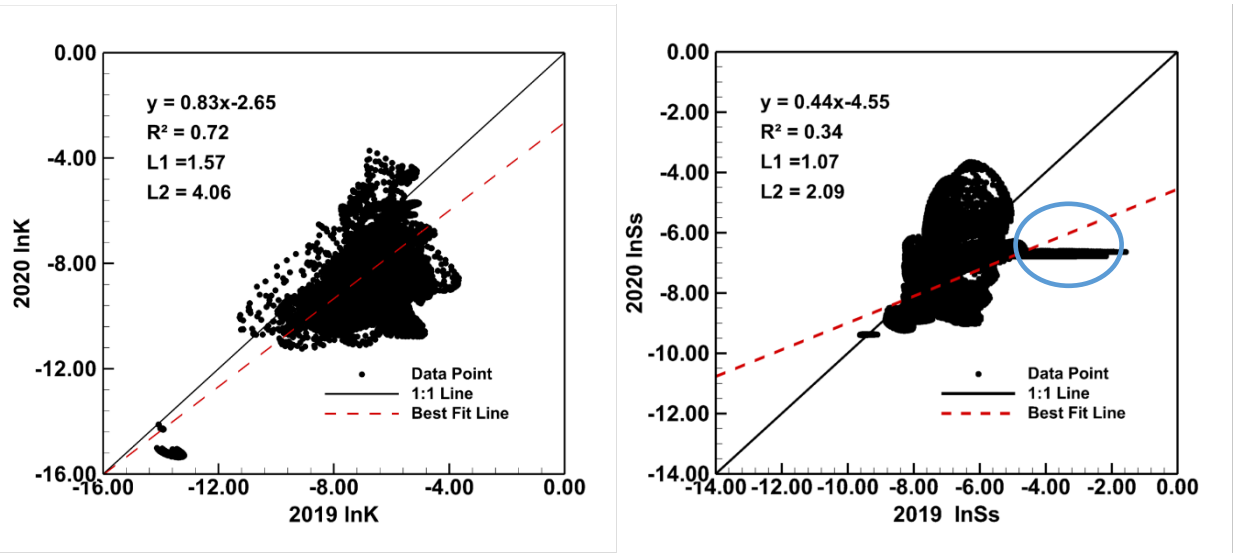

Fig. 20: *Scatter plots of the estimated K and S_s tomograms from the 2019 and 2020 HT surveys. The solid black line represents the 1:1 line, while the dashed red line represents the best fit line.*

Chapter 5

Conclusions and Recommendations

5.1 Conclusions

In this study, we conducted simulations of eight step-up pumping tests with different approaches, the Creager's empirical solution (shown in Equation 7), the effective parameter approach and the geological zonation approach, to investigate and evaluate their usefulness on delineating parameter fields and reproducing the groundwater flow field by coupling HGS with PEST. Twelve hydraulic tomography tests were monitored during 2019 and 2020. Data collected from 17 wells at the Narashino Site operated by Takenaka corporation in Japan were analyzed using THT analysis with the SSLE_loop code to characterizing the three-dimensional K and S_s fields and to investigate the influence of data points included in the inversion on the HT performance as well as to investigate the reproducibility of HT on delineating the parameter fields.

Our study resulted in the following conclusions:

1. As expected, the Creager's empirical solution has the worst performance when matching the simulated drawdowns to observed drawdowns. After estimating the effective hydraulic parameters by PEST, the results improved in terms of the quality of the fits. Yet, the effective parameter approach ignores the stratum/geological information and provides zero resolution in terms of heterogeneity.
2. The geological zonation approach can yield better results when plotting the drawdown versus time and scatter plots that comparing simulated and observed drawdowns. However, the

estimation of K and S_s in the Dc and Ds3 layers is not very reasonable, due to the lack of drawdown data in these two layers.

3. Although treating the medium to be isotropic (geological zonation approach; Case 1) performs better than Case2, consideration of the anisotropy does change the estimation of K and S_s value, and it improved the performance of the drawdown fit at several wells. Yet, consideration of the model calibration performance and the limitation of SLE algorithm, impact of the anisotropy is ignored, and the medium is treated to be isotropic for the HT analyses at the Narashino Site.
4. HT analyses using the SSLE_loop code are able to delineate three-dimensional K and S_s fields or tomograms with reasonable details that can reproduce the groundwater flow fields and used to predicted groundwater flow of an independent pumping event. With more tests or data sets involved in the code, the tomograms are able to show more details and the prediction of the independent tests that are not used in calibration effort becomes more accurate. In addition, when a data set which is considered to have noise or errors is input into SSLE_loop during the inversion process, the performance of the validation will deteriorate, as expected. When the errors are not obviously neglectable, the SSLE_loop code might not converge.
5. The removal of cut-off walls at the Narashino Site changes the boundary conditions for forward and inverse modeling of groundwater flow at the site. Several assumptions were made when inverting the data: (1) the water level was assumed to be horizontal before the test started; (2) the data were adjusted to eliminate the influence of the unknown pumping events or ambient groundwater gradient; (3) the boundary was assumed to be constant during the simulation, and (4) the water level change at the boundaries from different hydraulic tests was ignored.

6. The comparison of results from the 2019 and 2020 HT surveys revealed some discrepancies in the estimated values, even when the same geological information was provided, and the same code was used. It is important that some differences in results from 2019 and 2020 are expected due to the following factors: (1) different number of tests are analyzed (five tests for 2019 vs three tests for 2020); (2) different number of data sets are incorporated into the inverse model (715 for 2019 vs 219 for 2020); (3) different order of tests included in SSLE_loop; and (4) different testing conditions (e.g., differences in boundary conditions and background water levels, etc.).

5.2 Recommendations for Future Studies

Some recommendations for the future studies are listed below:

1. To further improve the 2020 HT analysis, the boundary condition should be considered more carefully. Instead of using a constant head boundary condition throughout the one-day pumping period, a shorter period can be simulated to minimize the influence of changing boundary heads during the simulation period.
2. The data were included into the SSLE_loop sequentially in this study. In the future, one should consider inverting data sets simultaneously;
3. The comparison of HT analyses of 2019 and 2020 tests could be further improved by utilizing the same number of pumping tests with data selected from similar times. The resulting tomograms can then be assessed with a number of independently conducted hydraulic tests and tracer tests conducted at the site;
4. In this study, the impact of open boreholes within the simulation domain was disregarded. In the future, the influence of open boreholes, packer locations, and other site features that could

potentially impact testing conditions should be more accurately incorporated into the model and studied to assess whether it is necessary to include such features in HT analyses.

5. In this study, a coarse description of geology was used as an initial guess in THT inversions. However, finer scale descriptions of the geology are also available at the site. It will be very informative to study the influence of resolution in geological characterization on HT results.

References

Batu, V. (1998) Aquifer Hydraulics: A Comprehensive Guide to Hydrogeologic Data Analysis. John Wiley and Sons, Inc., New York.

Berg, S. J., and Illman, W. A. (2011a). Capturing aquifer heterogeneity: Comparison of approaches through controlled sandbox experiments. *Water Resources Research*, *47*(9), 1–17. https://doi.org/10.1029/2011WR010429

Berg, S. J., and Illman, W. A. (2011b). Three-dimensional transient hydraulic tomography in a highly heterogeneous glaciofluvial aquifer-aquitard system. *Water Resources Research*, *47*(10). https://doi.org/10.1029/2011WR010616

Berg, S. J., and Illman, W. A. (2012). Improved predictions of saturated and unsaturated zone drawdowns in a heterogeneous unconfined aquifer via transient hydraulic tomography: Laboratory sandbox experiments. *Journal of Hydrology*, *470–471*, 172–183. https://doi.org/10.1016/j.jhydrol.2012.08.044

Berg, S. J., and Illman, W. A. (2013). Field study of subsurface heterogeneity with steady-state hydraulic tomography. *Ground Water*, *51*(1), 29–40. https://doi.org/10.1111/j.1745-6584.2012.00914.x

Berg, S. J., and Illman, W. A. (2015). Comparison of Hydraulic Tomography with Traditional Methods at a Highly Heterogeneous Site. *Ground Water*, *53*(1), 71–89. https://doi.org/10.1111/gwat.12159

Chen, J. L., Wen, J. C., Jim Yeh, T. C., Andrew Lin, K. Y., Wang, Y. L., Huang, S. Y., Ma, Y., Yu, C. Y., and Lee, C. H. (2019). Reproducibility of hydraulic tomography estimates and their predictions: A two-year case study in Taiwan. *Journal of Hydrology, 569*(October 2018), 117–134. https://doi.org/10.1016/j.jhydrol.2018.11.064

Cooper J r., H. H., and Jacob, C. E. (1946). A generalized graphical method for evaluating formation constants and summarizing well-field history. *Transactions, American Geophysical Union*, *27*, 526–534.

Creager, W.P., Justin, J.D., Hinds, J. (1945): *Engineering for dams, Earth, Rock-fill, Steel and Timber Dams, III,* Wiley, New York, 648–649.

Doherty, J. (2005). PEST: Model independent parameter estimation. Fifth edition of user manual. *Watermark Numerical Computing*.

Google. (n.d.-a). [Location of Narashino, Chiba, Japan]. Retrieved November 26, 2020, from https://earth.google.com/web/@35.65535979,139.5636889,78.7304918a,3042405.0033527 6d,35y,0h,0t,0r.

Google. (n.d.-b). [Location of Narashino Site]. Retrieved November 26, 2020, from https://earth.google.com/web/@35.70684739,140.07264415,33.8518367a,518.01081633d,3 5y,0h,0t,0r

Illman, W. A., Berg, S. J., and Zhao, Z. (2015). Should hydraulic tomography data be interpreted using geostatistical inverse modeling? A laboratory sandbox investigation. *Water Resources Research*. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Illman, W. A., Craig, A. J., and Liu, X. (2008). Practical issues in imaging hydraulic conductivity through hydraulic tomography. *Ground Water*, *46*(1), 120–132. https://doi.org/10.1111/j.1745-6584.2007.00374.x

Illman, W. A., Liu, X., Takeuchi, S., Yeh, T. C. J., Ando, K., and Saegusa, H. (2009). Hydraulic tomography in fractured granite: Mizunami Underground Research site, Japan. *Water Resources Research*, *45*(1), 1–18. https://doi.org/10.1029/2007WR006715

Illman, W. A., Zhu, J., Craig, A. J., and Yin, D. (2010). Comparison of aquifer characterization approaches through steady state groundwater model validation: A controlled laboratory sandbox study. *Water Resources Research*, *46*(4), 1–18. https://doi.org/10.1029/2009WR007745

Luo, N., Zhao, Z., Illman, W. A., and Berg, S. J. (2017). Comparative study of transient hydraulic tomography with varying parameterizations and zonations: Laboratory sandbox investigation. *Journal of Hydrology*, *554*, 758–779. https://doi.org/10.1016/j.jhydrol.2017.09.045

Mao, D., Yeh, T. C. J., Wan, L., Lee, C. H., Hsu, K. C., Wen, J. C., and Lu, W. (2013). Cross-correlation analysis and information content of observed heads during pumping in unconfined aquifers. *Water Resources Research*, *49*(2), 713–731. https://doi.org/10.1002/wrcr.20066

Mao, D., Yeh, T. C. J., Wan, L., Wen, J. C., Lu, W., Lee, C. H., and Hsu, K. C. (2013). Joint interpretation of sequential pumping tests in unconfined aquifers. *Water Resources Research*, *49*(4), 1782–1796. https://doi.org/10.1002/wrcr.20129

Rovey II, C. W. (1998). Digital simulation of the scale effect in hydraulic conductivity. *Hydrogeology Journal*, *6*(2), 216–225.

Rovey II, C. W., and Cherrkauer, D. S. (1995). Scale dependency of hydraulic conductivity measurements. *Ground Water*, *33*(5).

Sharmeen, R. (2011). *Hydraulic Tomography and Trichloroethene Dissolution in a Fractured Dolostone : Small Scale Laboratory Experiments*.

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 Resources Research*, *22*(13), 2069–2082. https://doi.org/10.1029/WR024i007p01209

Sun, R., Yeh, T. C. J., Mao, D., Jin, M., Lu, W., and Hao, Y. (2013). A temporal sampling strategy for hydraulic tomography analysis. *Water Resources Research*, *49*(7), 3881–3896. https://doi.org/10.1002/wrcr.20337

Theis, C. V. (1935). The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using ground-water storage. *Transactions, American Geophysical Union*, *16*, 519–524. https://doi.org/10.1017/CBO9781107415324.004

Therrien, R., McLaren, R. G., Sudicky, E. A., and Panday, S. M. (2010). HydroGeoSphere. A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. *Groundwater Simulations Group*, 457.

Tokyo Soil Research (2015), [(Tentative name) Geotechnical investigation (primary investigation) report accompanying the development of in-situ purification equipment for contaminated soil using a heating purification agent]. (No.125010033).

Wu, C. M., Yeh, T. C. J., Zhu, J., Tim, H. L., Hsu, N. S., Chen, C. H., and Sancho, A. F. (2005). Traditional analysis of aquifer tests: Comparing apples to oranges? Water Resources Research, 41(9), 1–12. https://doi.org/10.1029/2004WR003717

Xiang, J., Yeh, T. C. J., Lee, C. H., Hsu, K. C., and Wen, J. C. (2009). A simultaneous successive linear estimator and a guide for hydraulic tomography analysis. *Water Resources Research*, *45*(2), 1–14. https://doi.org/10.1029/2008WR007180

Xin, T. (2018). *Impact of geological conceptualization on the calibration of groundwater models at the Mannheim Wellfield, Kitchener, Ontario* (Vol. 10, Issue 2).

Yeh, T. C. J., and Liu, S. (2000). Hydraulic tomography: Development of a new aquifer test method. *Water Resources Research*, *36*(8), 2095–2105. https://doi.org/10.1029/2000WR900114

Yeh, T. C. J., and Zhang, J. (1996). A geostatistical inverse method for variably saturated flow in the vadose zone. *Water Resources Research*, *32*(9), 2757–2766. https://doi.org/10.1029/96WR01497

Yeh, T. J., Jin, M., and Hanna, S. (1996). An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields. *Water Resources Research*, *32*(1), 85–92.

Zha, Y., Yeh, T. C. J., Illman, W. A., Tanaka, T., Bruines, P., Onoe, H., and Saegusa, H. (2015). What does hydraulic tomography tell us about fractured geological media? A field study and synthetic experiments. *Journal of Hydrology*, *531*, 17–30. https://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., and Wen, J. C. (2016). An Application of Hydraulic Tomography to a Large-Scale Fractured Granite Site, Mizunami, Japan. *Groundwater*, *54*(6), 793–804. https://doi.org/10.1111/gwat.12421

Zha, Y., Yeh, T. C. J., Illman, W. A., Zeng, W., Zhang, Y., Sun, F., and Shi, L. (2018). A Reduced-Order Successive Linear Estimator for Geostatistical Inversion and its Application in Hydraulic Tomography. *Water Resources Research*, *54*(3), 1616–1632. https://doi.org/10.1002/2017WR021884

Zhang, J., and Yeh, T. C. J. (1997). An iterative geostatistical inverse method for steady flow in the vadose zone. *Water Resources Research*, *33*(1), 63–71. https://doi.org/10.1029/96WR02589

Zhang, Y., Wu, C., Hu, B. X., Yeh, T. C. J., Hao, Y., and Lv, W. (2019). Fine characterization of the effects of aquifer heterogeneity on solute transport: A numerical sandbox experiment. *Water*, *11*(11), 1–12. https://doi.org/10.3390/w11112295

Zhao, H. (2020). *The Importance of Fracture Geometry and Matrix Data on Transient Hydraulic Tomography in Fractured Rocks : Analyses of Synthetic and Laboratory Rock Block Experiments*.

Zhao, Z., and 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. *Journal of Hydrology*, *544*, 640–657. https://doi.org/10.1016/j.jhydrol.2016.12.004

Zhao, Z., and Illman, W. A. (2018). Three-dimensional imaging of aquifer and aquitard heterogeneity via transient hydraulic tomography at a highly heterogeneous field site. *Journal of Hydrology*, *559*, 392–410. https://doi.org/10.1016/j.jhydrol.2018.02.024

Zhao, Z., Illman, W. A., and Berg, S. J. (2016). On the importance of geological data for hydraulic tomography analysis: Laboratory sandbox study. *Journal of Hydrology*, *542*, 156–171. https://doi.org/10.1016/j.jhydrol.2016.08.061

Zhao, Z., Illman, W. A., Yeh, T.-C. J., Berg, S. J., and Mao, D. (2015). Validation of hydraulic tomography in an unconfined aquifer: A controlled sandbox study. *Water Resources Research*. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Zhu, J., and Yeh, T. C. J. (2005). Characterization of aquifer heterogeneity using transient hydraulic tomography. *Water Resources Research*, *41*(7), 1–10. https://doi.org/10.1029/2004WR003790

Zhu, J., and Yeh, T. C. J. (2006). Analysis of hydraulic tomography using temporal moments of drawdown recovery data. *Water Resources Research*, *42*(2), 1–11. https://doi.org/10.1029/2005WR004309

Appendix A

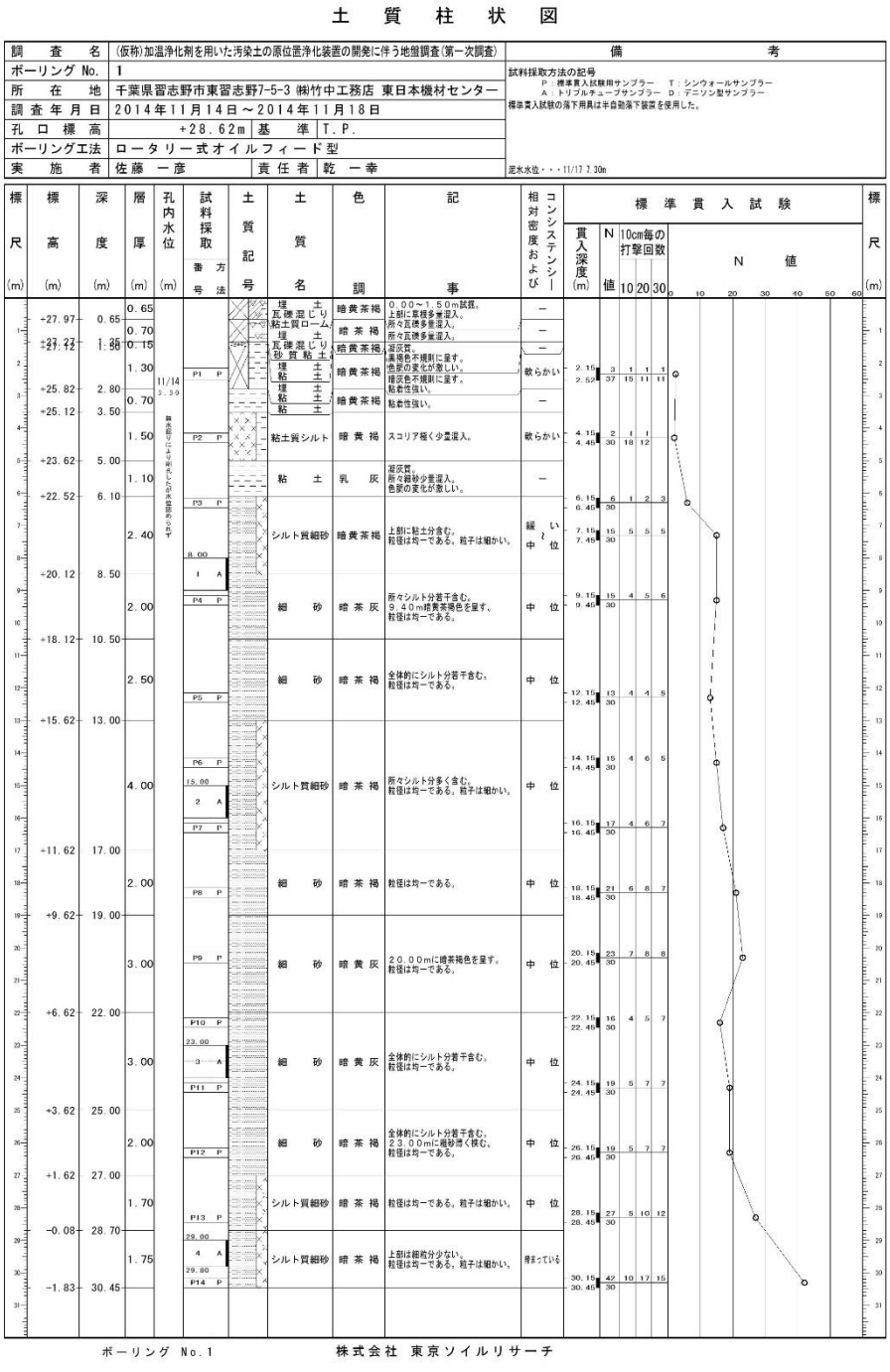

Fig. A1: *Borehole logs at Well o3.*

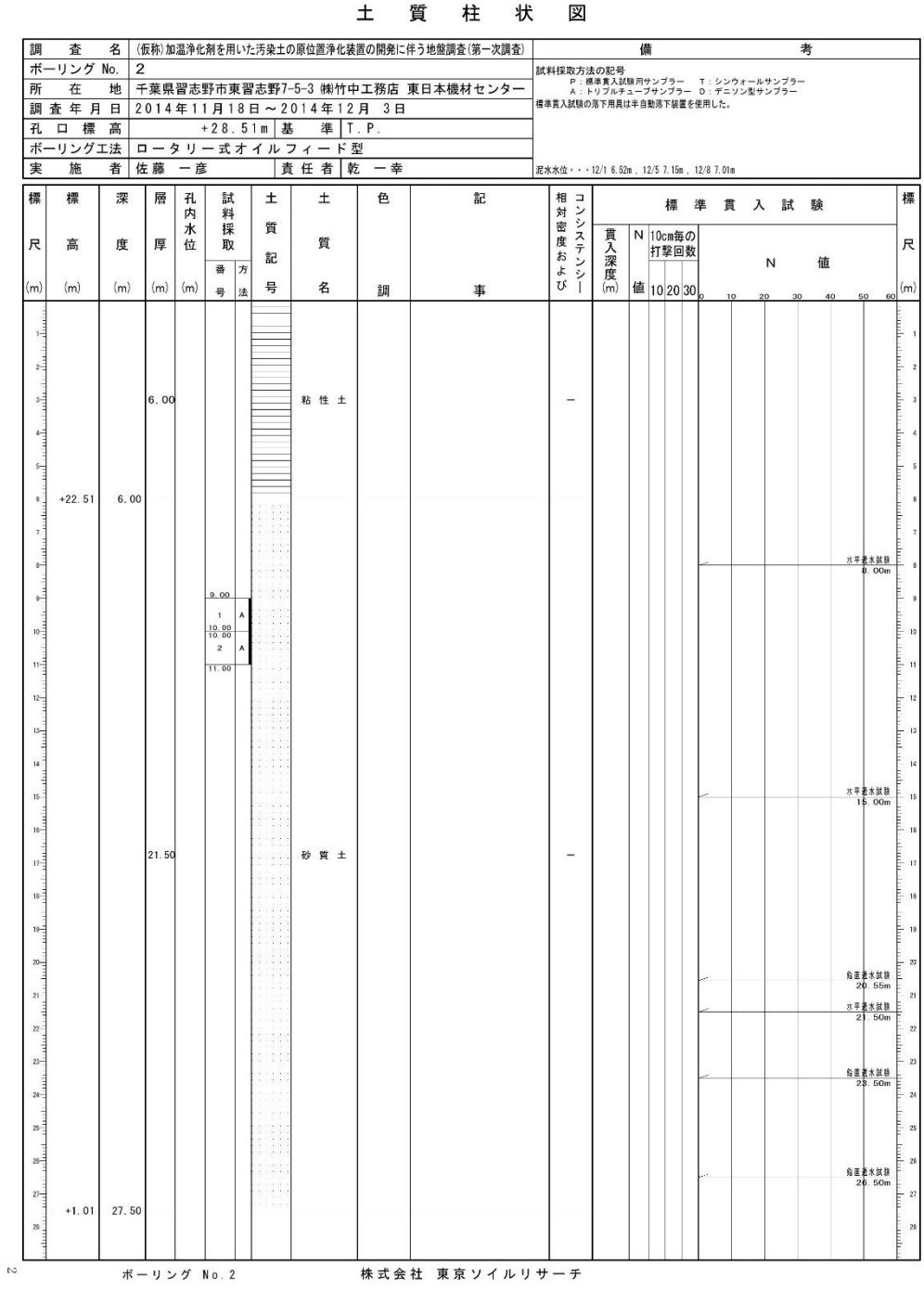

Fig. A2: *Borehole logs at Well o2.*

Appendix B

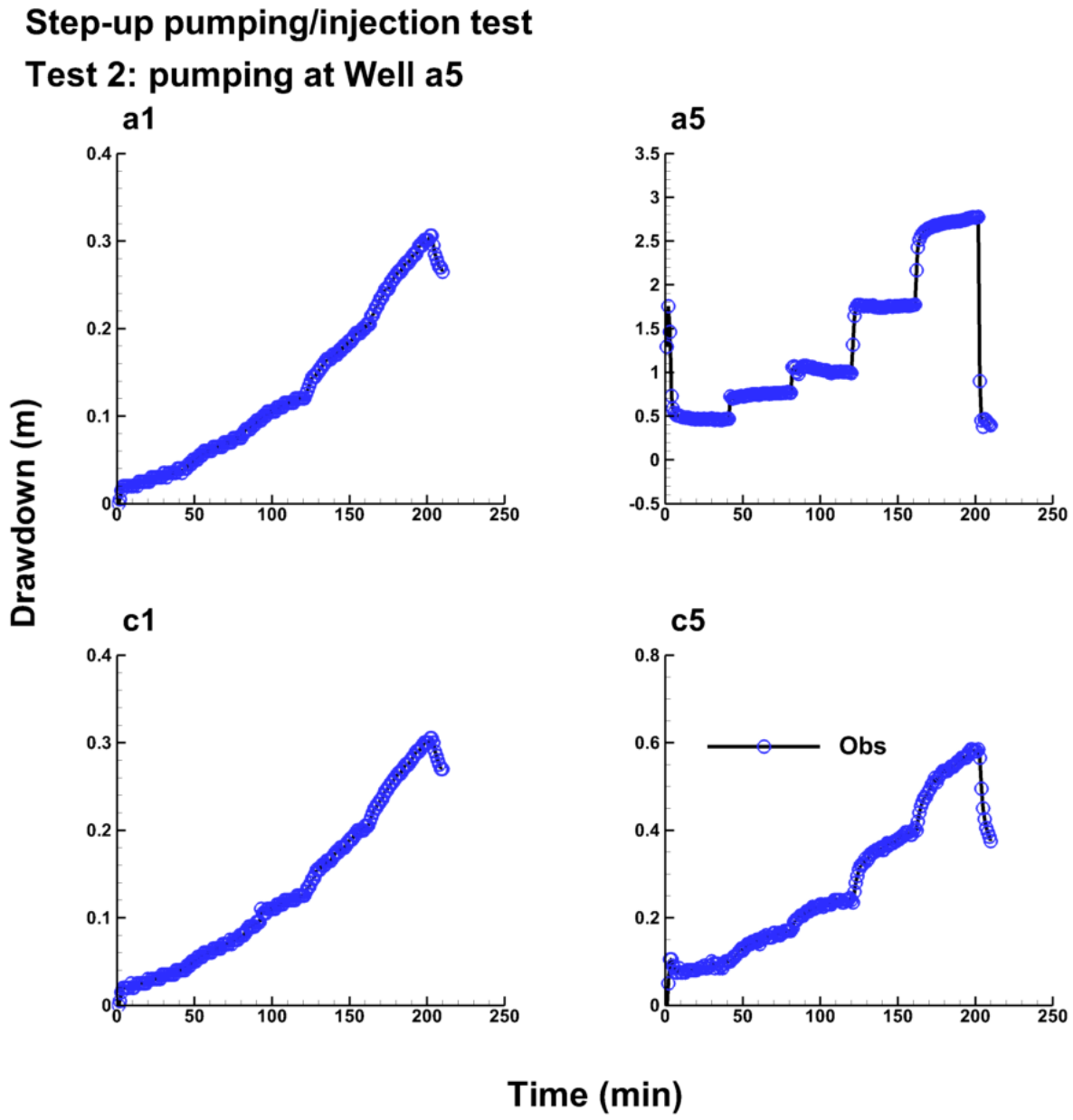

Fig. B1: *Observed drawdown curves of step-up pumping/injection Test 2 (pumping at Well a5). Open circles indicate the drawdown data used for inverse modeling.*

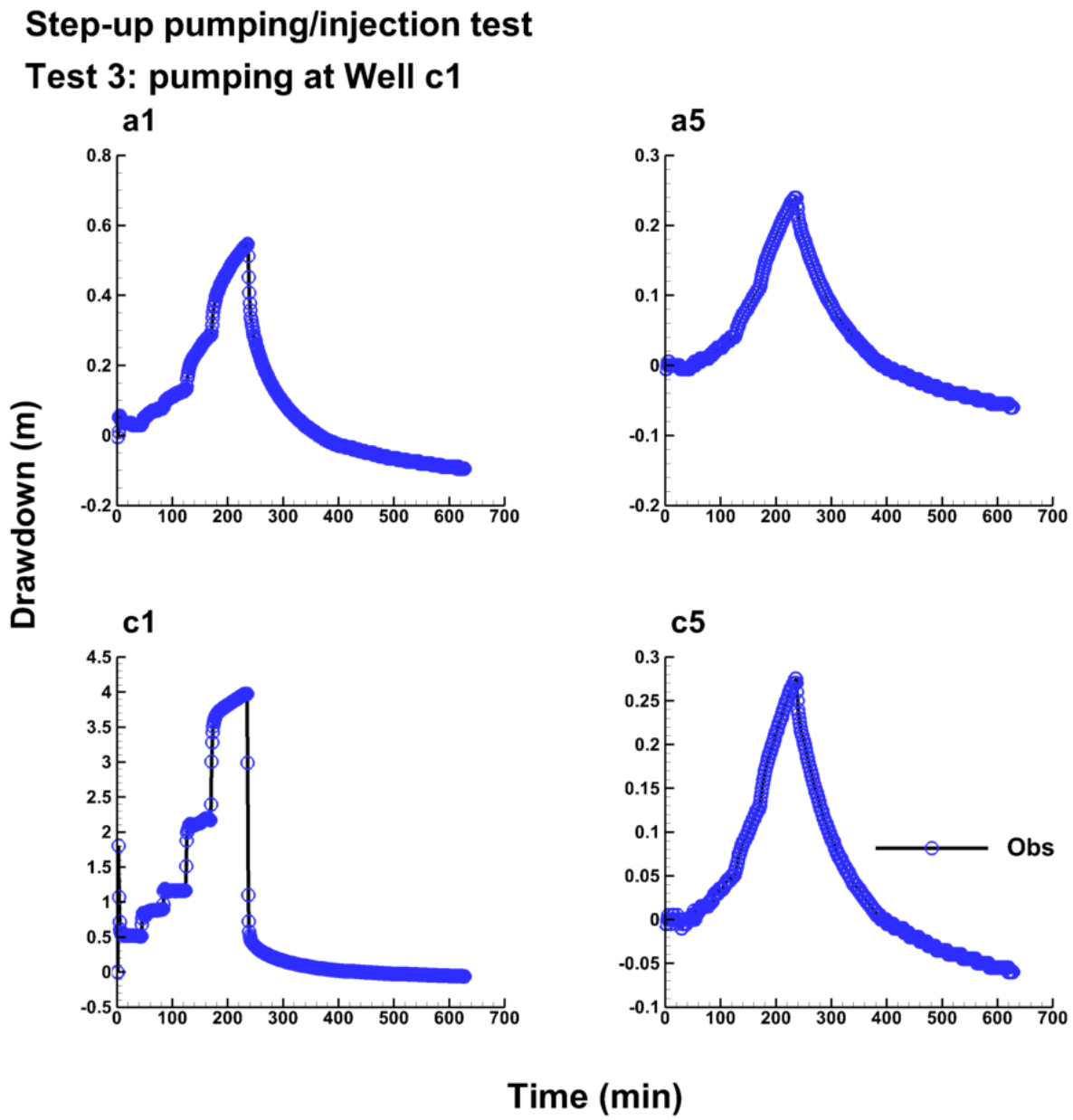

Fig. B2: *Observed drawdown curves of step-up pumping/injection Test 3 (pumping at Well c1). Open circles indicate the drawdown data used for inverse modeling.*

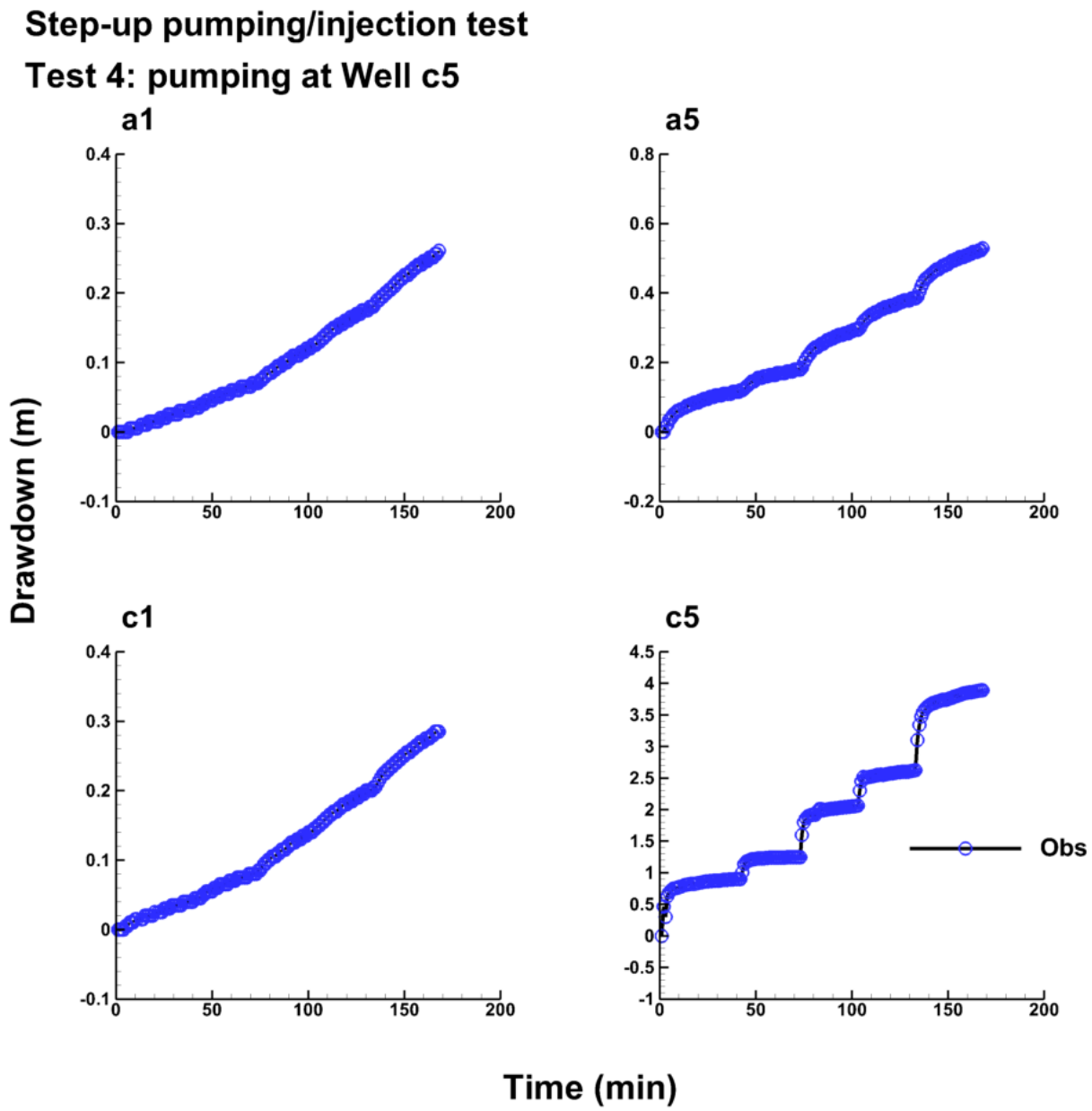

Fig. B3: *Observed drawdown curves of step-up pumping/injection Test 4 (pumping at Well c5). Open circles indicate the drawdown data used for inverse modeling.*

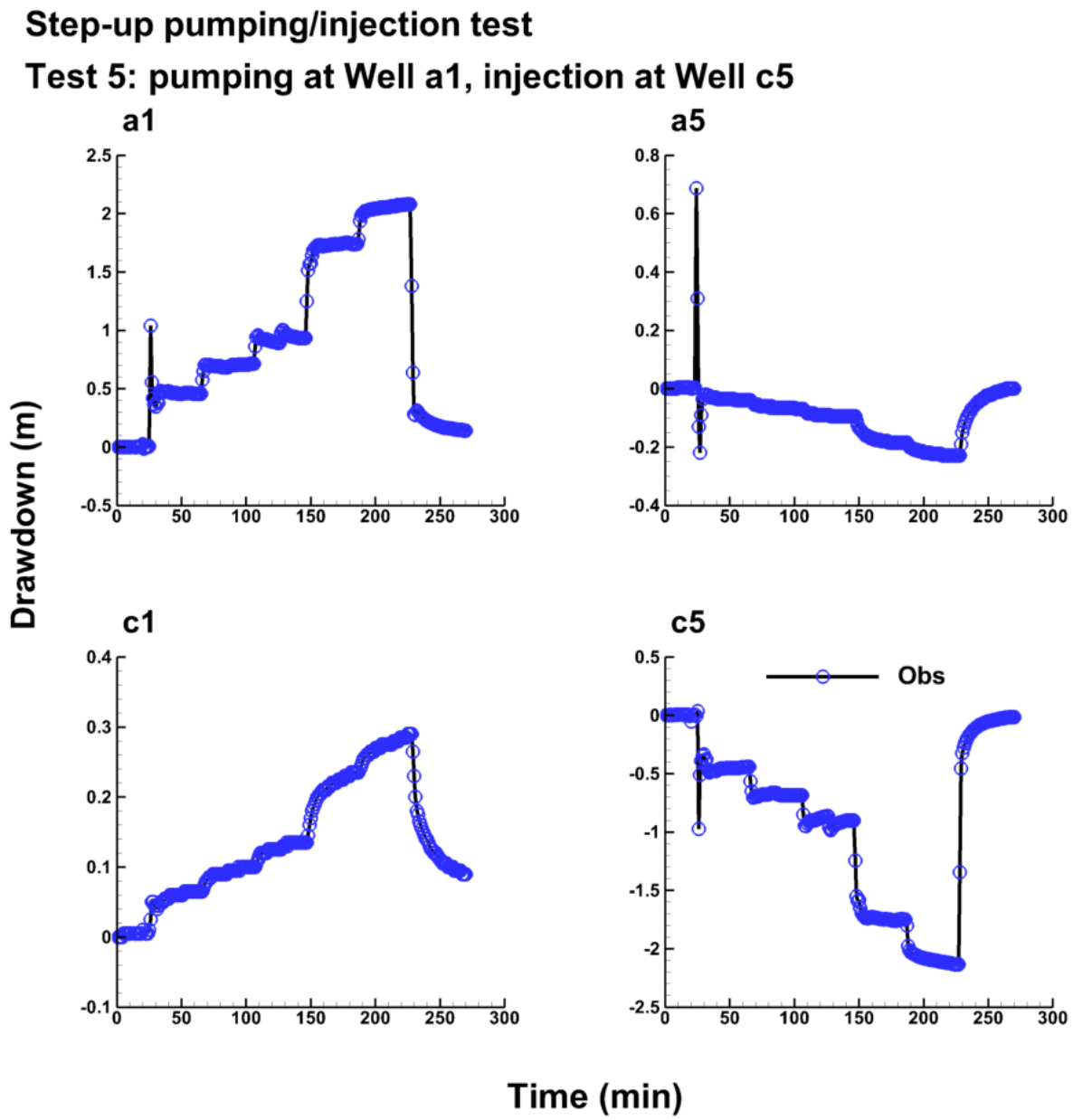

Fig. B4: *Observed drawdown curves of step-up pumping/injection Test 5 (pumping at Well a1, injection at Well c5). Open circles indicate the drawdown data used for inverse modeling.*

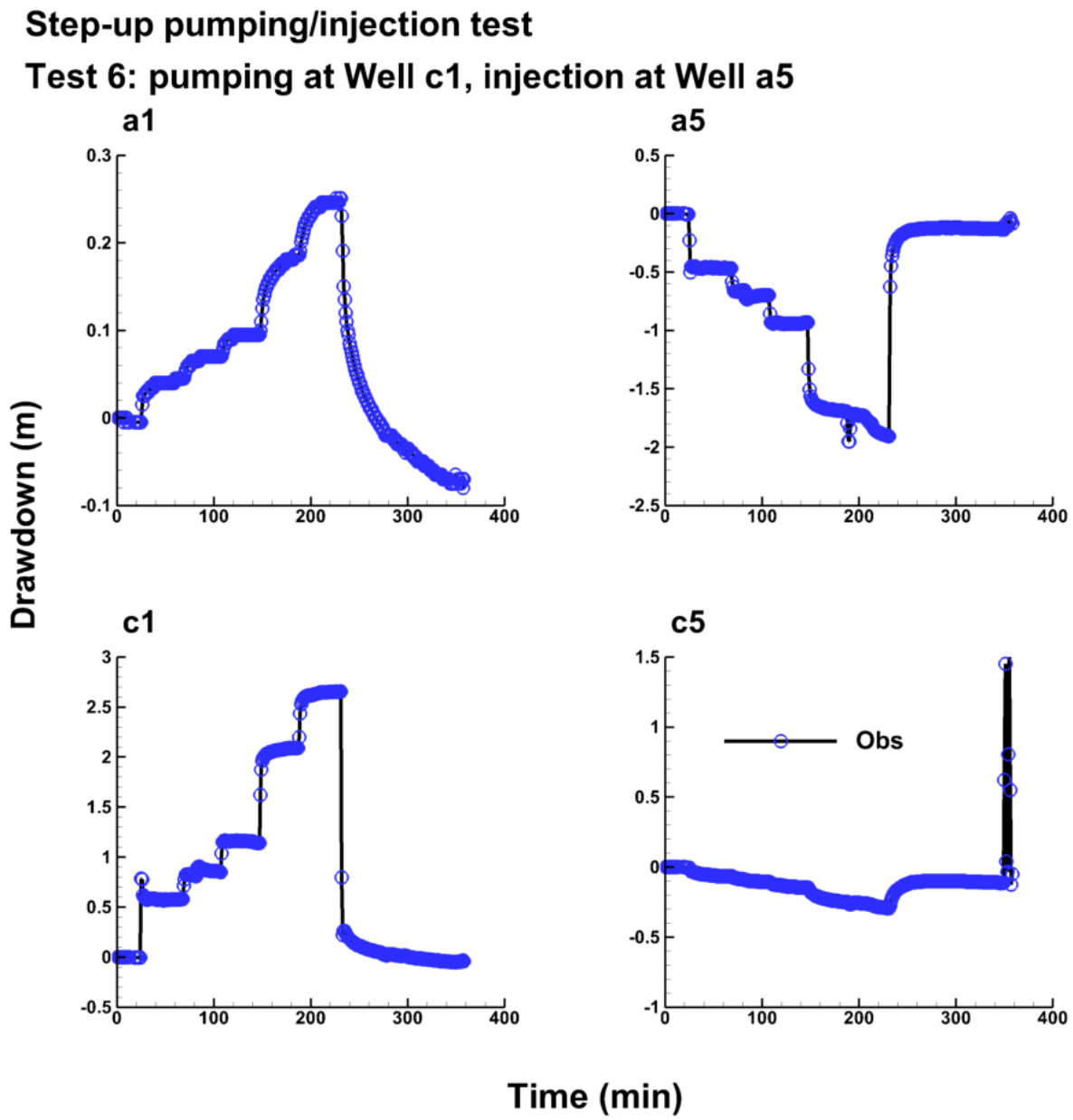

Fig. B5: *Observed drawdown curves of step-up pumping/injection Test 6 (pumping at Well c1, injection at Well a5). Open circles indicate the drawdown data used for inverse modeling.*

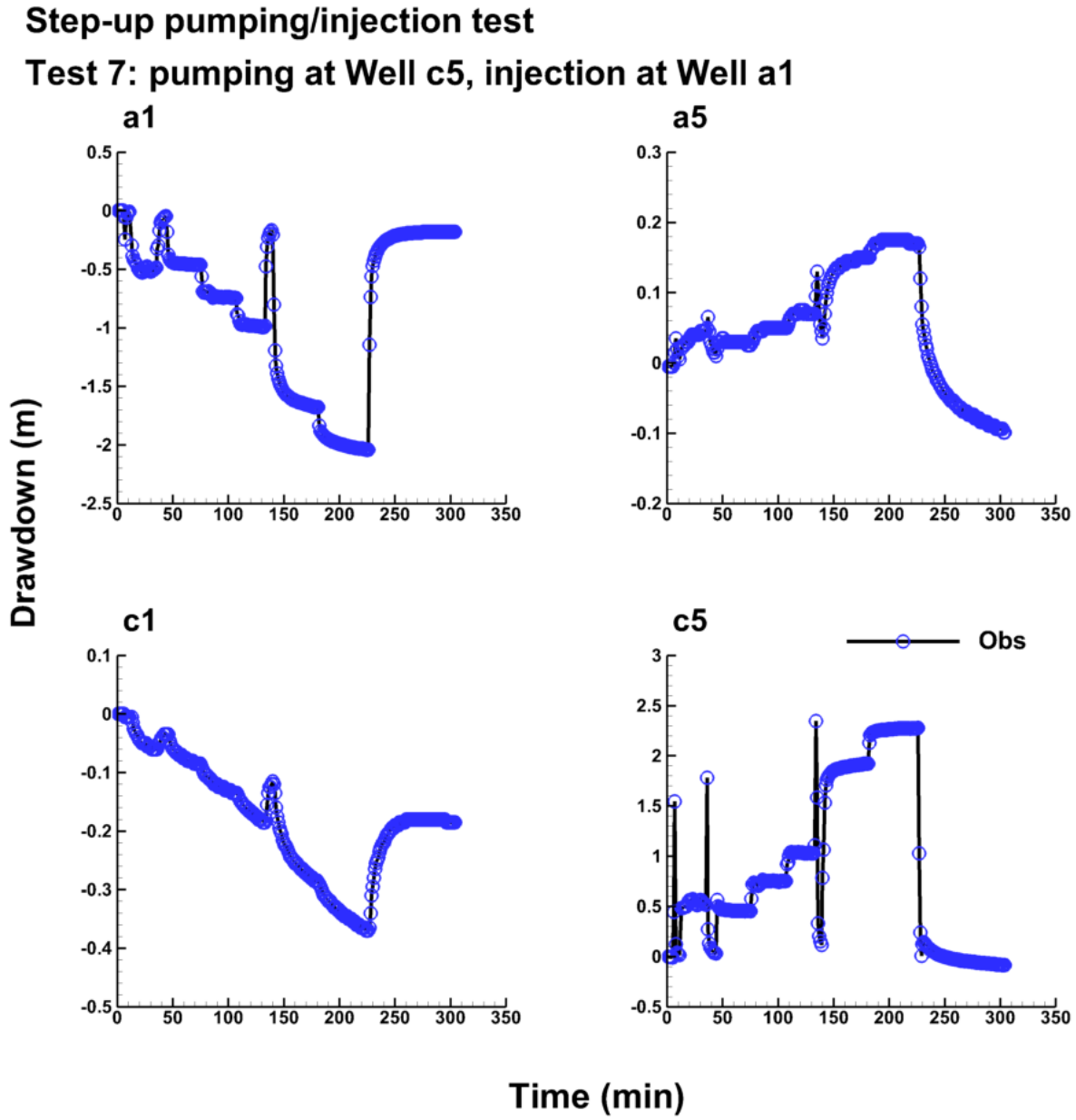

Fig. B6: *Observed drawdown curves of step-up pumping/injection Test 7 (pumping at Well c5, injection at Well a1). Open circles indicate the drawdown data used for inverse modeling.*

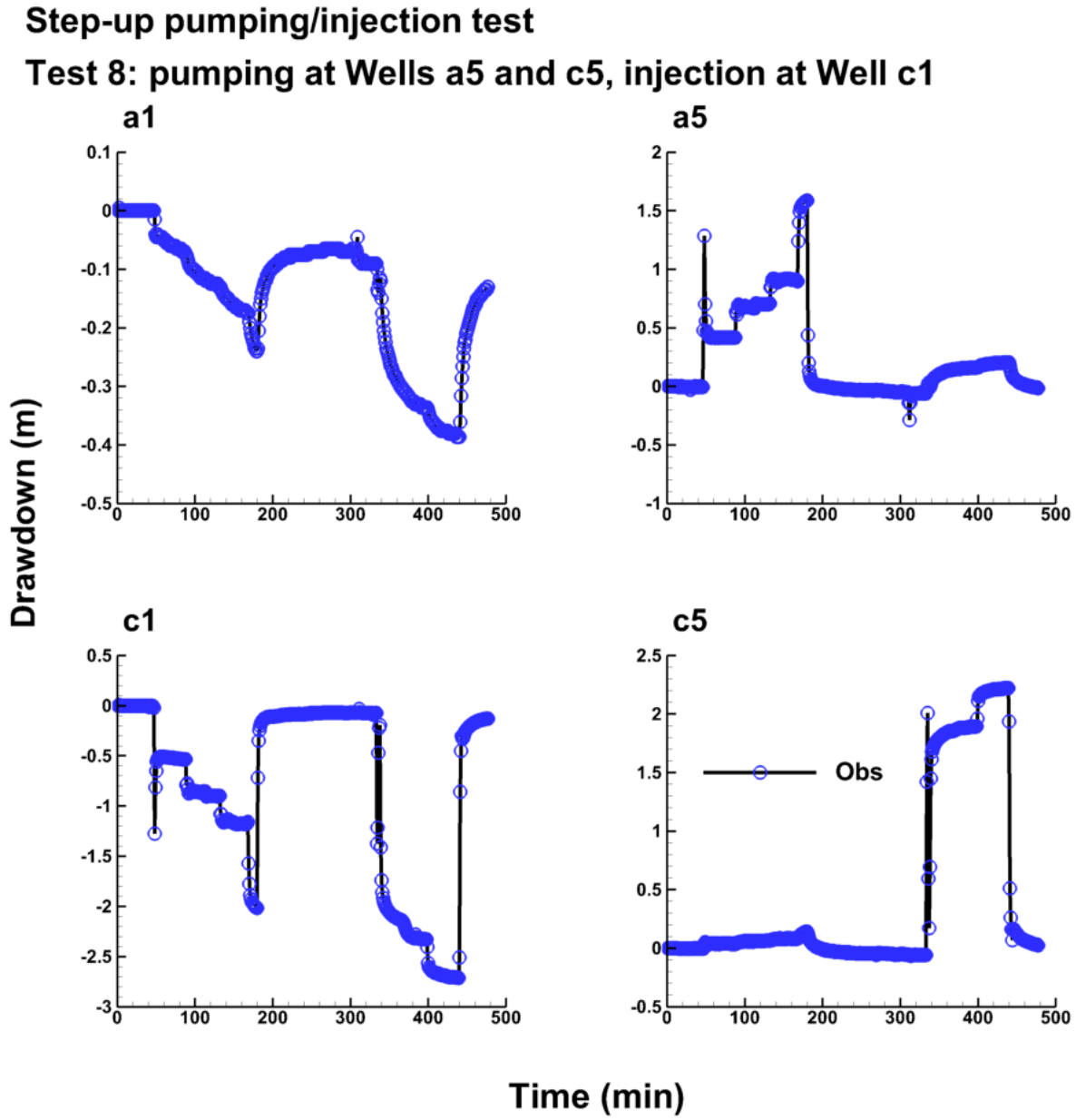

Fig. B7: *Observed drawdown curves of step-up pumping/injection Test 8 (pumping at Well a5 and c5, injection at Well c1). Open circles indicate the drawdown data used for inverse modeling.*

Appendix C

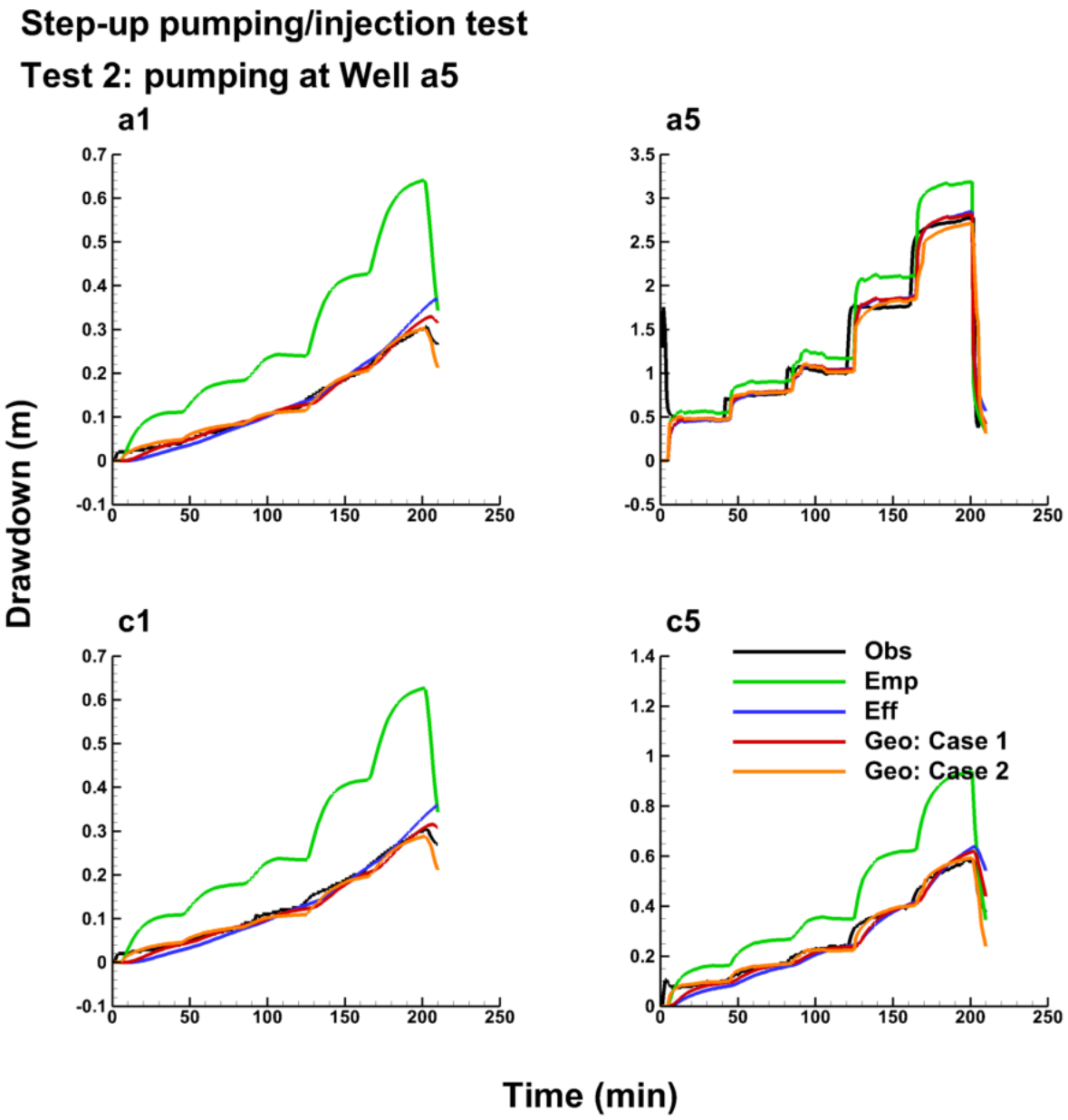

Fig. C1: *Drawdown vs. time plots for step-up pumping/injection test 2: pumping at Well a5 The black line represents the observed drawdowns, the green line represents the simulated drawdowns of the Creager's empirical solution, the blue line represents the simulated drawdowns of effective parameter approach, the red and orange line represent the simulated drawdowns of geological zonation approach: Case 1 and Case 2, respectively.*

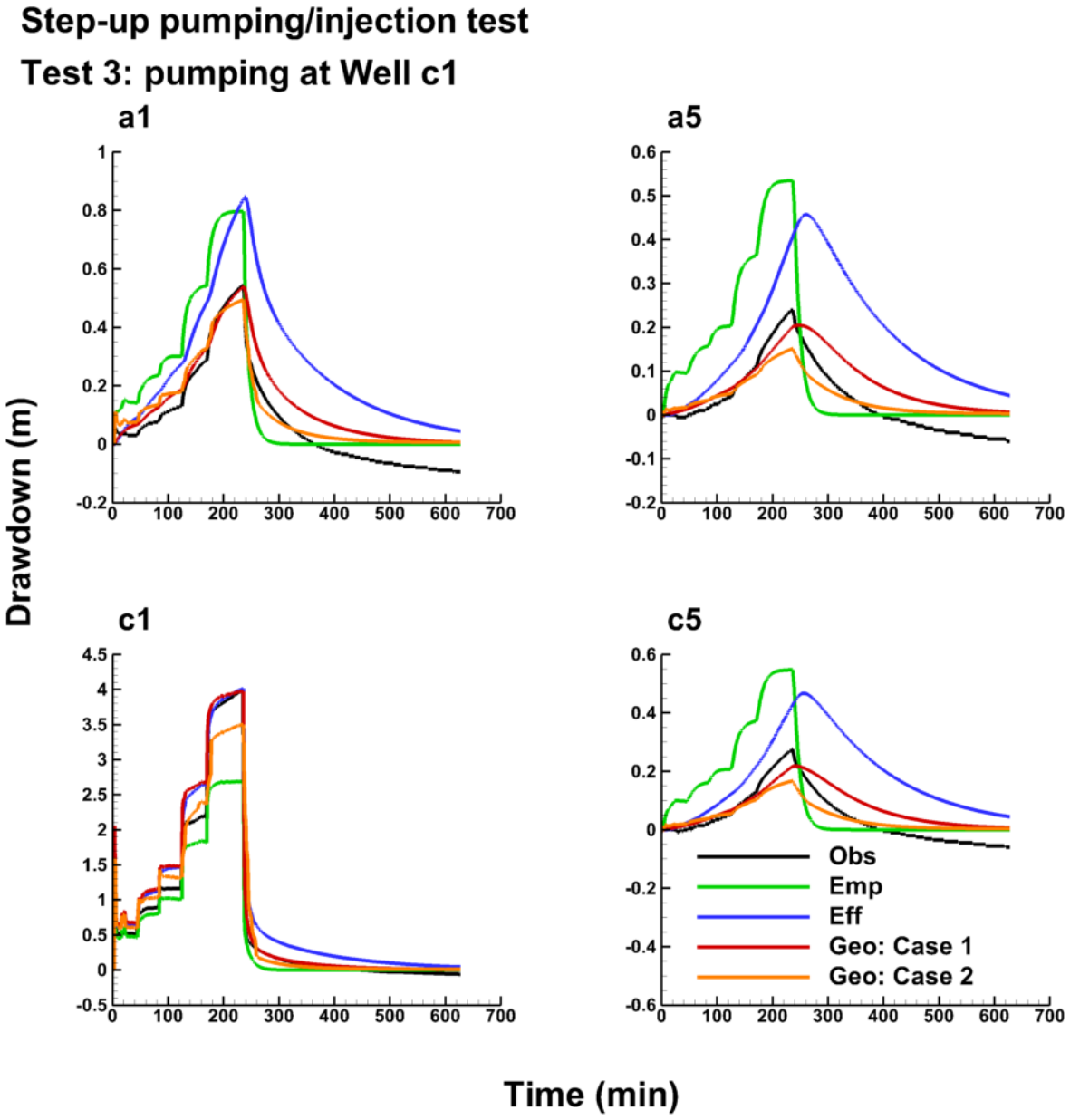

Fig. C2: *Drawdown vs. time plots for step-up pumping/injection test 3: pumping at Well c1. The black line represents the observed drawdowns, the green line represents the simulated drawdowns of the Creager's empirical solution, the blue line represents the simulated drawdowns of effective parameter approach, the red and orange line represent the simulated drawdowns of geological zonation approach: Case 1 and Case 2, respectively.*

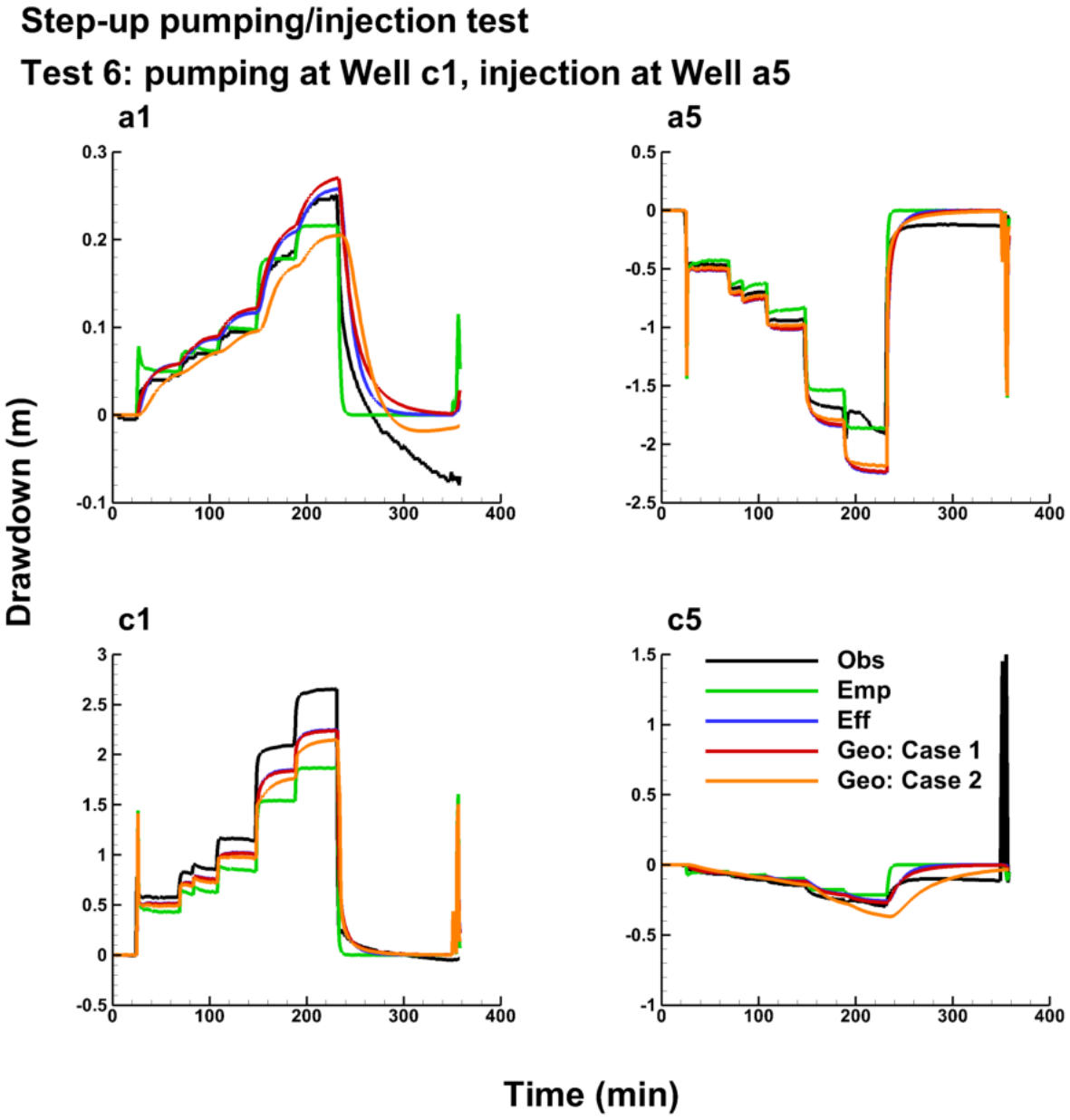

Fig. C3: *Drawdown vs. time plots for step-up pumping/injection test 6: pumping at Well c1, injection at Well a5. The black line represents the observed drawdowns, the green line represents the simulated drawdowns of the Creager's empirical solution, the blue line represents the simulated drawdowns of effective parameter approach, the red and orange line represent the simulated drawdowns of geological zonation approach: Case 1 and Case 2, respectively.*

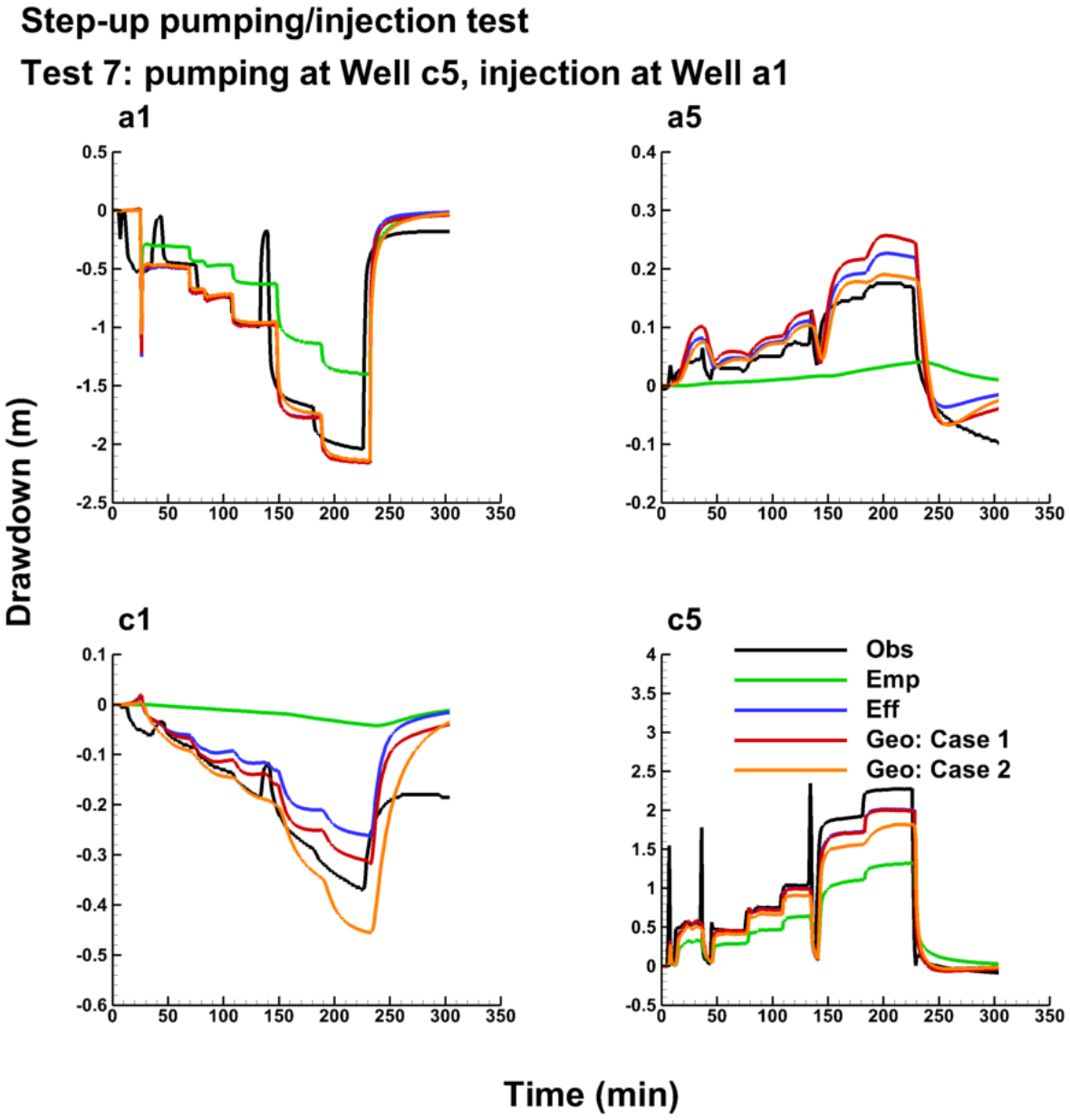

Fig. C4: *Drawdown vs. time plots for step-up pumping/injection test 7: pumping at Well c5, injection at Well a1. The black line represents the observed drawdowns, the green line represents the simulated drawdowns of the Creager's empirical solution, the blue line represents the simulated drawdowns of effective parameter approach, the red and orange line represent the simulated drawdowns of geological zonation approach: Case 1 and Case 2, respectively.*

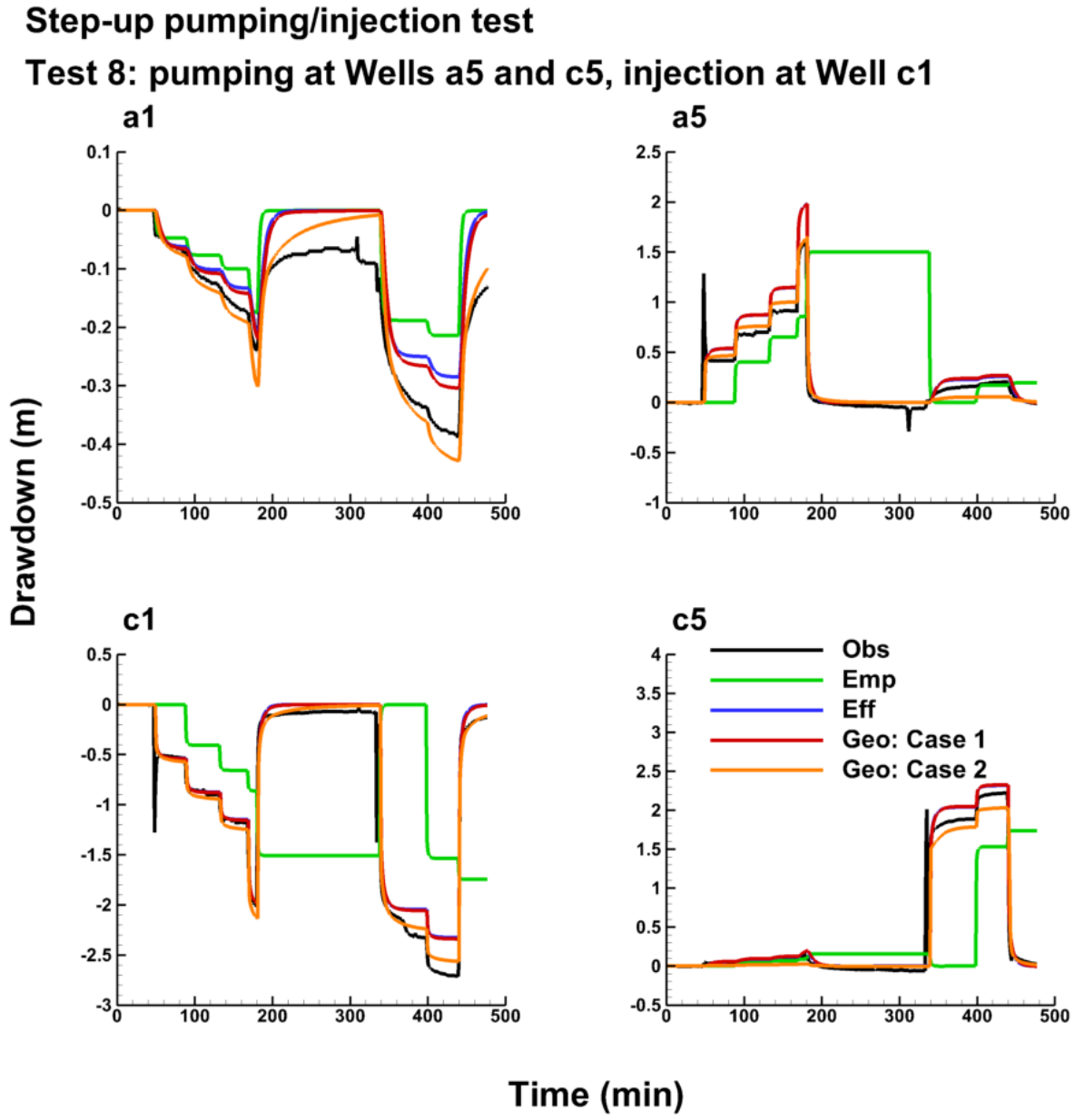

Fig. C5: *Drawdown vs. time plots for step-up pumping/injection test 8: pumping at Well a5 and c5, injection at Well c1. The black line represents the observed drawdowns, the green line represents the simulated drawdowns of the Creager's empirical solution, the blue line represents the simulated drawdowns of effective parameter approach, the red and orange line represent the simulated drawdowns of geological zonation approach: Case 1 and Case 2, respectively.*

Appendix D

Table D1: *Summary of estimated K and S_s values and their 95% confidence intervals of Test 1 to Test 8 for the geological zonation approach: Case 1 (isotropic).*

Case 1	Parameter			95% Percent Confidence Interval	
				Lower Limit	Upper Limit
Test 1	*K* (m/min)	Dc	1.84E-03	1.63E-11	2.08E+05
		Ds1	7.43E-04	8.83E-05	6.24E-03
		Ds2	5.33E-03	2.96E-03	9.60E-03
		Ds3	1.32E-03	5.65E-04	3.07E-03
	S_s (1/m)	DC	9.63E-05	5.93E-83	1.56E+74
		Ds1	1.57E-03	5.25E-04	4.67E-03
		Ds2	2.79E-04	2.95E-05	2.64E-03
		Ds3	8.12E-05	2.83E-10	2.33E+01
Test 2	*K* (m/min)	Dc	1.73E-03	5.64E-19	5.34E+12
		Ds1	7.20E-04	1.85E-06	2.81E-01
		Ds2	4.17E-03	7.31E-04	2.38E-02
		Ds3	2.40E-03	7.71E-04	7.48E-03
	S_s (1/m)	DC	1.02E-04	6.44E-35	1.60E+26
		Ds1	2.30E-04	1.10E-07	4.81E-01
		Ds2	1.10E-03	2.01E-04	6.01E-03
		Ds3	5.45E-05	1.73E-31	1.71E+22
Test 3	*K* (m/min)	Dc	7.58E-07	7.58E-307	7.58E+293
		Ds1	1.08E-03	6.54E-04	1.79E-03
		Ds2	1.22E-03	7.26E-04	2.06E-03
		Ds3	6.83E-03	9.63E-04	4.85E-02
	S_s (1/m)	DC	1.55E-06	1.72E-167	1.40E+155
		Ds1	1.03E-03	8.07E-04	1.30E-03
		Ds2	1.07E-04	7.85E-06	1.47E-03
		Ds3	2.70E-06	1.73E-263	4.20E+251
Test 4	*K* (m/min)	Dc	4.40E-03	1.87E-09	1.03E+04
		Ds1	2.98E-03	7.82E-04	1.13E-02
		Ds2	2.10E-04	2.00E-12	2.21E+04
		Ds3	1.80E-03	1.16E-47	2.77E+41
	S_s (1/m)	DC	1.11E-04	1.09E-85	1.13E+77
		Ds1	2.94E-03	8.13E-04	1.07E-02
		Ds2	2.12E-04	4.92E-16	9.10E+07
		Ds3	1.18E-04	7.89E-157	1.78E+148

Table D1: *(continued)*

Case 1	Parameter			95% Percent Confidence Interval	
				Lower Limit	Upper Limit
Test 5	K (m/min)	Dc	2.34E-03	1.61E-07	3.39E+01
		Ds1	1.22E-03	4.57E-05	3.24E-02
		Ds2	4.99E-03	1.92E-03	1.30E-02
		Ds3	1.85E-03	6.37E-04	5.38E-03
	S_s (1/m)	DC	7.15E-03	5.52E-08	9.27E+02
		Ds1	1.40E-04	1.32E-06	1.50E-02
		Ds2	3.58E-04	1.57E-05	8.15E-03
		Ds3	1.37E-02	7.08E-03	2.67E-02
Test 6	K (m/min)	Dc	1.36E-03	3.09E-11	5.95E+04
		Ds1	1.04E-03	2.33E-06	4.64E-01
		Ds2	2.95E-03	1.05E-04	8.34E-02
		Ds3	1.56E-03	3.47E-04	7.05E-03
	S_s (1/m)	DC	1.08E-04	8.13E-118	1.45E+109
		Ds1	1.17E-03	2.81E-04	4.90E-03
		Ds2	3.24E-04	1.31E-07	8.01E-01
		Ds3	9.62E-05	1.84E-22	5.04E+13
Test 7	K (m/min)	Dc	2.01E-03	1.87E-13	2.16E+07
		Ds1	2.54E-03	2.75E-04	2.34E-02
		Ds2	1.56E-03	1.42E-05	1.72E-01
		Ds3	7.97E-04	5.93E-05	1.07E-02
	S_s (1/m)	DC	1.10E-04	3.75E-37	3.25E+28
		Ds1	5.29E-04	1.10E-05	2.55E-02
		Ds2	1.80E-04	2.36E-09	1.37E+01
		Ds3	1.30E-04	2.27E-17	7.48E+08
Test 8	K (m/min)	Dc	2.45E-04	1.36E-53	4.42E+45
		Ds1	2.67E-03	9.90E-04	7.22E-03
		Ds2	5.39E-04	4.03E-06	7.23E-02
		Ds3	2.71E-01	2.52E-118	2.93E+116
	S_s (1/m)	DC	1.77E-04	8.11E-20	3.88E+11
		Ds1	5.70E-04	1.42E-04	2.28E-03
		Ds2	3.22E-05	2.45E-24	4.24E+14
		Ds3	1.19E-04	1.19E-304	1.19E+296

Table D2: *Summary of estimated K and S_s values and their 95% confidence intervals of Test 1 to Test 8 for the geological zonation approach: Case 2 (anisotropic).*

Case 2	Parameter			95% Percent Confidence Interval	
				Lower Limit	Upper Limit
Test 1	K_x/ K_y (m/min)	Dc	2.32E-07	1.41E-08	3.83E-06
		Ds1	1.67E-03	1.48E-03	1.90E-03
		Ds2	3.55E-02	2.76E-02	4.56E-02
		Ds3	4.84E-04	6.71E-05	3.50E-03
	K_z (m/min)	Dc	2.75E-08	4.04E-09	1.87E-07
		Ds1	8.56E-04	7.03E-04	1.04E-03
		Ds2	4.82E-04	3.29E-04	7.06E-04
		Ds3	3.53E-03	1.54E-03	8.10E-03
	S_s (1/m)	DC	1.49E-03	3.41E-05	6.53E-02
		Ds1	7.43E-05	2.59E-05	2.13E-04
		Ds2	5.40E-03	3.34E-03	8.72E-03
		Ds3	8.03E-03	4.72E-03	1.37E-02
Test 2	K_x/ K_y (m/min)	Dc	3.71E-06	1.81E-07	7.58E-05
		Ds1	1.54E-03	9.52E-04	2.49E-03
		Ds2	2.94E-02	1.58E-02	5.46E-02
		Ds3	3.12E-05	1.65E-31	5.88E+21
	K_z (m/min)	Dc	2.58E-07	4.30E-24	1.55E+10
		Ds1	5.91E-04	2.83E-04	1.23E-03
		Ds2	8.63E-04	4.14E-04	1.80E-03
		Ds3	2.56E-03	4.94E-09	1.33E+03
	S_s (1/m)	DC	1.39E-04	4.07E-31	4.77E+22
		Ds1	2.03E-05	7.74E-07	5.34E-04
		Ds2	1.40E-03	5.40E-04	3.61E-03
		Ds3	4.34E-05	2.07E-24	9.14E+14
Test 3	K_x/ K_y (m/min)	Dc	1.93E-06	1.93E-306	1.93E+294
		Ds1	1.29E-03	1.26E-03	1.32E-03
		Ds2	1.56E-03	1.06E-03	2.29E-03
		Ds3	6.67E-05	9.86E-06	4.52E-04
	K_z (m/min)	Dc	2.66E-07	2.66E-307	2.66E+293
		Ds1	6.55E-04	6.24E-04	6.88E-04
		Ds2	5.02E-03	3.10E-03	8.15E-03
		Ds3	1.00E-03	1.40E-04	7.17E-03
	S_s (1/m)	DC	1.15E-02	2.96E-04	4.47E-01
		Ds1	3.57E-05	1.37E-05	9.26E-05
		Ds2	1.52E-04	1.87E-05	1.24E-03
		Ds3	1.56E-03	1.02E-03	2.40E-03

Table D2: *(continued)*

Case 2	Parameter			95% Percent Confidence Interval	
				Lower Limit	Upper Limit
Test 4	K_x/ K_y (m/min)	Dc	1.58E-06	1.58E-306	1.58E+294
		Ds1	1.62E-03	1.49E-03	1.76E-03
		Ds2	2.62E-03	1.56E-04	4.39E-02
		Ds3	2.37E-03	3.93E-07	1.43E+01
	K_z (m/min)	Dc	1.00E-06	1.00E-306	1.00E+294
		Ds1	1.62E-03	1.37E-03	1.92E-03
		Ds2	2.62E-03	1.27E-04	5.42E-02
		Ds3	2.37E-03	2.00E-18	2.79E+12
	S_s (1/m)	DC	3.10E-03	3.47E-43	2.78E+37
		Ds1	3.24E-04	1.69E-06	6.22E-02
		Ds2	3.24E-04	3.61E-18	2.91E+10
		Ds3	3.24E-04	4.30E-34	2.44E+26
Test 5	K_x/ K_y (m/min)	Dc	3.31E-07	1.57E-07	7.01E-07
		Ds1	9.30E-04	8.42E-04	1.03E-03
		Ds2	5.41E-03	3.62E-03	8.10E-03
		Ds3	1.00E-01	8.63E-02	1.16E-01
	K_z (m/min)	Dc	1.00E-06	7.93E-07	1.26E-06
		Ds1	2.40E-04	2.16E-04	2.65E-04
		Ds2	8.16E-02	5.86E-02	1.14E-01
		Ds3	1.27E-02	5.29E-03	3.03E-02
	S_s (1/m)	DC	2.91E-02	1.65E-02	5.11E-02
		Ds1	5.68E-04	4.66E-04	6.93E-04
		Ds2	5.40E-04	2.15E-04	1.36E-03
		Ds3	1.00E-02	7.71E-03	1.30E-02
Test 6	K_x/ K_y (m/min)	Dc	7.02E-07	5.06E-08	9.73E-06
		Ds1	2.80E-04	1.49E-04	5.25E-04
		Ds2	1.27E-02	1.20E-02	1.36E-02
		Ds3	1.04E-02	7.21E-04	1.50E-01
	K_z (m/min)	Dc	5.76E-07	1.05E-07	3.17E-06
		Ds1	8.26E-05	4.23E-05	1.61E-04
		Ds2	2.35E-02	1.71E-02	3.22E-02
		Ds3	4.81E-03	6.46E-04	3.57E-02
	S_s (1/m)	DC	1.76E-02	3.88E-04	7.97E-01
		Ds1	1.60E-04	8.35E-05	3.05E-04
		Ds2	9.00E-03	1.97E-03	4.11E-02
		Ds3	2.78E-04	1.77E-07	4.36E-01

Table D2: *(continued)*

Case 2	Parameter			95% Percent Confidence Interval	
				Lower Limit	Upper Limit
Test 7	K_x/ K_y (m/min)	Dc	1.53E-07	1.53E-307	1.53E+293
		Ds1	1.81E-03	1.57E-03	2.10E-03
		Ds2	1.81E-03	4.98E-05	6.56E-02
		Ds3	5.57E-03	2.93E-03	1.06E-02
	K_z (m/min)	Dc	9.99E-07	2.20E-07	4.54E-06
		Ds1	3.95E-04	2.92E-04	5.34E-04
		Ds2	2.25E-02	8.76E-04	5.78E-01
		Ds3	9.92E-03	3.65E-07	2.70E+02
	S_s (1/m)	DC	3.10E-03	3.52E-05	2.73E-01
		Ds1	3.25E-04	2.31E-04	4.57E-04
		Ds2	6.42E-04	1.84E-05	2.25E-02
		Ds3	5.90E-04	9.55E-08	3.64E+00
Test 8	K_x/ K_y (m/min)	Dc	4.33E-07	7.03E-11	2.67E-03
		Ds1	1.41E-03	1.38E-03	1.44E-03
		Ds2	1.28E-03	5.59E-04	2.94E-03
		Ds3	2.12E-03	1.27E-03	3.54E-03
	K_z (m/min)	Dc	2.35E-07	4.97E-12	1.11E-02
		Ds1	1.51E-03	1.44E-03	1.59E-03
		Ds2	9.52E-03	4.33E-03	2.09E-02
		Ds3	9.49E-03	9.80E-07	9.19E+01
	S_s (1/m)	DC	1.07E-02	3.88E-07	2.93E+02
		Ds1	7.74E-05	5.67E-05	1.06E-04
		Ds2	1.33E-03	6.43E-04	2.75E-03
		Ds3	1.99E-03	8.56E-04	4.63E-03

Appendix E

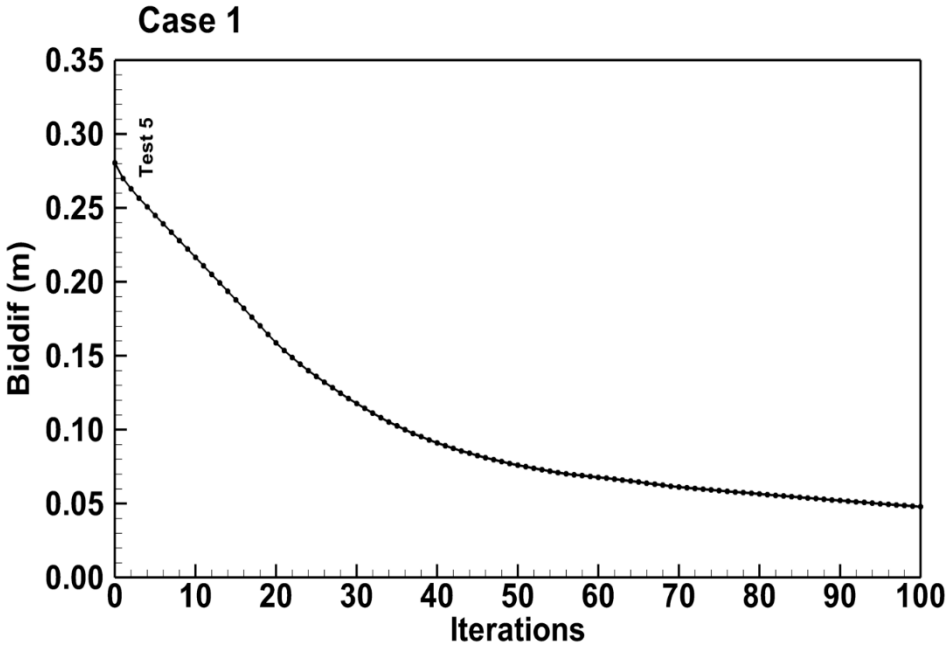

Fig. E1: *Biggest difference of head misfit during iterations from Case 1. Note: "Biddif" in the plot represents the biggest difference of head misfit during iterations.*

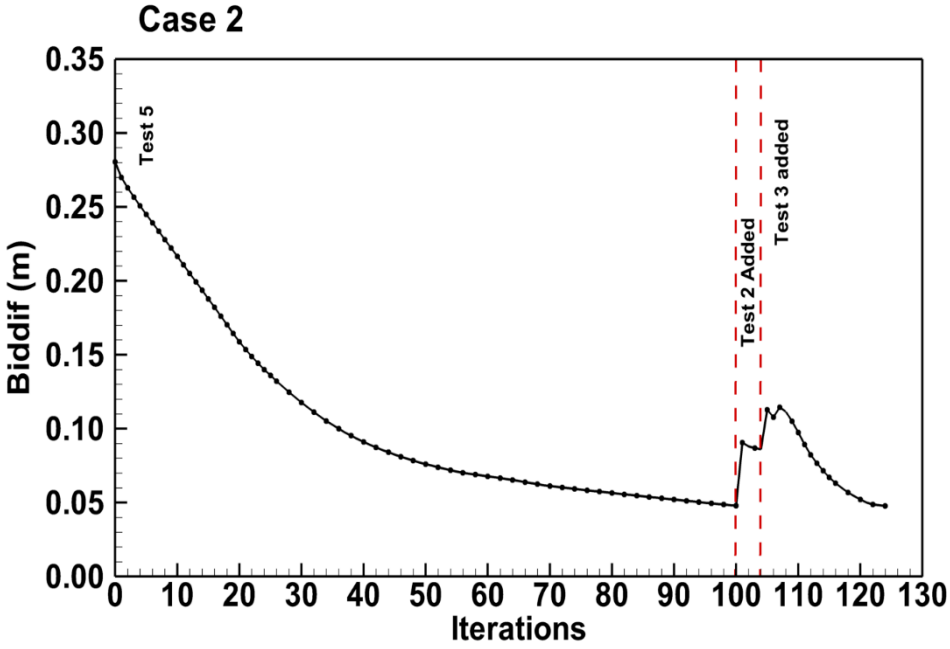

Fig. E2: *Biggest difference of head misfit during iterations from Case 2. Note: "Biddif" in the plot represents the biggest difference of head misfit during iterations.*

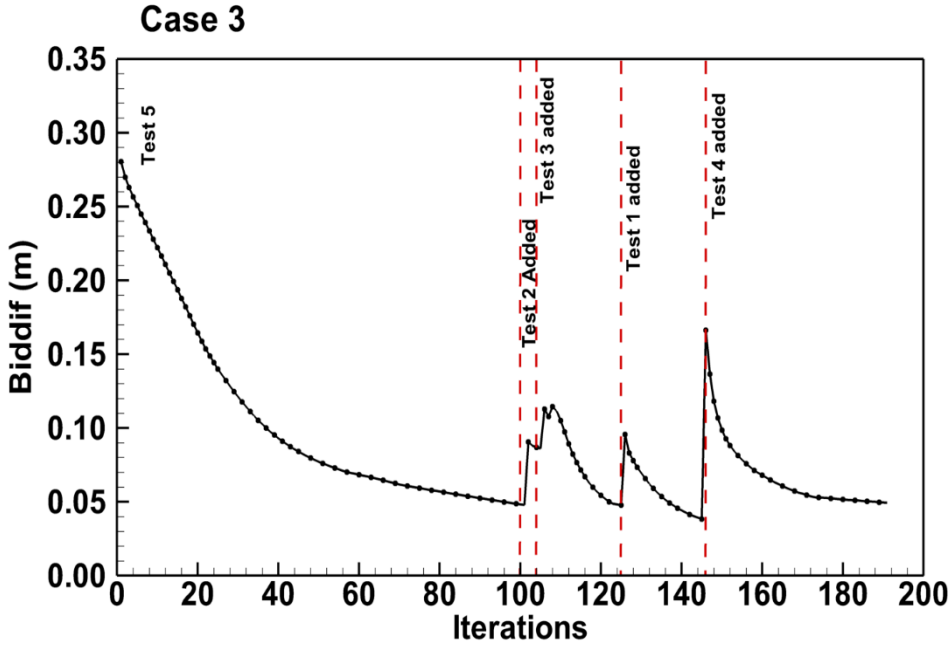

Fig. E3: *Biggest difference of head misfit during iterations from Case 3. Note: "Biddif" in the plot represents the biggest difference of head misfit during iterations.*

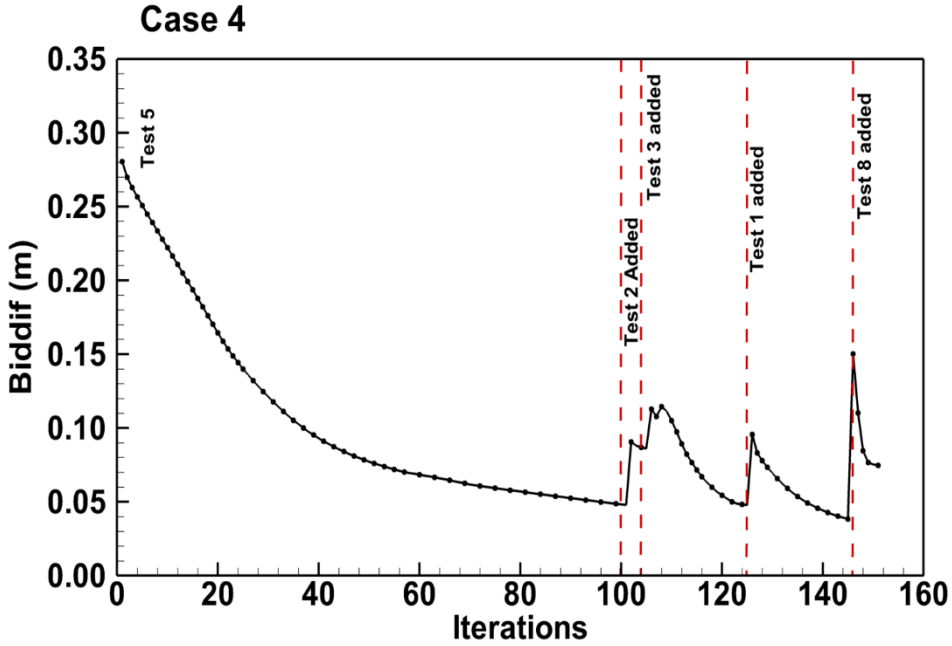

Fig. E4: *Biggest difference of head misfit during iterations from Case 4. Note: "Biddif" in the plot represents the biggest difference of head misfit during iterations.*

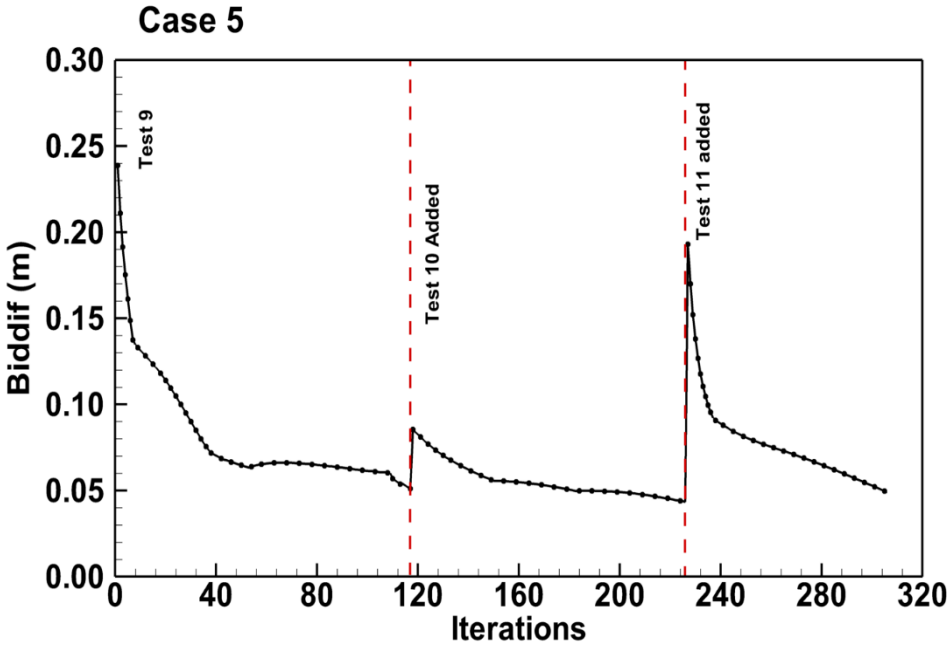

Fig. E5: *Biggest difference of head misfit during iterations from Case 5. Note: "Biddif" in the plot represents the biggest difference of head misfit during iterations.*

Appendix F

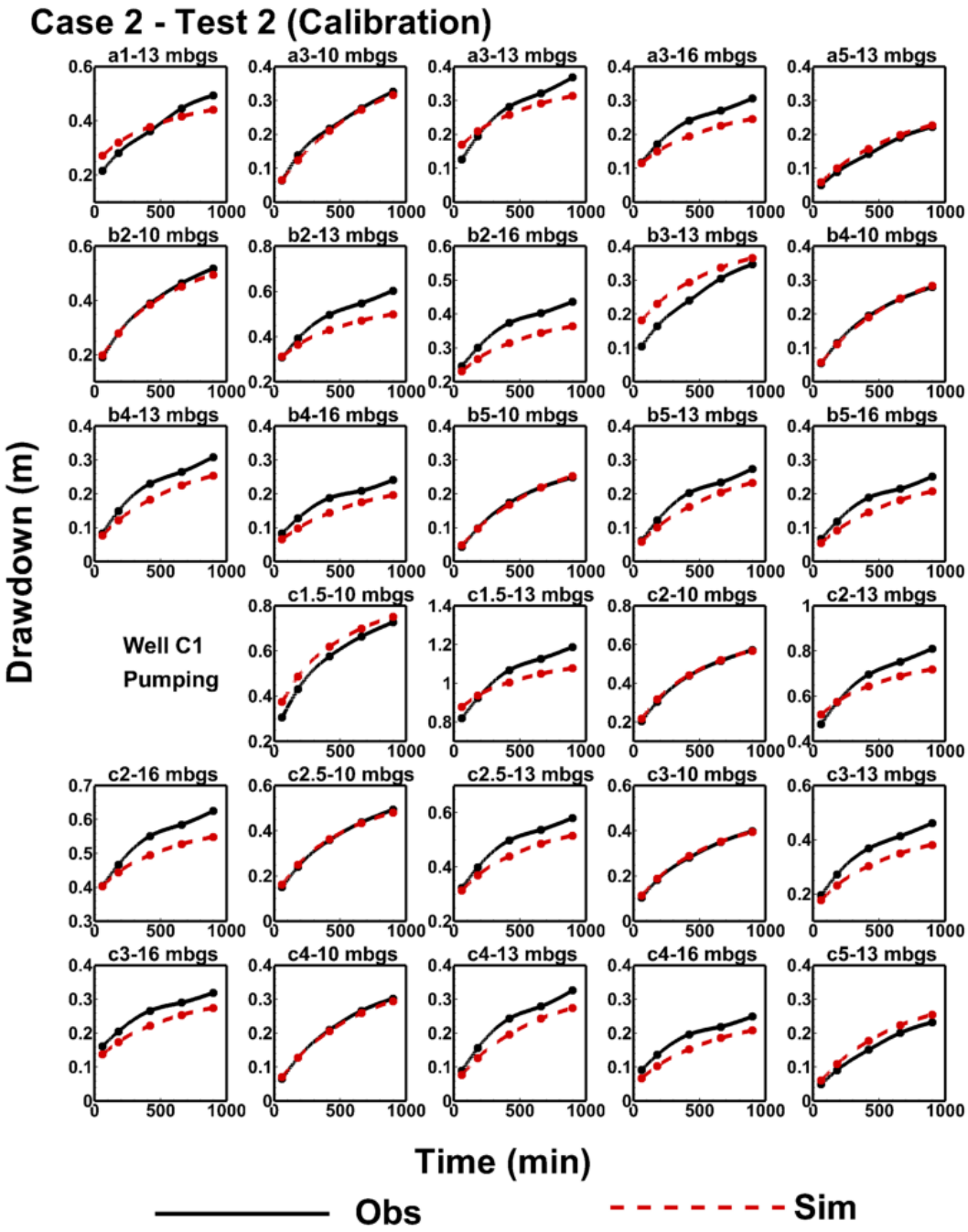

Fig. F1: *Observed and simulated drawdowns of Case 2 calibration results versus time curves at observation ports during HT Test 2 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 2. Pumping wells used in this HT test are listed in the plot as well.*

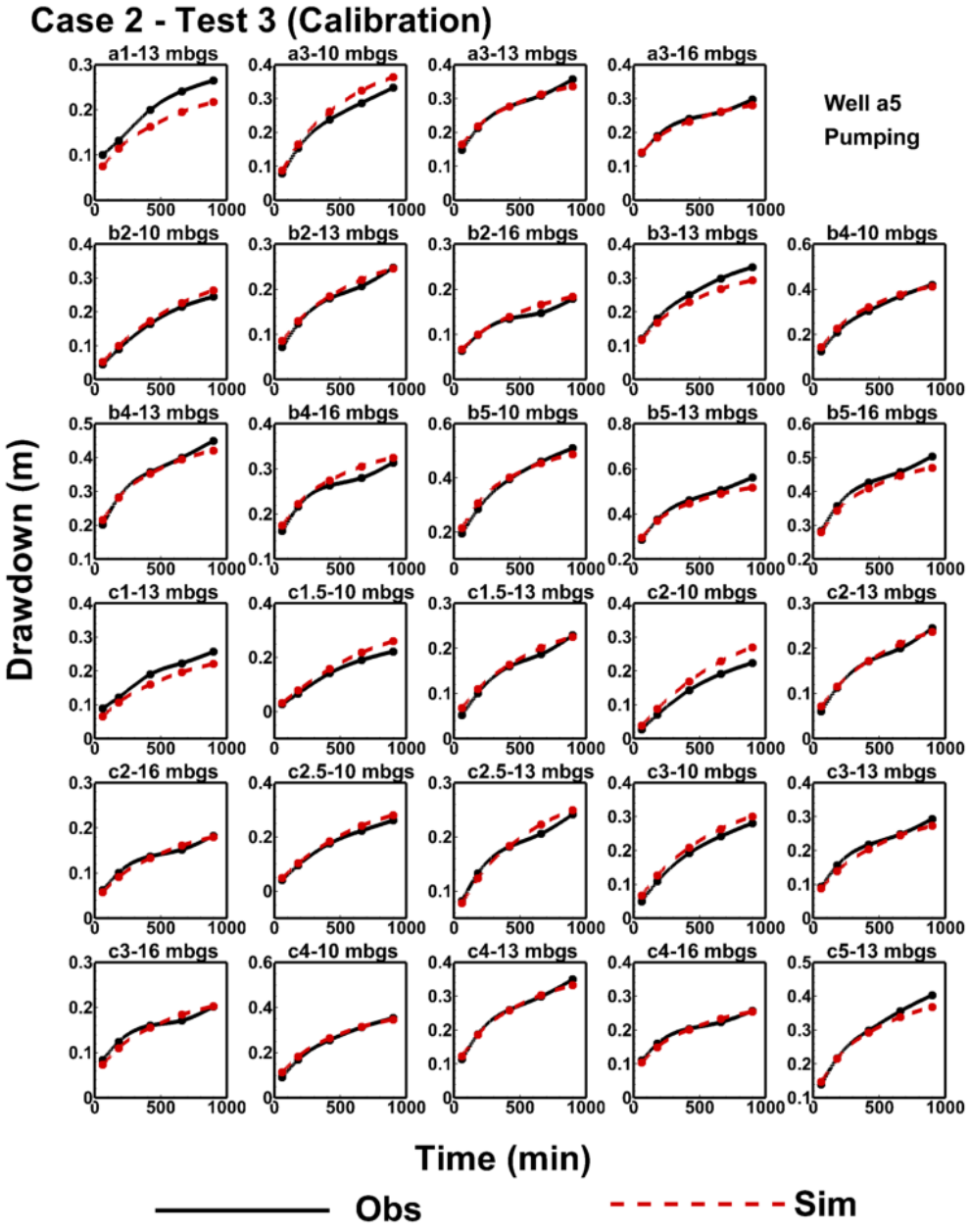

Fig. F2: *Observed and simulated drawdowns of Case 2 calibration results versus time curves at observation ports during HT Test 3 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 2. Pumping wells used in this HT test are listed in the plot as well.*

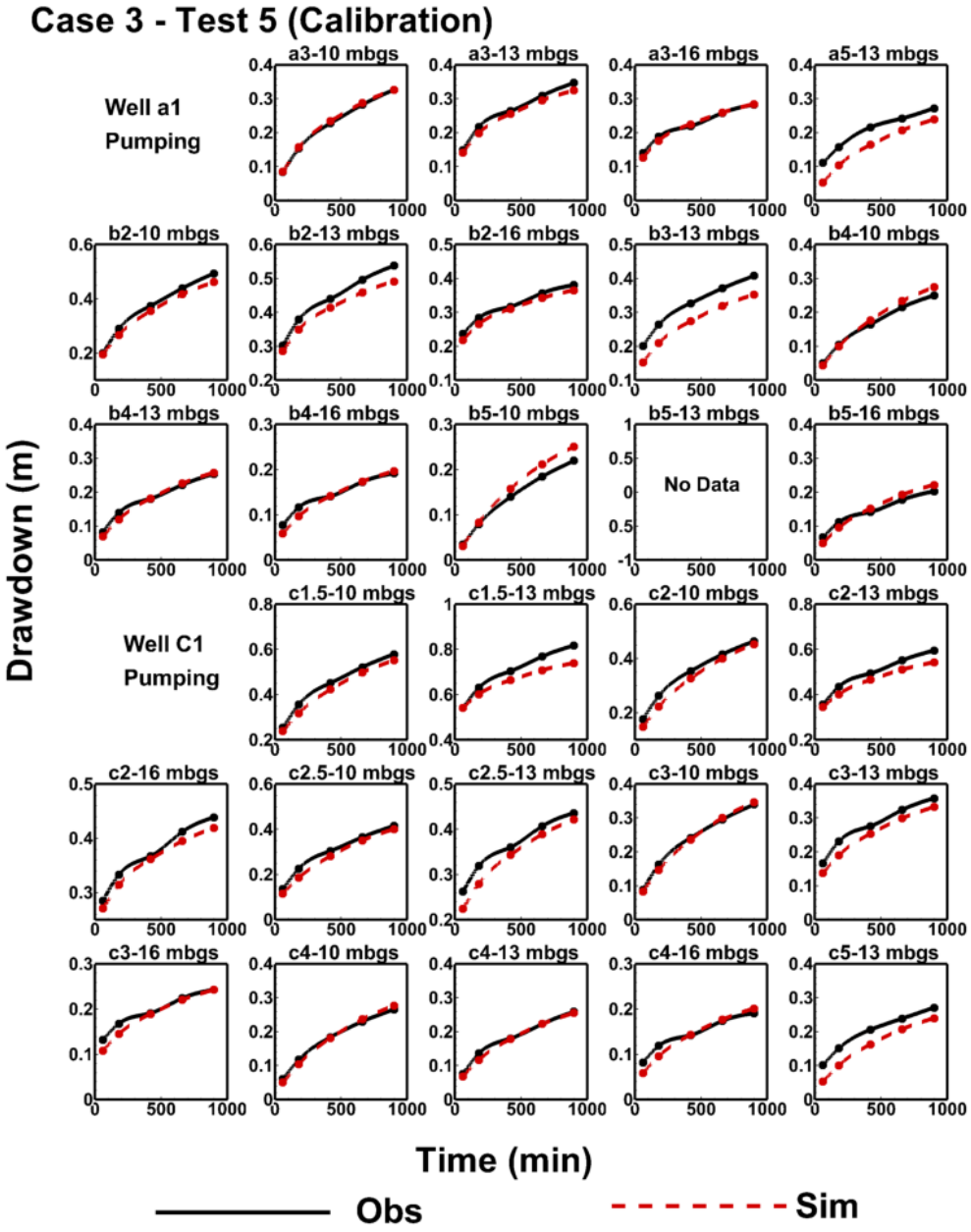

Fig. F3: *Observed and simulated drawdowns of Case 3 calibration results versus time curves at observation ports during HT Test 5 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 3. Note: "No data" indicates that data was not collected in that well during corresponding HT test. Pumping wells used in this HT test are listed in the plot as well.*

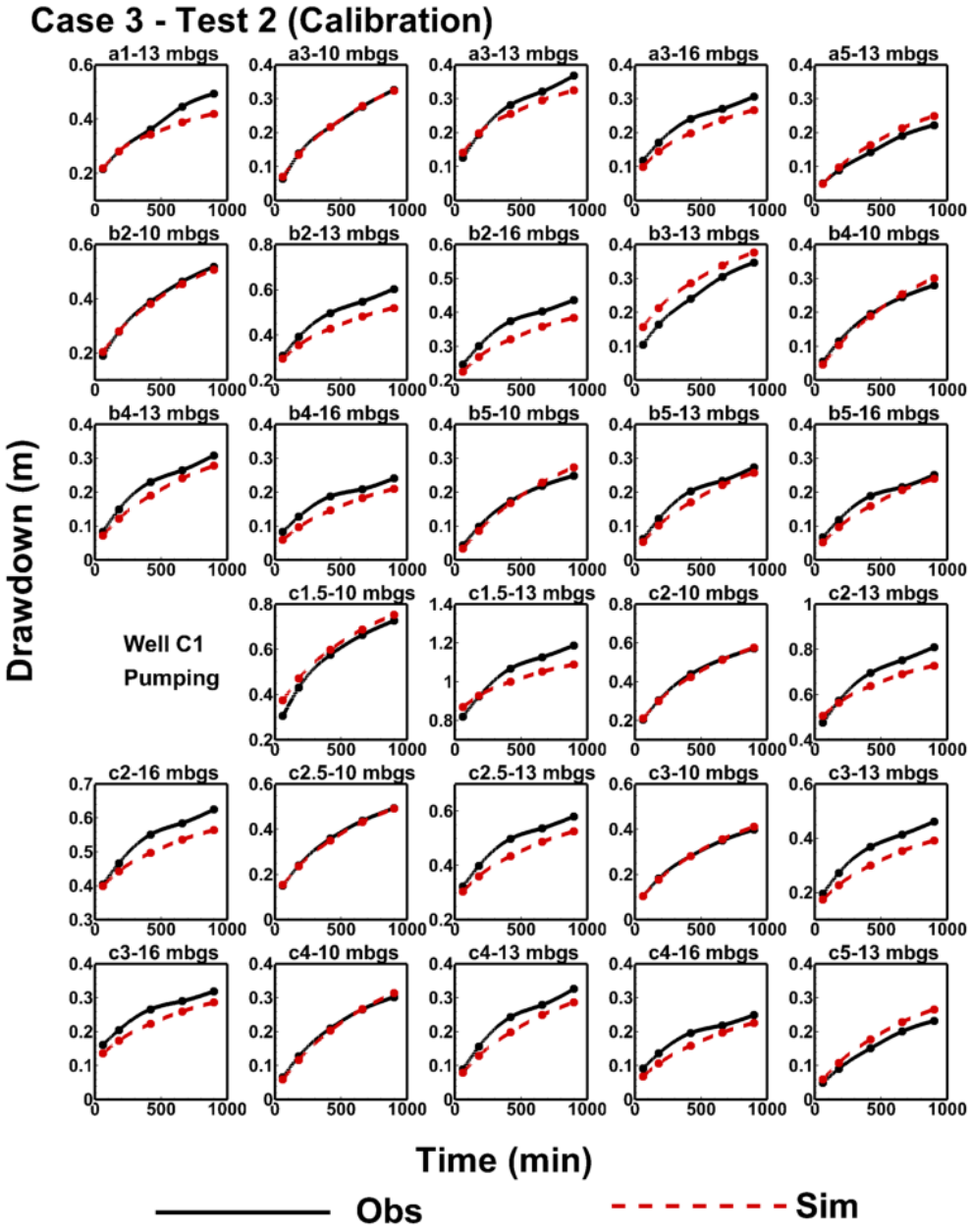

Fig. F4: *Observed and simulated drawdowns of Case 3 calibration results versus time curves at observation ports during HT Test 2 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 3. Pumping wells used in this HT test are listed in the plot as well.*

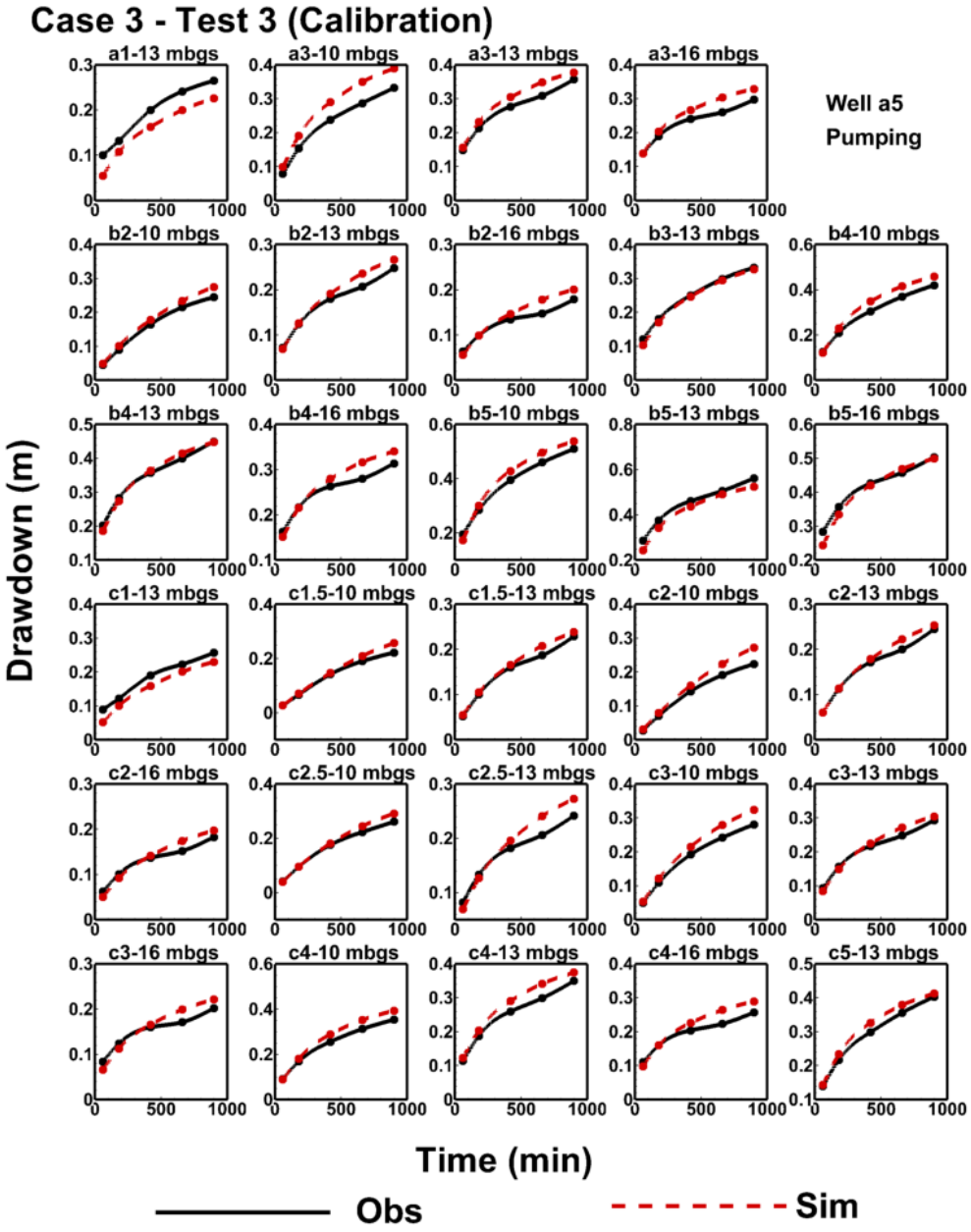

Fig. F5: *Observed and simulated drawdowns of Case 3 calibration results versus time curves at observation ports during HT Test 3 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 3. Pumping wells used in this HT test are listed in the plot as well.*

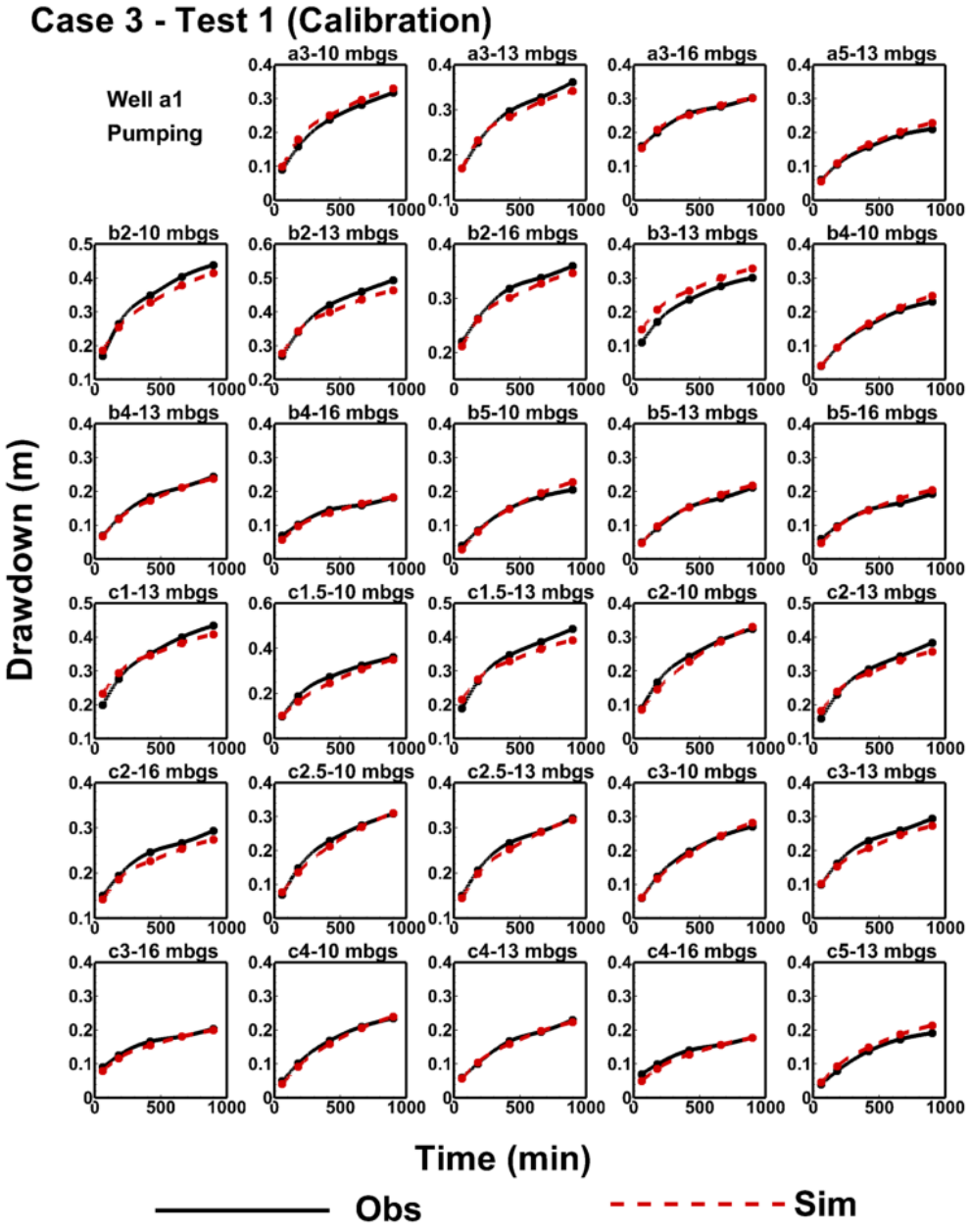

Fig. F6: *Observed and simulated drawdowns of Case 3 calibration results versus time curves at observation ports during HT Test 1 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 3. Pumping wells used in this HT test are listed in the plot as well.*

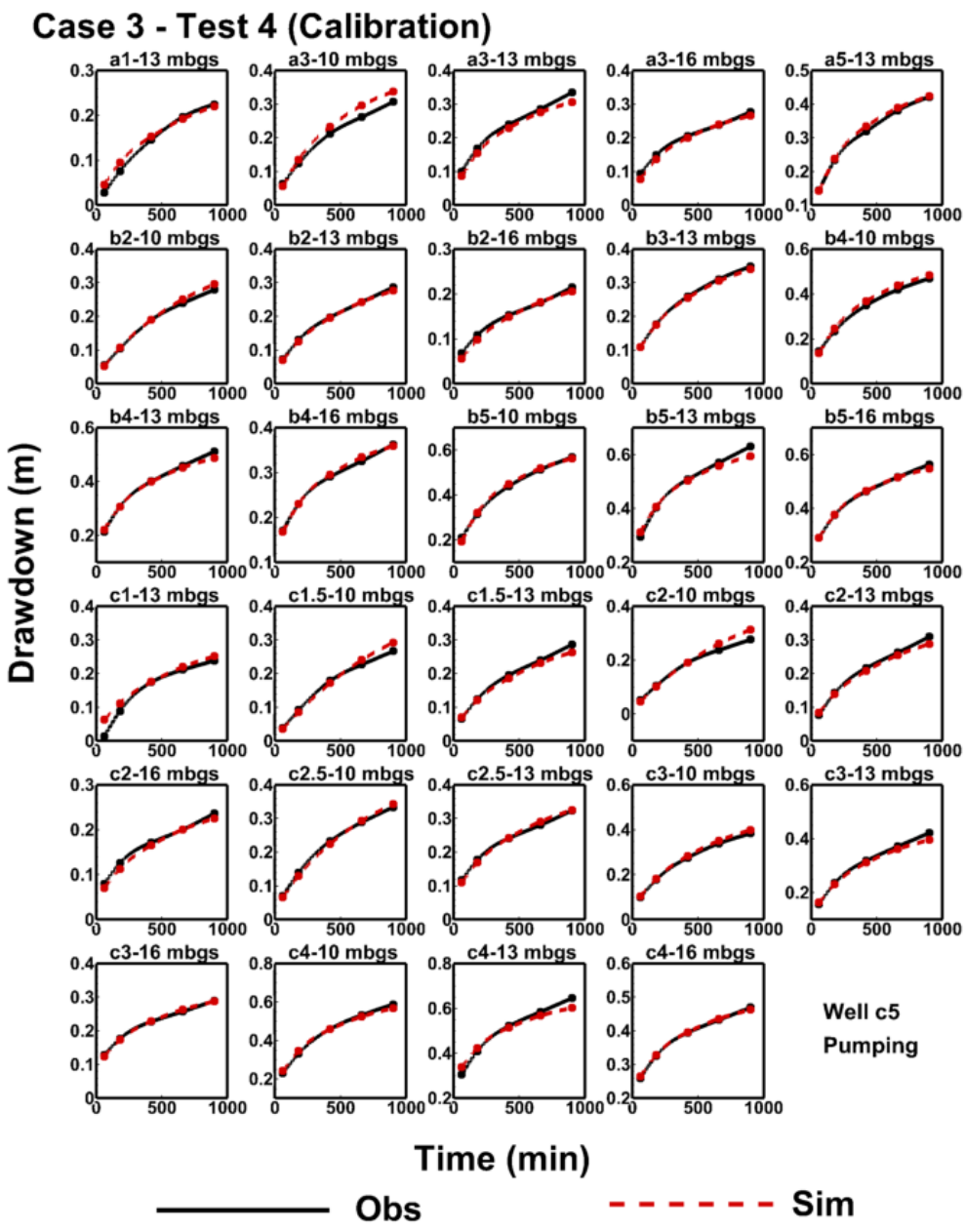

Fig. F7: *Observed and simulated drawdowns of Case 3 calibration results versus time curves at observation ports during HT Test 4 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 3. Pumping wells used in this HT test are listed in the plot as well.*

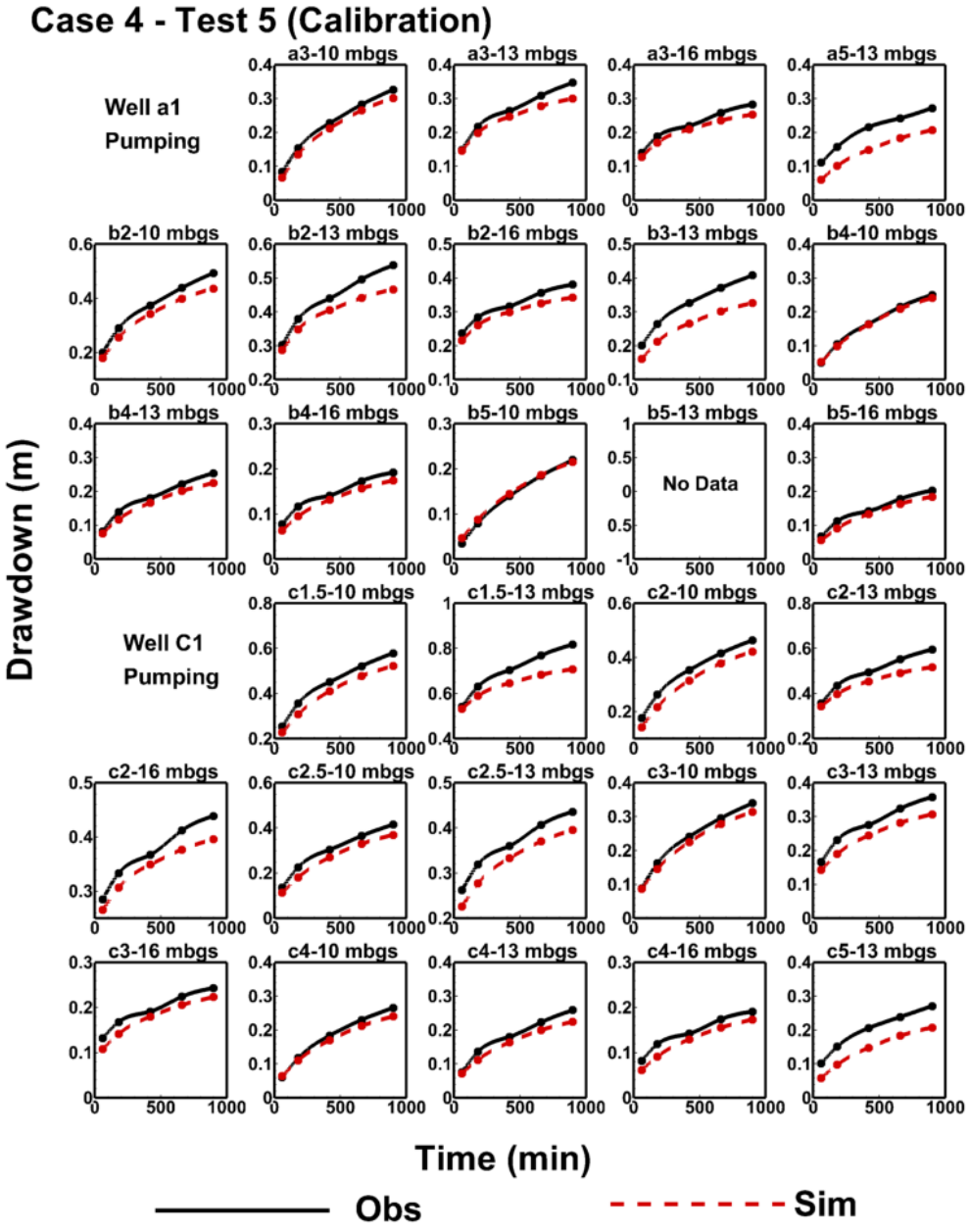

Fig. F8: *Observed and simulated drawdowns of Case 4 calibration results versus time curves at observation ports during HT Test 5 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 4. Note: "No data" indicates that data was not collected in that well during corresponding HT test. Pumping wells used in this HT test are listed in the plot as well.*

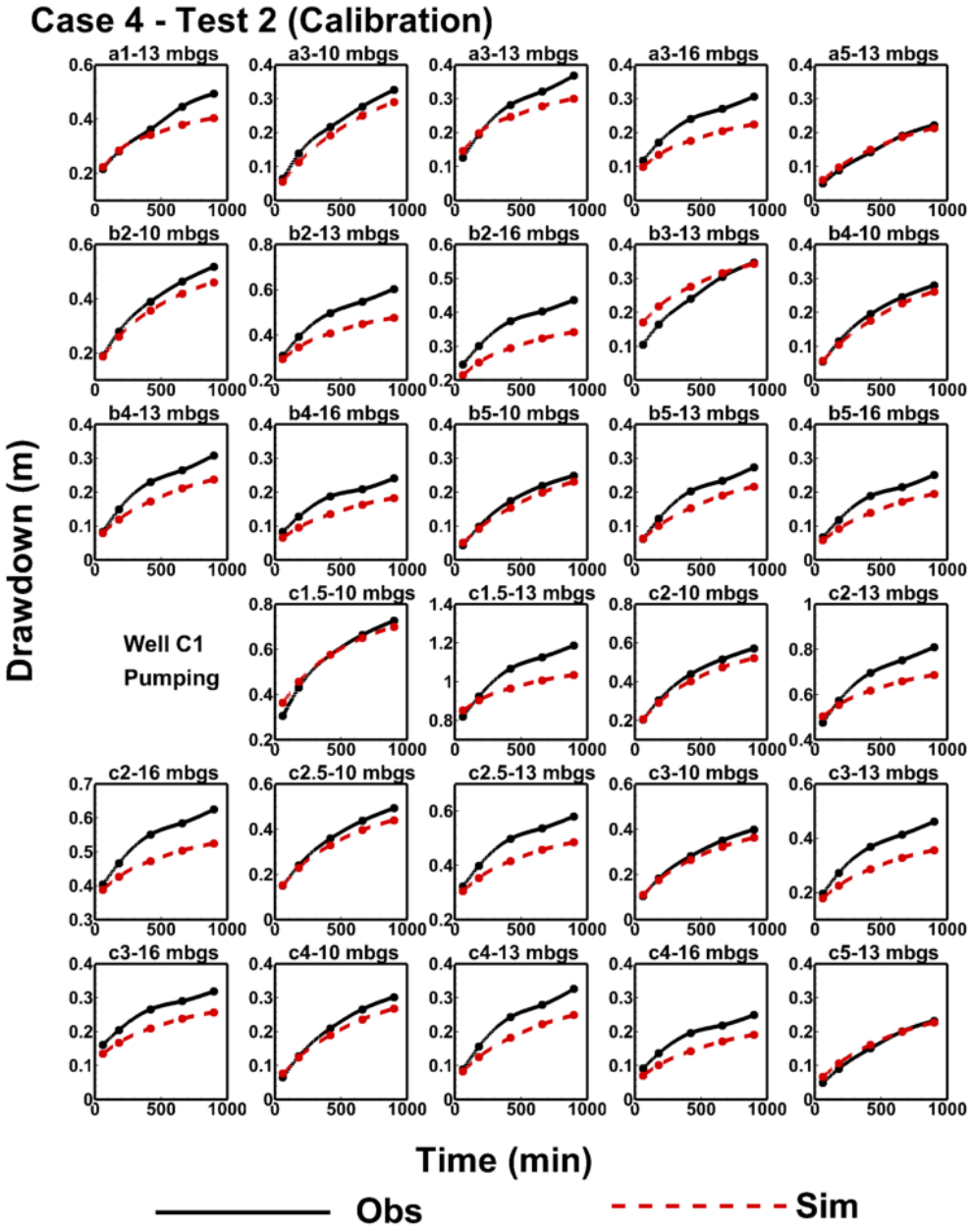

Fig. F9: *Observed and simulated drawdowns of Case 4 calibration results versus time curves at observation ports during HT Test 2 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 4. Pumping wells used in this HT test are listed in the plot as well.*

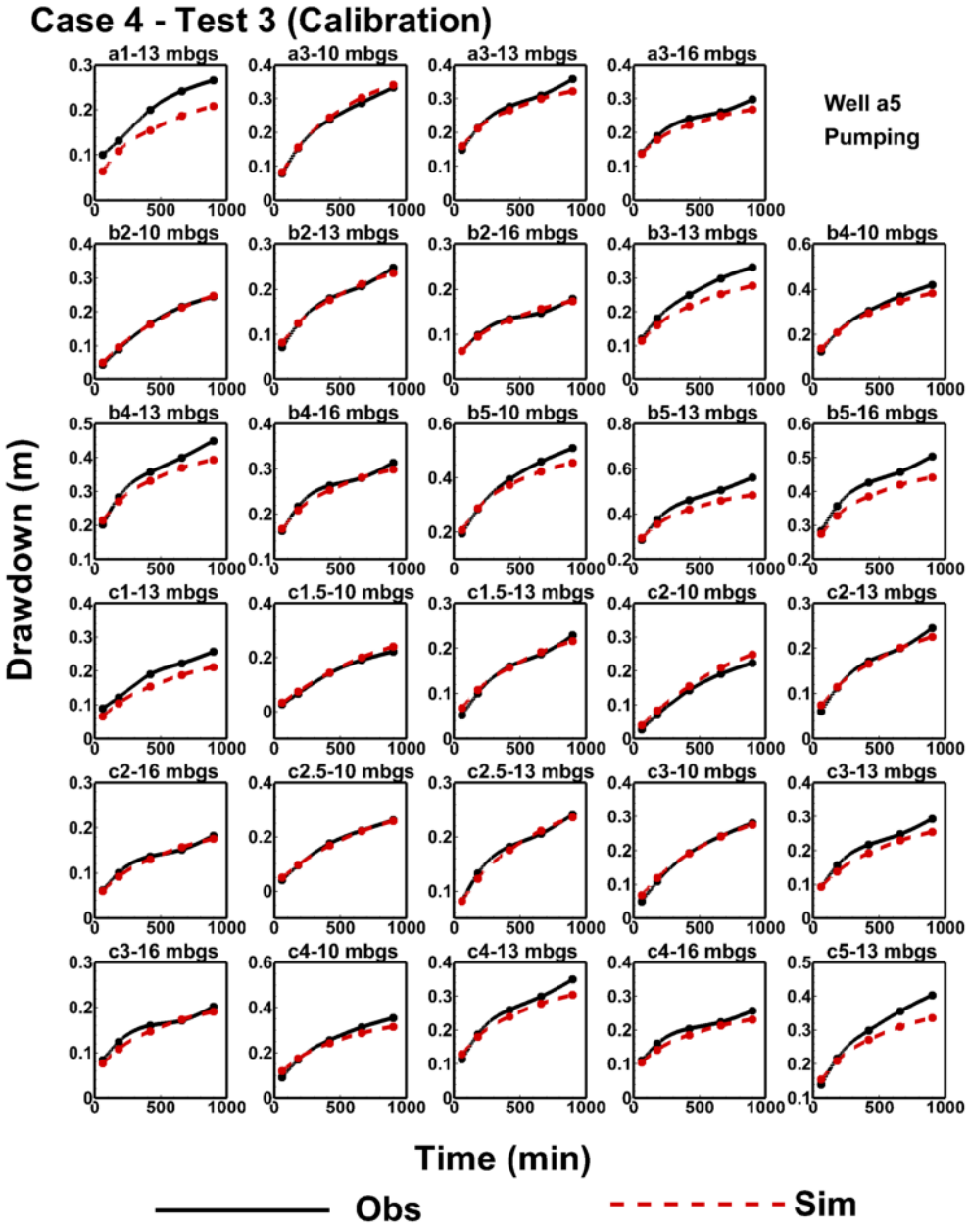

Fig. F10: *Observed and simulated drawdowns of Case 4 calibration results versus time curves at observation ports during HT Test 3 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 4. Pumping wells used in this HT test are listed in the plot as well.*

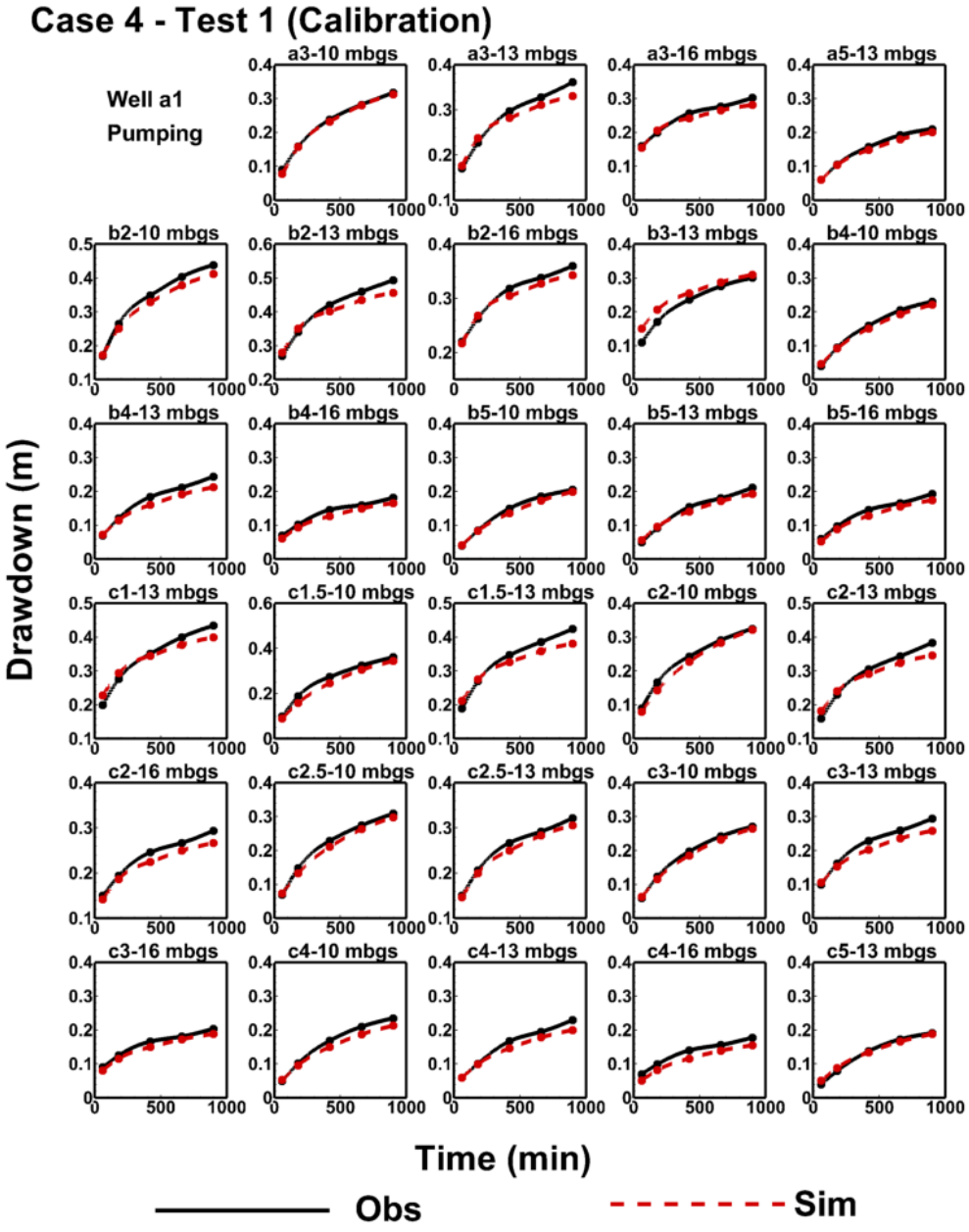

Fig. F11: *Observed and simulated drawdowns of Case 4 calibration results versus time curves at observation ports during HT Test 1 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 4. Pumping wells used in this HT test are listed in the plot as well.*

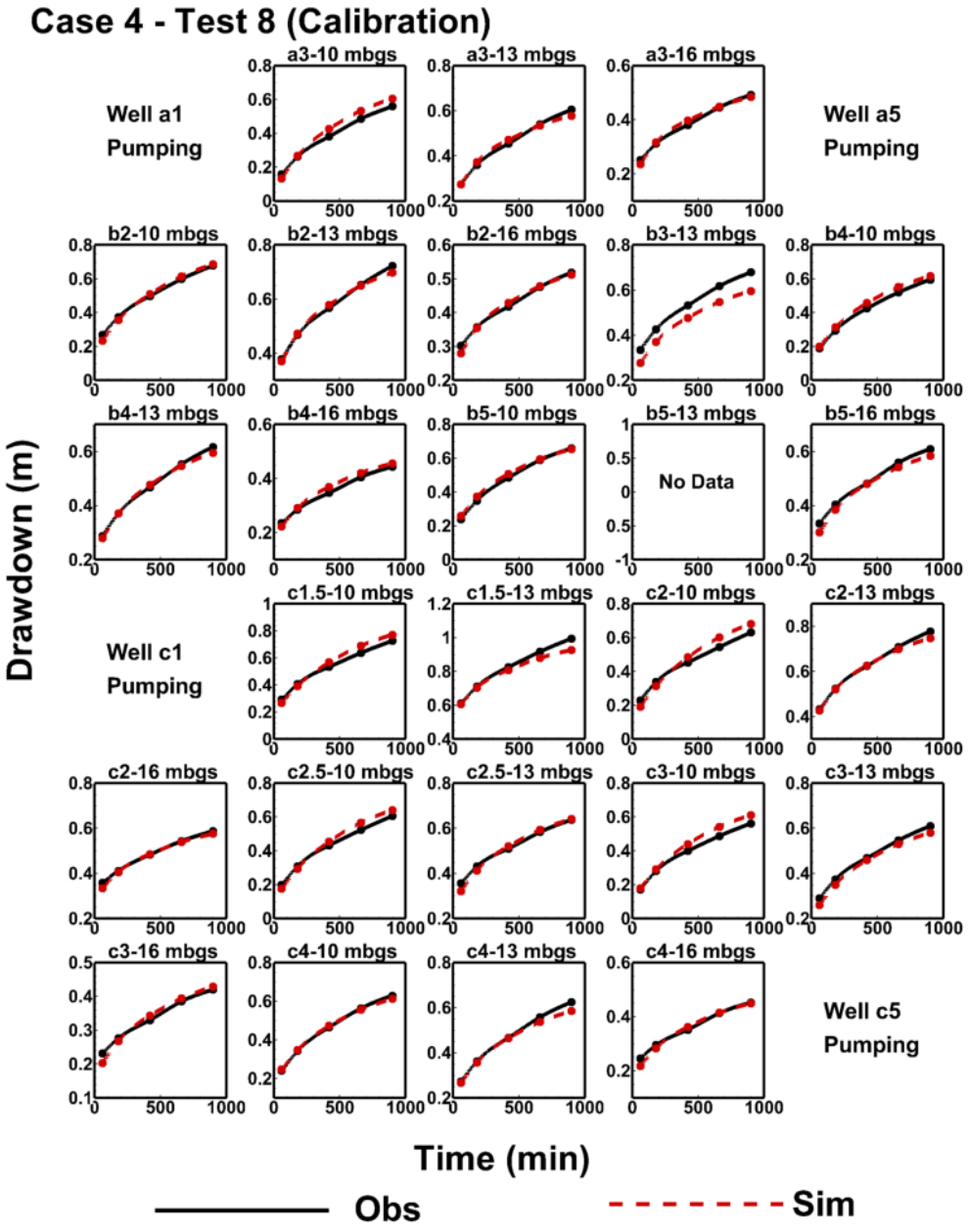

Fig. F12: *Observed and simulated drawdowns of Case 4 calibration results versus time curves at observation ports during HT Test 8 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 4. Note: "No data" indicates that data was not collected in that well during corresponding HT test. Pumping wells used in this HT test are listed in the plot as well.*

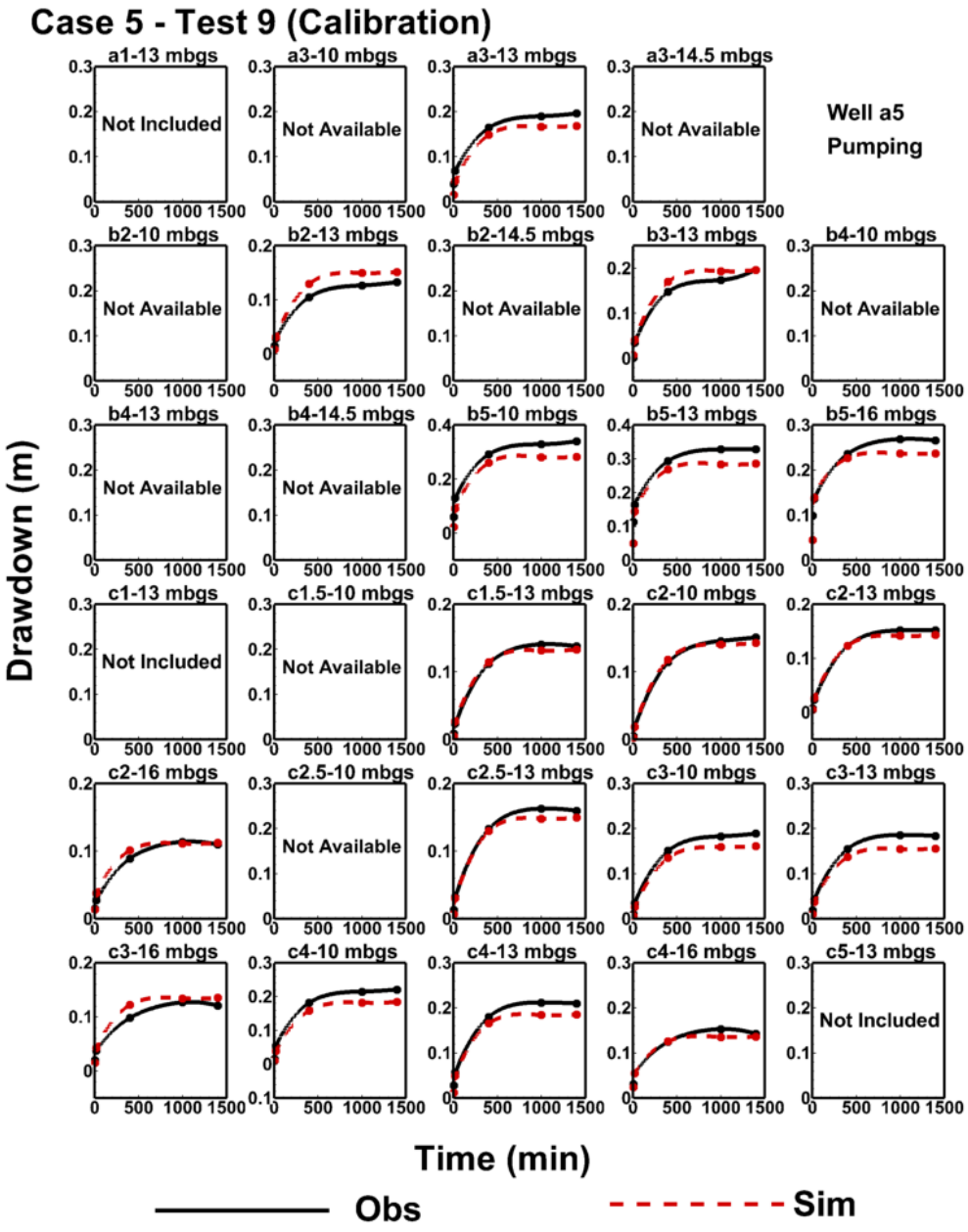

Fig. F13: *Observed and simulated drawdowns of Case 5 calibration results versus time curves at observation ports during HT Test 9 conducted in 2020. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 5. Note: "Not Included" indicates that data collected at corresponding well were not included in the inversion process, while "Not Available" indicates that data was not collected in that well during corresponding HT test. Pumping wells used in this HT test are listed in the plot as well.*

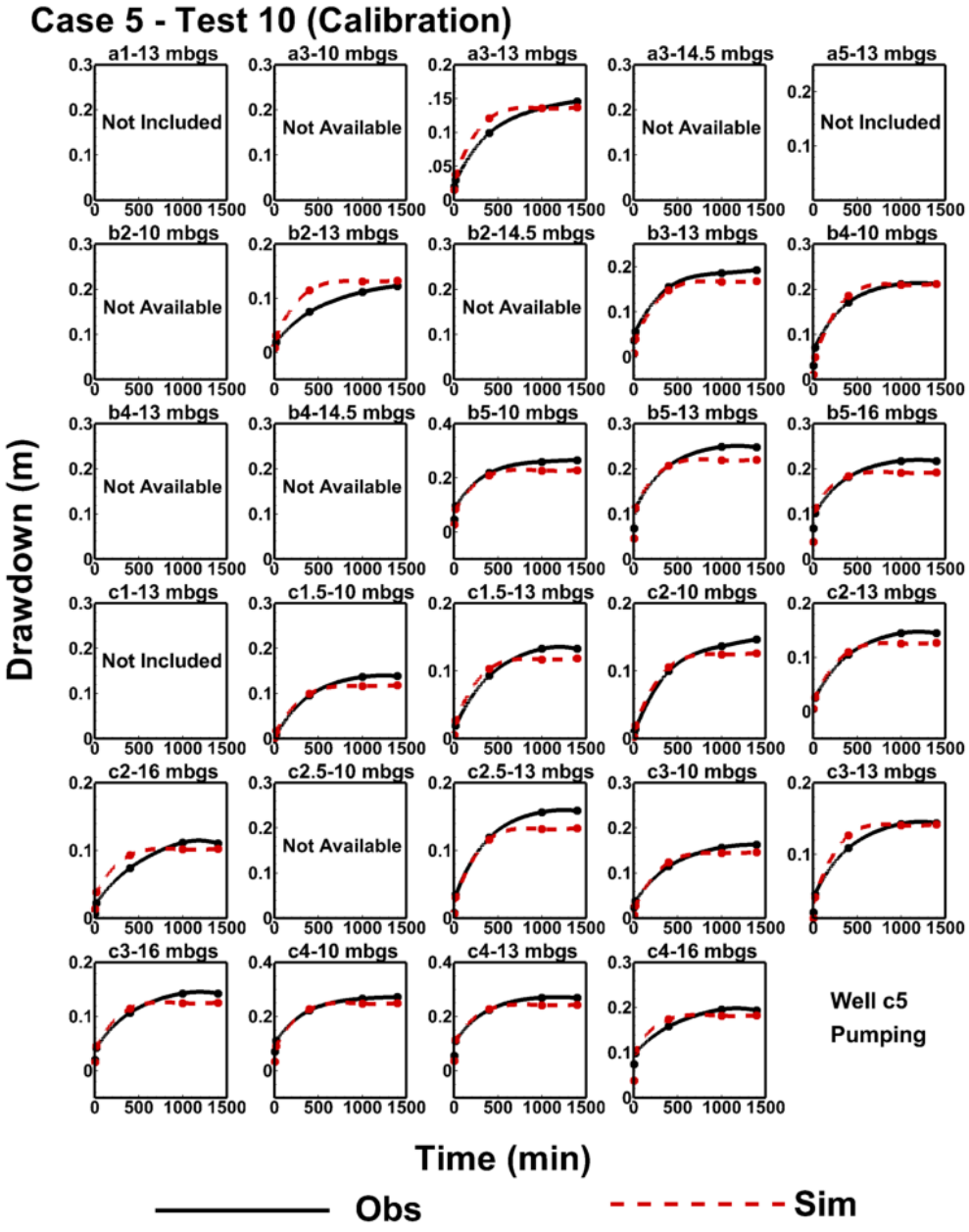

Fig. F14: *Observed and simulated drawdowns of Case 5 calibration results versus time curves at observation ports during HT Test 10 conducted in 2020. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 5. Note: "Not Included" indicates that data collected at corresponding well were not included in the inversion process, while "Not Available" indicates that data was not collected in that well during corresponding HT test. Pumping wells used in this HT test are listed in the plot as well.*

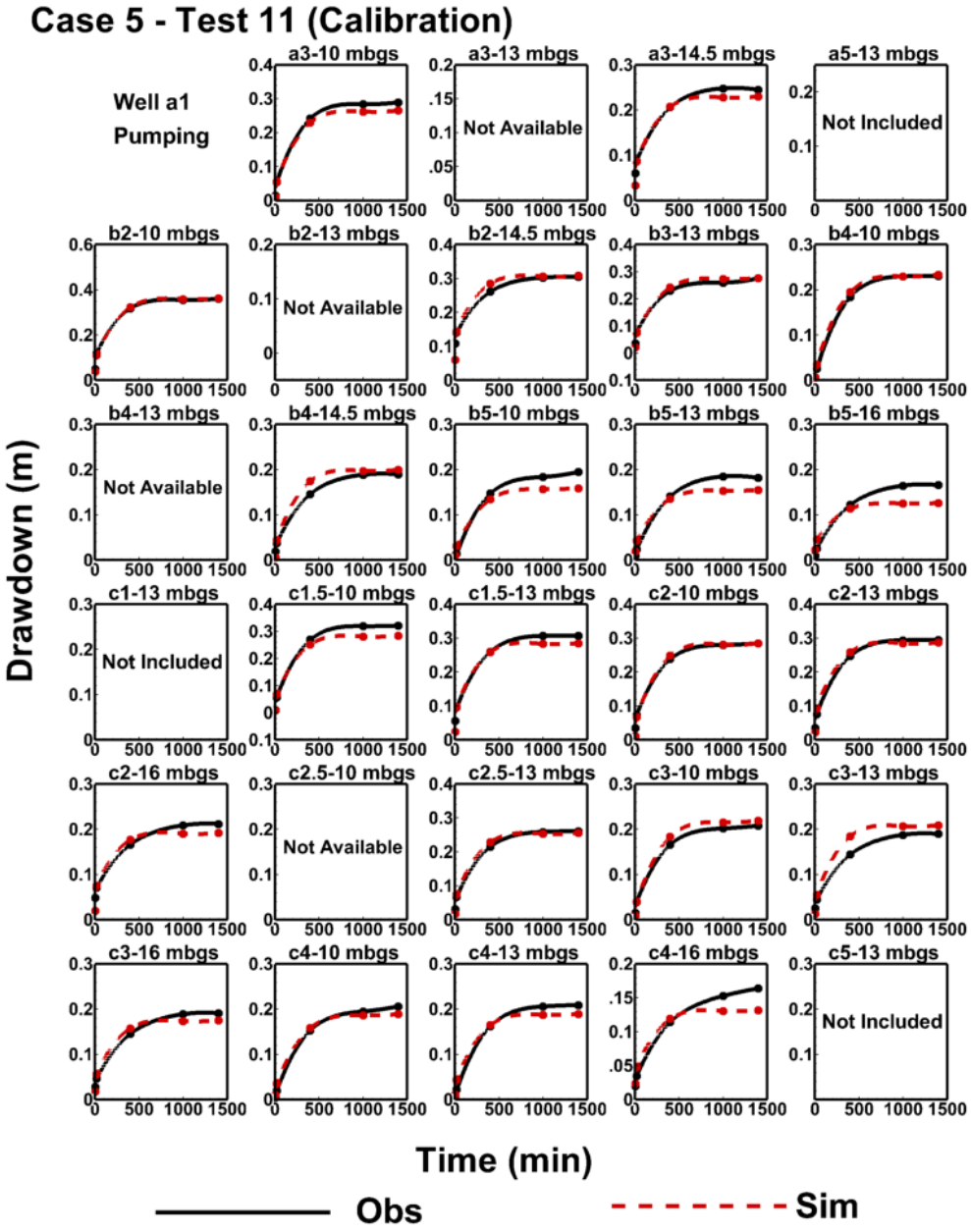

Fig. F15: *Observed and simulated drawdowns of Case 5 calibration results versus time curves at observation ports during HT Test 11 conducted in 2020. The solid black line represents the observed drawdowns, while the dashed red line represents the calibrated drawdowns of Case 5. Note: "Not Included" indicates that data collected at corresponding well were not included in the inversion process, while "Not Available" indicates that data was not collected in that well during corresponding HT test. Pumping wells used in this HT test are listed in the plot as well.*

Appendix G

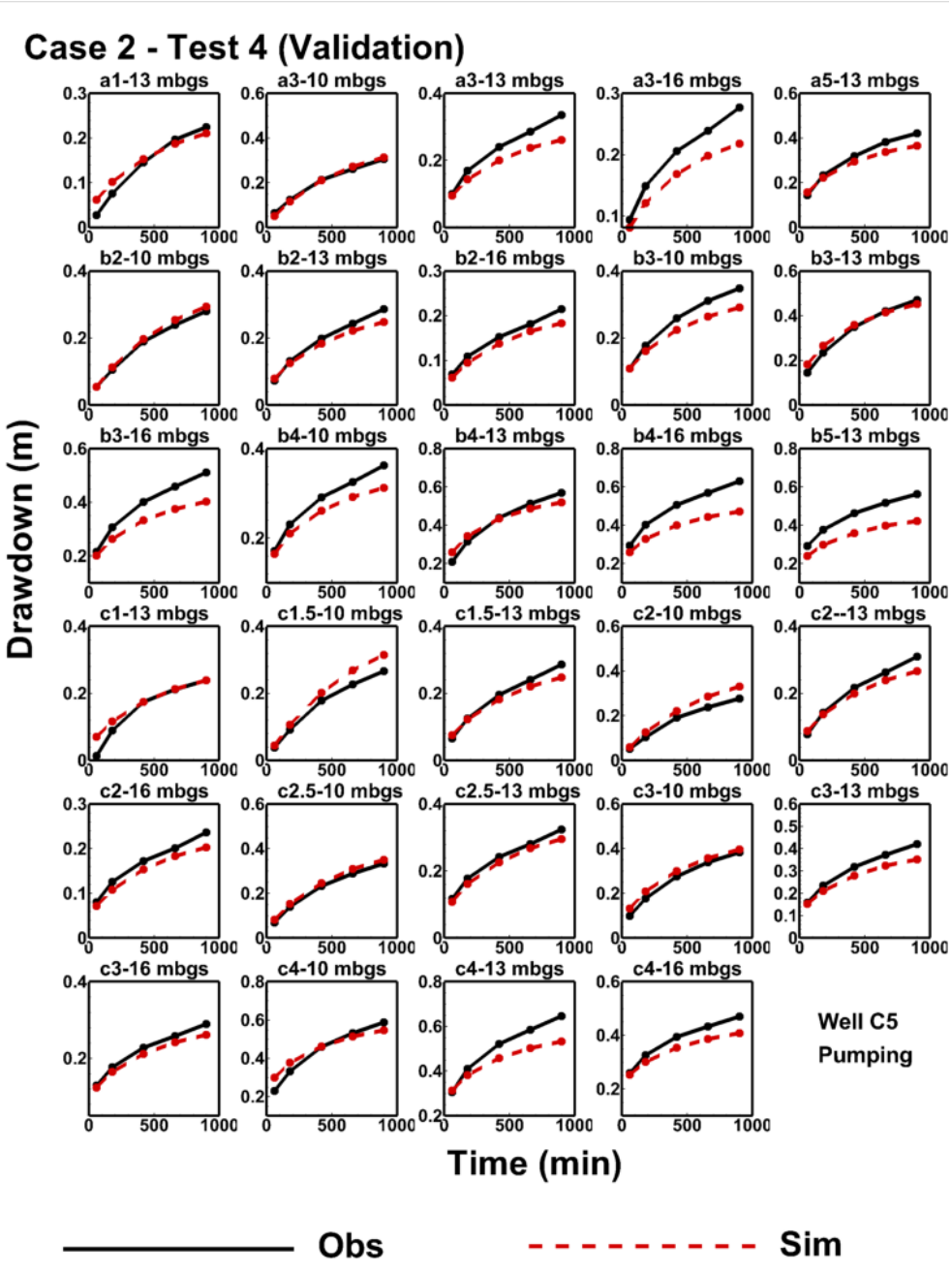

Fig. G1: *Observed and simulated drawdowns of Case 2 validation results versus time curves at observation ports during HT Test 4 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the simulated drawdowns of Case 2. Pumping wells used in this HT test are listed in the plot as well.*

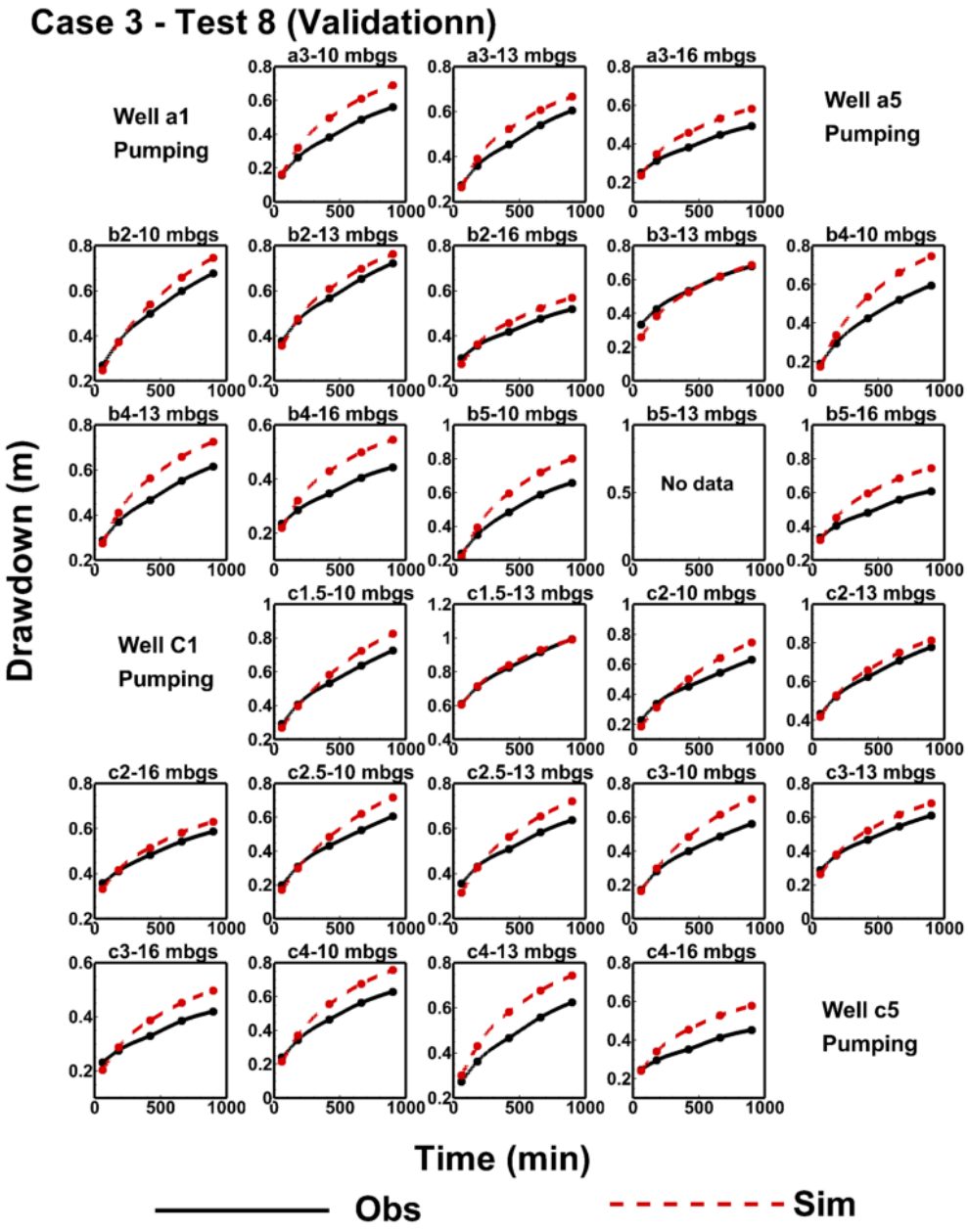

Fig. G2: *Observed and simulated drawdowns of Case 3 validation results versus time curves at observation ports during HT Test 8 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the simulated drawdowns of Case 3. Note: "No data" indicates that data was not collected in that well during corresponding HT test. Pumping wells used in this HT test are listed in the plot as well.*

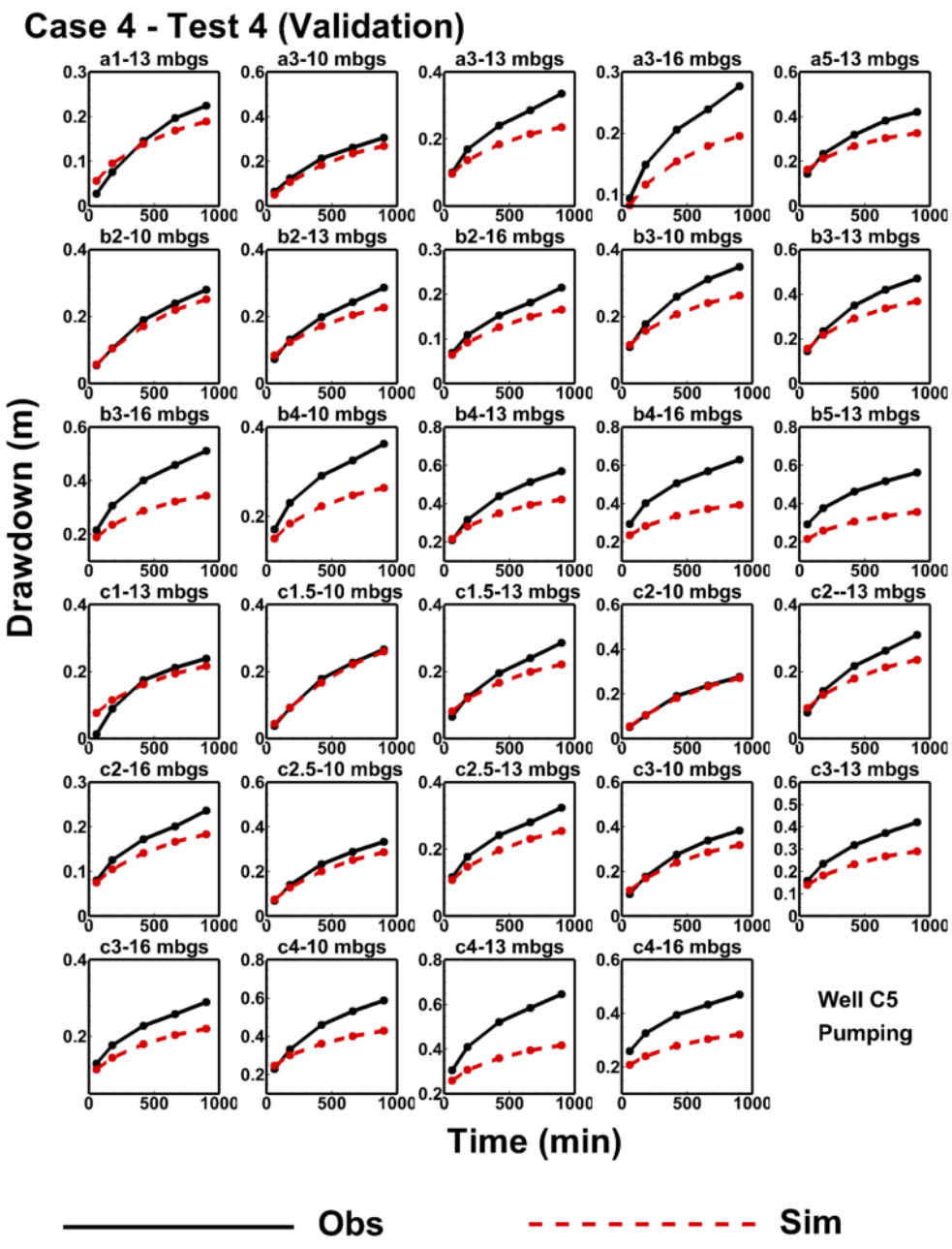

Fig. G3: *Observed and simulated drawdowns of Case 4 validation results versus time curves at observation ports during HT Test 4 conducted in 2019. The solid black line represents the observed drawdowns, while the dashed red line represents the simulated drawdowns of Case 4. Pumping wells used in this HT test are listed in the plot as well.*

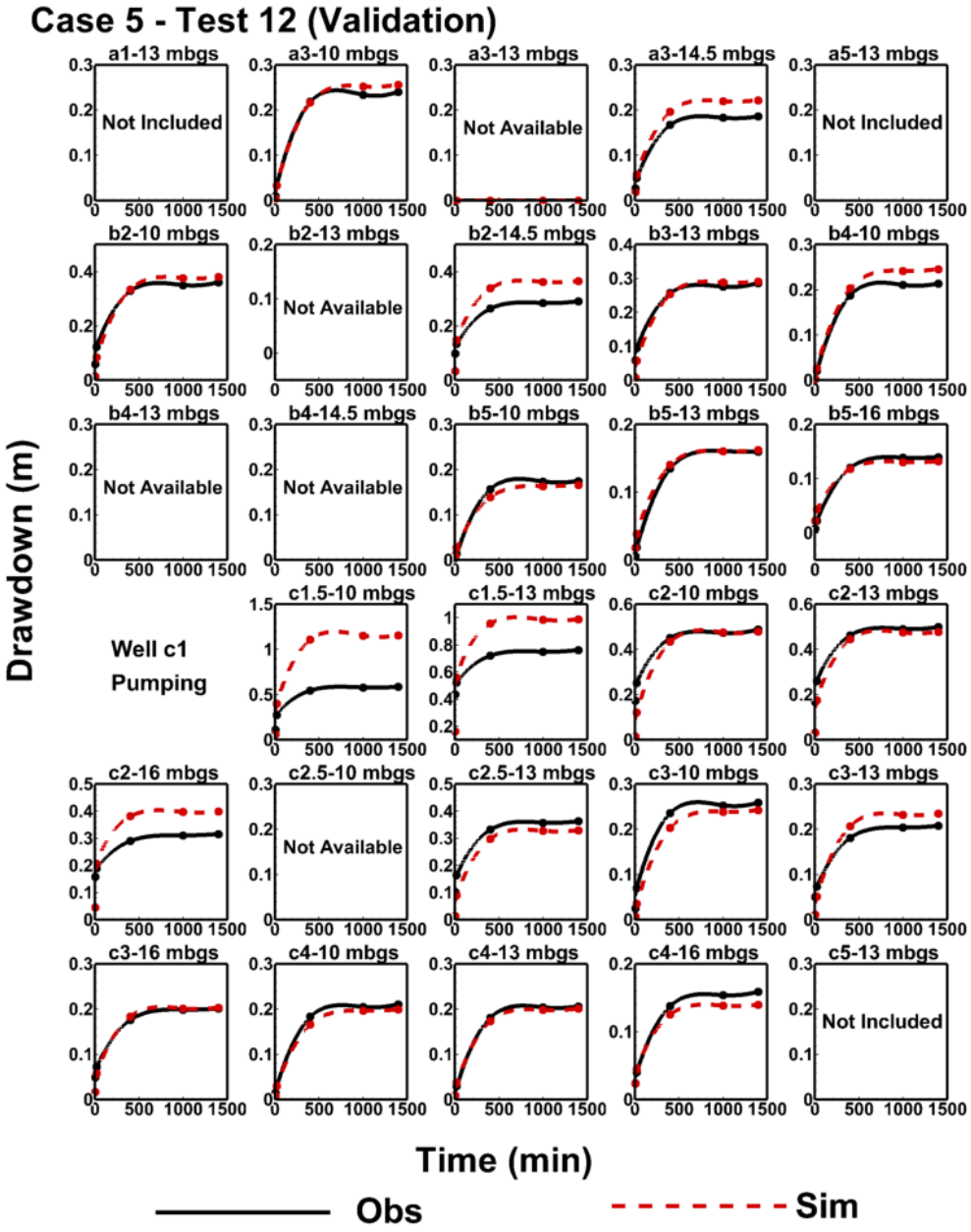

Fig. G4: *Observed and simulated drawdowns of Case 5 validation results versus time curves at observation ports during HT Test 12 conducted in 2020. The solid black line represents the observed drawdowns, while the dashed red line represents the simulated drawdowns of Case 5. Note: "Not Included" indicates that data collected at corresponding well were not included in the inversion process, while "Not Available" indicates that data was not collected in that well during corresponding HT test. Pumping wells used in this HT test are listed in the plot as well.*