Evaluation of Hydraulic Conductivity Collected by Various Approaches at a Highly Heterogeneous Field Site

by

Dongwei Sun

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Earth Sciences

Waterloo, Ontario, Canada, 2023

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

Significant research efforts have been conducted over the last several decades to better understand the groundwater flow and subsurface contaminant transport. It has been found that building a groundwater model for remediation decision-making requires an accurate delineation of spatial variation in hydraulic conductivity (K) and specific storage (S_s). Currently, numerous methods are available for site characterization. Traditional methods such as grain size analyses, permeameter and slug tests can provide point-scale estimates of K, while large-scale estimates from pumping tests are widely used for water-supply and water-quality investigations. However, when the degree of local heterogeneity increases, the necessary number of K increases dramatically, which presents a challenge to conventional methods. As a consequence, Direct Push (DP) based methods have been developed as efficient alternatives to conventional well-based approaches to provide K variability for shallow, unconsolidated aquifers. Hydraulic Profiling Tool (HPT) is one of the novel DP approaches designed for high-resolution site characterization with a test interval of about 1.5 cm. Various site-dependent formulae can be utilized to convert data collected during the HPT surveys into K estimates over a limited range. More recently, inverse modeling approaches of varying degrees of parametrization have become one of the most promising techniques to map hydrostratigraphic spatial variations between boreholes and identify heterogeneity characteristics with a level of detail never before possible. Many comparisons of diverse approaches have been performed, but there is no consensus on which approach yields parameters that are representative for field sites. The main objective of this study is to evaluate K estimates obtained via various site characterizations methods including: (1) grain size analyses; (2) falling head permeameter tests; (3) slug tests; (4) HPT with three different formulae; (McCall and Christy, 2020; Borden et al., 2021; and Zhao and Illman, 2022b)

(5) inverse modeling based on a geological zonation approach, and (6) a highly parametrized transient hydraulic tomography (THT) approach. The performance of each approach is first qualitatively analyzed by comparing it with site geology. A 19-layer geological model and forward groundwater model are employed to further assess various methods by simulating seven independent pumping tests that are not used for model calibration under both steady-state and transient-state conditions. Results reveal that the highly parametrized THT analysis with prior geological information yields the best results in model validation under both steady and transient states, and the generated K field revealed the most salient features of inter- and intra-layer heterogeneity. In contrast, traditional methods yield biased prediction of drawdowns, while HPT methods are primarily constraint by the limited range of estimates, especially for low permeable materials.

Acknowledgements

The HPT surveys conducted by Geoprobe Systems, Geotech Solutions Ltd. and the University of Waterloo (UW) at the NCRS were a result of discussions at the NovCare meeting held at the University of Waterloo during the summer of 2019. We are very grateful to Geoprobe Systems and Geotech Solutions Ltd. for visiting UW and training our staff and students to conduct the highly informative survey at the NCRS. The HPT data were collected by Aaron Vandenhoff and Ning Luo from the University of Waterloo, Wes McCall from Geoprobe Inc., and Jeff Bibbings from Geotech Inc. We thank these individuals for the use of HPT data as originally described in Sun et al. (2022).

I would like to thank Dr. Walter Illman for providing me with this invaluable learning experience and involving me in this important project. I extremely appreciate his guidance throughout this process. I would also like to thank my committee members, Dr. Andre Unger and Dr. David Rudolph for their advice.

This work would not have been possible without the collaborative efforts of Dr. Steven J. Berg, Matthew Alexander, and Dr. Zhanfeng Zhao, whose hard work at our research site continuously produced high-quality field and laboratory datasets, as well as excellent hydraulic tomography results that are vital to this work.

Finally, I received superb assistance from two Ph.D. candidates Ning Luo and Chenxi Wang. This work would not have gone so smoothly without their fruitful discussions on my progress presentations.

Dedication

This work is dedicated to my brilliant girlfriend, Xin Chen. You are that woman who transformed my imperfections into perfections. Your support through this process and tolerance of the usually long time spent without me will never be forgotten.

Table of Contents

Author's Declaration ii

Abstract iii

Acknowledgements v

Dedication vi

List of Figures ix

List of Tables xi

Chapter 1. Introduction 1

Chapter 2. Description of Field Site and Data Used for Analysis 8

2.1 Site description and hydrogeology 8

2.2 Available field data and the 19-layer geological model 9

Chapter 3. Description of Various Subsurface Characterization Approaches 13

3.1 Conventional methods 13

3.1.1 Grain Size Analysis 13

3.1.2 Permeameter Test 15

3.1.3 Slug Tests 16

3.2 Direct Push Methods 17

3.2.1 HPT methods 18

3.3 Inverse modeling methods 21

3.3.1 Calibration of the geological model 21

3.3.2 Highly parameterized transient hydraulic tomography (THT) 22

3.3.3 Averaged THT geological model 23

Chapter 4. Description of Groundwater Model and Experimental Design 24

Chapter 5. Statistical Analysis of Hydraulic Conductivity Collected from Various Site Characterization Approaches 26

5.1 Direct comparison of *K* estimates 26

5.2 Descriptive statistical analysis of zonation populated *K* 30

Chapter 6. Evaluation and Validation of Hydraulic Conductivity from Various Subsurface Characterization Methods Using a Groundwater Model 35

6.1 Comparison of *K* distributions 35

6.2 Results from model validation 37

6.3 Discussion 46

Chapter 7. Summary and Conclusions 51

References 55

Appendix A. Additional Tables 69

Appendix B. Additional Figures 77

List of Figures

Figure 1. a) Schematic configuration of wells used in the study at the NCRS including the CMT and PW network and nine NC wells where geological data are obtained, as well as 11 HPT profiles. Grey dashed lines represent four geological cross sections A-A', B-B', C-C' and D-D' as presented in Figure 2. b) 3D perspective of wells and DP locations around the 15m × 15m well clustering area shown as the blue dashed area in Figure 1a), corresponding well screen and pumping locations, bentonite sealings and high-resolution HPT survey intervals. 11

Figure 2. Cross-sectional view of the 19-layer geological zonation model with CMT and PW screened intervals shown in cross sections C-C' and D-D'. Cross section of A-A' and B-B' is available in Figure S1 in the Supplementary Information section. The 19 layers represent 7 different material types as indicated in the stratigraphic index. Specifically, the 19 layers are clay (1, 4, 8, 12, 16, 18), silt and clay (17, 19), silt (2, 7, 10, 14), sandy silt (6, 9, 13), silt and sand (5), sand (3, 11) and sand and gravel (15). 12

Figure 3. Vertical profiles of $\log 10K$ $(\mathrm{m/s})$ estimates from grain size analyses using three models, permeameter tests, slug tests, HPT with three formulae, as well as the highly parameterized THT analysis along site stratigraphy at borehole, DP, and elemental locations... 26

Figure 4. Box-and-whisker plots of *K* estimates from various site characterization methods for 19 layers of the geological model. 31

Figure 5. KG estimates from various site characterization approaches for 19 layers of the geological model. 33

Figure 6. *K* distributions at the NCRS from various site characterization approaches. Positions of CMT and PW wells (red lines) along with their screened intervals (black colour) as well as HPT loggings (dashed pink lines) are shown on each subfigure..... 35

Figure 7. Scatterplots of observed versus simulated drawdowns from various *K* characterization approaches for model validation under steady state conditions..... 41

Figure 8. Scatterplots of observed versus simulated drawdowns from various *K* characterization approaches for model validation under transient conditions. 42

Figure 9. *L*1 norms of observed versus simulated drawdowns from various *K* characterization approaches through the forward simulations of seven independent pumping tests under transient conditions. 43

Figure 10. Simulated drawdown curves from forward simulations with various *K* characterization approaches at a) CMT and b) PW wells from PW1-3. Observed drawdown curves are plotted as black curves and the data selected for computation of model performance metrics and scatterplots are solid black circles. 45

List of Tables

Table 1. Descriptive statistics of K from various site characterization approaches at the NCRS.29

Table 2. Summary of KG estimates for the 19-layer geological model from various site characterization approaches. 34

Chapter 1. Introduction

Significant research efforts have been expended over the last several decades to better understand groundwater flow and subsurface contaminant transport. Considerable progress has been achieved both in advancing theories and developing models (e.g., Hazen, 1911; Hvorslev, 1951; Freeze & Cherry, 1977; Puckett et al., 1985; Barr, 2001; Zhu & Yeh, 2005; Xiang et al., 2009; McCall & Christy, 2020; Borden et al., 2021; and Zhao & Illman, 2022b). Groundwater flow patterns, contaminant transport and their subsurface distributions have been found to be primarily governed by the spatial distribution of hydraulic conductivity (K) and specific storage (S_s), but the accurate delineation of such parameters is very difficult in complex groundwater flow systems due to strong heterogeneity of the geological media (Sudicky & Huyakorn, 1991). Besides, the development of accurate groundwater flow models for remediation decision-making is critically dependent on the ability to provide adequate and precise hydraulic parameter estimates. Inaccurate hydraulic parameter estimates will lead to poor groundwater flow and solute transport predictions (Berg & Illman, 2011, 2013, 2015; Zhao & Illman, 2017, 2018; Zhao et al., 2022). In addition, as it was clearly demonstrated by Rehfeldt et al. (1992) and Sudicky et al. (2010), to accurately forecast the migration of groundwater flow, the required number of K increases significantly as the level of local heterogeneity increases, which presents a challenge to implementing traditional site characterization techniques.

Traditional methods such as empirical-relation-based grain size analysis, laboratory permeameter analysis of core samples, slug test and pumping test have been used in water-supply investigations for several decades. However, most of them are not capable of providing reliable and sufficient information about local heterogeneity efficiently (Butler, 2005; Alexander, 2009). For example, laboratory analyses of core samples such as grain size analysis and

permeameter test can provide small-scale estimates of K based on various analytical solutions at sampling locations. However, they are usually time-consuming, notwithstanding the low sample recovery rate at high permeable zone, deviation from using repacked samples and different lab conditions than nature in situ conditions (Klute & Dirksen, 1986; White, 1988). Moreover, the information between boreholes cannot be easily delineated without interpolating point-scale measurements by geostatistical methods such as kriging (Alexander et al., 2011; Berg & Illman, 2011b). Slug tests are usually conducted to provide small-scale K and S_s estimates of materials in the close proximity of screened interval, however, considerable care must be taken as the quality of well development activity will have a significant impact on the K and Ss estimates (Beckie and Harvey, 2002; Butler, 1997, 2002). In addition, using solutions that ignore inertial mechanisms such as the typical form of Hvorslev (1951) can lead to a significant overestimation of K (Butler et al. 2003). The K and S_s estimates from pumping tests are averaged parameters over a large volume, with zero information regarding the assessment of spatial variation (Leven & Dietrich, 2006). Moreover, conventional type curve and straight-line methods assume a homogeneous medium, which presents a challenge in analyzing complex groundwater systems that have strong heterogeneity (Theis, 1935; Cooper and Jacob, 1946). Additionally, if a pumping test is being performed in a contaminated aquifer, the management of the contaminated water pumped from the subsurface can lead to significant treatment costs. Therefore, to better capture subsurface heterogeneity, it is critically necessary to advance traditional characterization methodologies to be more efficient and precise in both time and space aspects.

During the past two decades, various direct push (DP) methods, such as DP slug test (DPST), DP permeameter (DPP), DP injection logger (DPIL) and hydraulic profiling tool (HPT), have been proposed as efficient alternatives to conventional well-based approaches to provide

vertical profiles of *K* variability for shallow, unconsolidated aquifers (Hinsby et al., 1992; Stienstra and van Deen, 1994; Lowry et al., 1999; Butler, 2002; Dietrich et al., 2008; McCall and Christy, 2010; Geoprobe, 2015). However, most DP approaches except for DPP are sensitive to skin effects as the probe is advanced into the formation, while it offers no information between test intervals and only provides small-scale *K* estimates (~ 0.4m) (Liu et al., 2012, 2019). The HPT, which is similar to DPIL, has been recognized as one novel DP approach to rapidly obtain high-resolution (~1.5cm) *K* profiles based on the ratio of water injection rate and corrected down-hole water pressure measured in situ (McCall and Christy, 2020). Existing formulae can only convert the measured HPT data into *K* estimates, and such models are usually site-dependent (Borden et al., 2021; Zhao and Illman, 2022b). The cross-sectional view of HPT pressure logs provides an opportunity to interpret heterogeneity between each survey (Zhao & Illman, 2022a). However, the assessment should be conducted with caution, as solely relying on pressure logs to delineate stratigraphic boundaries can be sometimes misleading (McCall & Christy, 2020).

Most approaches described here can only provide *K* variations in the immediate vicinity of a well or DP location, reliable information away from or between the boreholes is hard to obtain. A dense network of multilevel-sampler tracer tests can be used to reveal *K* variations between existing wells by analyzing a large number of point measurements where plume migration can be examined, but they are commonly not practical under limited time, budget and effort (Freyberg, 1986; Hess et al., 1992). As a result, geostatistical and inverse modeling methods such as kriging, effective parameter approach, stochastic inverse methods, calibration of the geological model, and hydraulic tomography (HT) have been developed and heavily tested through a number of synthetic (e.g., Yeh & Liu, 2000; Bohling et al., 2002; Xiang et al., 2009;

Zhu & Yeh, 2005; Zhao and Illman, 2022b), laboratory (e.g., Liu et al., 2007; Berg and Illman, 2011a; Zhao et al., 2016, 2022; Luo et al., 2017; Jiang et al., 2022) and field studies (Jones et al., 2008; Bohling & Butler, 2010; Alexander et al., 2011; Berg and Illman, 2011b, 2013, 2015; Zhao and Illman, 2017, 2018, 2022a; Liu et al., 2021). Kriging, however, normally relies on point-scale K measurements only, and the created K field is generally not calibrated to geological data. Calibration of geological models through trial-and-error or automated methods can produce representative values of K if the zonation is accurate. HT can invert all pumping tests sequentially or simultaneously, while treating the medium to be heterogeneous and highly parameterized (Berg & Illman, 2015; Zhao et al., 2016; Luo et al., 2017; Zhao & Illman, 2017, 2018). As a consequence, HT has become one of the most promising approaches in terms of mapping hydrostratigraphic spatial variations between boreholes and revealing salient heterogeneity features with a degree of detail that has never been attainable (Berg and Illman, 2011, 2013, 2015; Zhao and Illman, 2018; Tong et al., 2021; Luo et al., 2022; Ning et al., 2022; Zhao et al., 2022). This type of approach usually depends on the joint inverse modelling of pressure change or drawdown data measured at multiple monitoring wells during a series of pumping/injection tests with varying stressed intervals. Steady-state hydraulic tomography (SSHT) can provide K estimates, while transient hydraulic tomography (THT) can provide both K and S_s estimates. Despite most of the advantages, HT could generate non-unique solutions that can deviate from the true K and S_s fields. In addition, smooth distributions of K and S_s heterogeneity can be obtained at locations with insufficient monitoring data (Bohling & Butler, 2010; Berg & Illman, 2011b, 2013). Therefore, the accurate depiction of subsurface heterogeneity via HT is dependent on data fusion. As Zhao et al. (2016) and Luo et al. (2017) demonstrated through laboratory synthetic sandbox studies, providing geological models as prior

mean *K* distributions for HT is useful in terms of introducing regularization and preserving stratigraphic features. While calibrations of inaccurate zonation models can lead to unrealistic tomograms, the accuracy of parameters generated from such integration is heavily dependent on the ability to construct an accurate geological model from a limited number of boreholes.

Based on diverse data collection and interpretation approaches, three generations of site characterization methodologies, have so far been widely employed. The first generation is described as traditional methods including grain size analyses, permeameter, slug and pumping tests. The second generation is designated as DP approaches including DPST, DPP, DPIL and HPT. The third generation is designated as inverse modeling methods with various degrees of model parametrization including calibration of geological model and HT. However, the dilemma frequently encountered by hydrogeologists stems from the adoption of a variety of techniques to assess hydraulic properties. A significant amount of research has been conducted to examine the effectiveness of different approaches (Alexander et al., 2011; Berg & Illman, 2015; Brauchler et al., 2013; Butler, 2005; Butler et al., 2007; Chapuis et al., 2005; Cheong et al., 2008; de Marsily et al., 2005; Hussain & Nabi, 2016; Ishaku et al., 2011; Liu et al., 2012; Rosas et al., 2014; Song et al., 2009; Vienken & Dietrich, 2011; Zhao & Illman, 2018). Specifically, Ishaku et al. (2011) utilized various empirical formulae to analyze grain size sieve results. Alexander et al. (2011) compared many traditional methods including grain size analysis, permeameter, slug and pumping tests. Liu et al. (2012) assessed multiple DP approaches including DPST, DPP and DPIL. Zhao & Illman (2018) evaluated numerous inverse modeling methods such as effective parameter approach, calibration of geological model, and THT. While only a few of the studies have compared cross-generation approaches, Butler et al. (2007) assessed the first two generations including grain size analysis, DPST, and DPP. Brauchler et al. (2013) compared the

last two generations including DPIL and HT. However, only qualitative and quantitative analyses have been conducted to analyze the estimated *K*, and there is no consensus on which approach yields parameters that are representative of a field site. As a result, a rigorous assessment designed to validate *K* estimations from cross-generation methodologies in three-dimensions (3-D) through steady state and transient state drawdown data is still lacking.

The main objective of this study is to evaluate three generations of *K* estimates obtained from various data types including grain size data, repacked core samples, head change data from slug tests, flow injection rate and water pressure from DP surveys with the HPT, as well as pumping/injection rates and drawdown data from pumping/injection tests at the well-studied North Campus Research Site (NCRS) underlain by a multiple aquifer-aquitard system consisting of highly heterogeneous glaciofluvial deposits. In particular, we choose what we believe to be a large sample of representative, widely utilized and most promising site characterization methods in terms of dealing with heterogeneity. Approaches that are evaluated include: (1) grain size analysis; (2) permeameter test; (3) slug test; (4) HPT methods; (5) calibration of a geological model; and (6) highly parameterized THT analysis. It is crucial to understand that each approach differs in terms of the scale and resolution at which heterogeneity is captured as well as the types and quantity of data that they rely on. Since there is no information about the true *K* field and most traditional and DP methods cannot provide S_s estimates, the ability to predict drawdowns under steady state is the first metric utilized for evaluation. The best way to assess *K* estimates from HT, according to Liu et al. (2007) and Berg and Illman (2011a), is the independent prediction of pumping tests or other drawdown-inducing events that have not been used during model calibration. As a result, a 3-D forward groundwater model was developed using HydroGeoSphere (HGS) (Aquanty, 2019) for forward simulations of steady-state drawdown data

from seven independent pumping tests that were not used for K estimation by any method evaluated in this study. Then, transient forward simulations were performed for simulations of transient drawdown data from the same pumping tests. This research was designed to examine three generations of site characterization techniques to answer the question of which method can provide sufficient information to accurately forecast groundwater flow through heterogeneous geological media. The complex glaciofluvial and heterogeneous field conditions are frequently seen in southern Ontario, as a result, the study findings may provide some valuable information about other locations with comparable subsurface characteristics.

Chapter 2. Description of Field Site and Data Used for Analysis

2.1 Site description and hydrogeology

The NCRS is situated in Waterloo, Ontario, Canada, on the campus of the University of Waterloo (UW). The primary topographic feature of the area is Waterloo Moraine which is a highly heterogeneous mixture of glaciofluvial deposits and tills. Deposits around and below the surface are mostly an outcome of advances and retreats of the Laurentide ice sheet lobes during glaciations. Tills covering and concealing the bedrock are laid down directly by the ice, mixing all sizes of materials from clay to boulders (Karrow, 1993).

Karrow (1979) drilled a 50-meter-long borehole to obtain a continuous core sampling of the materials down into the bedrock. According to the drilling report, below the top organic soil is a thin silt layer, followed by the Tavistock till which is composed of sandy-to-clay silt but only exists as erosional remnants. This till is underlined by a 3-meter-thick sand sequence, followed by the silty clay Maryhill till and dense Catfish Creek till, which consist of silty sand and stony silt. The Catfish Creek till extends approximately 20 meters below the ground surface and has been treated to be the hydraulic barrier of the field site (Alexander, 2009). Subsequent work by Sebol (2000) and Alexander et al. (2011) has revealed that the primary characteristic of the site is the alternating and interfingering aquifer-aquitard system which is made up of two high-K units separated by a discontinuous low-K layer. The lower aquifer consists of sandy gravel, while the upper aquifer is comprised of sand to sandy silt. Hydraulic connections are known to be provided by the low K layer in between, and the aquifer can be considered to be semi-confined. Silts and clays were discovered above and below the aquifer zone. Local stratigraphy is discontinuous with the presence of stratigraphic windows, although there are hints of layering, neither of these features spans continuously across the entire study site.

2.2 Available field data and the 19-layer geological model

The schematic configuration of the wells used in the study at the NCRS is shown in Figure 1a. The blue dashed box represents a nine-well pumping and observation network. During the study of Alexander et al. (2011), four continuous multichannel tubing wells (CMT 1 – 4), each with seven observation ports and a pumping well (PW1) screened at eight different elevations were installed at the site. Continuous soil samples were collected with recovery rates ranging from 68.9% to 83.2% during well installations. To provide a comprehensive *K* profile for each borehole, 270 gain size analyses and 471 falling head permeameter tests were carried out by Alexander et al. (2011). 28 slug tests were also performed at each monitoring port of the CMT systems. Later, two multi-screened wells (PW3, PW5) and two well clusters (PW2, PW4) were installed and described in Berg and Illman (2011). 15 additional slug tests were performed at PW1, PW3 and PW5 by Xie (2015) using various analytical models. Nine pumping tests (PW1-3, PW1-4, PW1-5, PW3-3, PW3-4, PW4-3, PW5-3, PW5-4, and PW5-5) were conducted mainly within aquifer layers to conduct a Hydraulic Tomography (HT) survey. Zhao and Illman (2017) then ran six additional pumping/injection tests (PW1-1, PW1-6, PW1-7, PW2-3, PW3-1, and PW5-1) with longer durations to stress the aquitard units, and details of the 15 pumping tests were summarized in Zhao and Illman (2017). Additional permeameter tests have also been performed in all CMT and pumping wells. To date, a total of 270 grain size analyses, 642 permeameter analyses of core samples, 43 slug tests and 15 pumping and injection tests were performed within the CMT and PW system. Additionally, two cross-hole flowmeter measurements were conducted at PW 1-3 and PW 5-3, while monitoring vertical flow at other pumping wells. Various geophysical surveys were also performed at the NCRS, geoprobe direct-push (DP) surveys were first conducted in April of 2015 at eight EC wells to obtain electrical

conductivity (EC) profile. In late July and early August of 2019, Sun et al. (2022) carried out 11 HPT surveys (HPT1, HPT2, HPT3, HPT4. HPT5. HPT6, HPT6-2, HPT7, HPT8, HPT9, and HPT10). Geonics EM39 and Gamma39 were utilized to obtain EC and gamma profile from four CMT and all PW wells in September of 2022. Figure 1b is the 3D perspective view of wells and DP locations around the 15m × 15m well clustering area, corresponding pumping and observation locations, bentonite sealings and high-resolution HPT survey intervals.

Figure 2 is the cross-sectional view (directions of cross sections are indicated on Figure 1a) of the 3D geological zonation model containing 19 different layers representing seven different material types created by Zhao and Illman (2017) for the NCRS. The geological model is $70m \times 70m \times 17m$ in extent and is constructed by the commercial software Leapfrog Geo which is a modeling tool using a volumetric algorithm called dual kriging to quickly construct models directly from various data types (ARANZ Geo. Limited, 2015). Locations of the CMT and PW wells and screened intervals associated are shown in the C-C' and D-D' cross sections in Figure 2, and A-A' and B-B' cross sections in Figure S1 in the Supplementary Information section. The geological model has been created by interpolating the drill logs from the CMT and PW wells as well as additional nine wells (NC3, NC6, NC30, NC45, NC65, NC69, NC75, NC104, NC105; see Figure 1) described by Sebol (2000). The interpolated geology between boreholes based on known geology is a reasonable representation of the site. The complex and truncated layering of different soil types indicates the highly heterogeneous nature of the glaciofluvial deposit at the NCRS. The two main aquifer units are separated by various discontinuous low-permeable materials. In addition, most of the features pitch out and do not continue along the modelled extent, which matches the vast majority of collected core logs (Sebol, 2000).

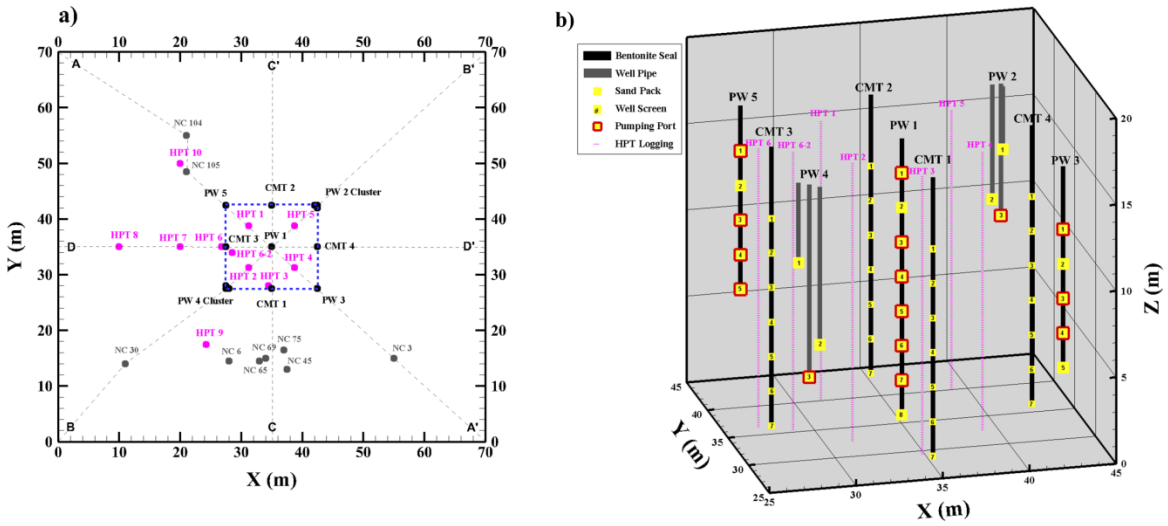


Figure 1. a) Schematic configuration of wells used in the study at the NCRS including the CMT and PW network and nine NC wells where geological data are obtained, as well as 11 HPT profiles. Grey dashed lines represent four geological cross sections A-A', B-B', C-C' and D-D' as presented in Figure 2. b) 3D perspective of wells and DP locations around the 15m × 15m well clustering area shown as the blue dashed area in Figure 1a), corresponding well screen and pumping locations, bentonite sealings and high-resolution HPT survey intervals.

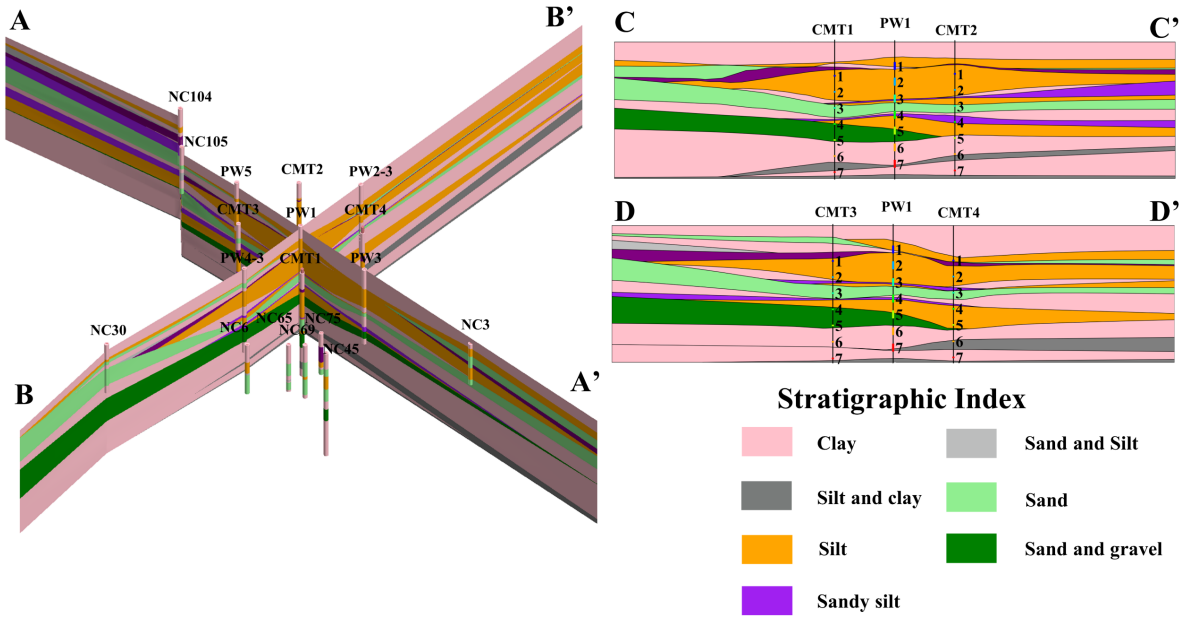


Figure 2. Cross-sectional view of the 19-layer geological zonation model with CMT and PW screened intervals shown in cross sections C-C' and D-D'. Cross section of A-A' and B-B' is available in Figure S1 in the Supplementary Information section. The 19 layers represent 7 different material types as indicated in the stratigraphic index. Specifically, the 19 layers are clay (1, 4, 8, 12, 16, 18), silt and clay (17, 19), silt (2, 7, 10, 14), sandy silt (6, 9, 13), silt and sand (5), sand (3, 11) and sand and gravel (15).

Chapter 3. Description of Various Subsurface Characterization Approaches

3.1 Conventional methods

3.1.1 Grain Size Analysis

The first method widely used for *K* estimation is grain size analysis. Many empirical formulae have been developed to establish relationships between *K* and particle size statistics (Ishaku et al., 2011; Rosas et al., 2014; Hussain & Nabi, 2016). This method is cost-efficient compared to other conventional approaches when it comes to obtaining rapid estimates of *K*, which avoids the need of conducting permeameter tests through core samples and the efforts to install wells for multiple slug tests. However, performing sieve analyses to obtain grain size distributions of multiple samples can be time-consuming. Alexander (2009) obtained 270 *K* estimates from grain size analyses using the empirical formulae of Puckett et al. (1985) and Hazen (1911). Between formulas, the *K* generated are obviously independent as the Puckett et al. (1985)'s formula, with a mean of 2.04E-06 m/s consistently generate greater conductivity than those estimated by the Hazen (1911)'s formula, which produced a mean of 2.34E-08 m/s.

The heterogeneous subsurface geological condition at the NCRS causes significant challenges to the analysis as most of the equations described in the literature were developed based on highly permeable materials, especially sand (Krüger, 1918; Terzaghi, 1925; Krumbein & Monk, 1943; Kozeny, 1953). Therefore, it is hard to determine if one dedicated empirical relationship is suitable for various unconsolidated materials. To confront the appropriateness of numerous relationships for a complex site, three different models are chosen to estimate *K* based on various material types.

Specifically, we chose Hazen (1911)'s model which is commonly used for estimating *K* of coarse-grained materials:

$$K = Cd_{10}^2 \quad (1)$$

where C is a coefficient depending on the grain size distributions of sand and d_{10} is the grain size (cm) in which 10% of the sample is finer.

Puckett et al. (1985) 's relationship was chosen to analyze fine grained materials since the equation only depends on clay content:

$$K = 0.0000436 \times e^{-0.1975 \times \%cl} \quad (2)$$

where $\%cl$ is the percentage of samples finer than 0.002 mm.

Barr (2001)'s formula was selected for silty materials assuming $d_{10} = 2r$, where r is the pore throat diameter (cm):

$$K = \frac{\rho g}{\mu} \times \frac{1}{(36)5C_s^2} \times \frac{n^3}{(1-n)^2} \times d_{10}^2 \quad (3)$$

where ρ is the density of the fluid (g/cm^3), g is the gravitational constant (cm/s^{-2}), u is the dynamic viscosity (Pa/s) of the fluid, n is the porosity and C_s is a shape factor ranging from 1 to 1.3 depending on the surface area, where 1 represents spherical grains, while 1.3 is used for angular grains. In addition, the fluid density and viscosity are temperature dependent and can be adjusted by:

$$\rho = 3.1 \times 10^{-8}\, T^3 - 7.0 \times 10^{-6}\, T^2 + 4.19 \times 10^{-5}\, T + 0.99985 \quad (4)$$

$$u = -7.0 \times 10^{-8}\, T^3 + 1.002 \times 10^{-5}\, T^2 - 5.7 \times 10^{-4}\, T + 0.0178 \quad (5)$$

where T is the water temperature in degrees Celsius ($°C$). The mean annual groundwater temperature of 7 $°C$ in waterloo area is utilized for calculation (Alexander et al, 2011). The porosity can be calculated through the empirical relationship given by Vukovic and Soro (1992):

$$n = 0.255 \times (1 + 0.83^{C_u}) \quad (6)$$

where C_u is the uniformity coefficient which can be calculated by:

$$C_u = \frac{d_{60}}{d_{10}} \quad (7)$$

where d_{60} is the grain size (cm) in which 60% of the sample is finer. A constant value of 1.15 was assigned to C_s in equation (3) to calculate *K* for all silt materials. The computational burden of utilizing various models is ameliorated by utilizing HydrogeoSieveXL (Devlin, 2015) which is a utility designed to give hydrogeologists a rapid and comprehensive way to rapidly obtain *K* estimates from grain size analyses data.

3.1.2 Permeameter Test

Another conventional method for understanding the characteristics of subsurface formation is to conduct laboratory analyses of repacked or original samples taken during well installation and borehole logging. Hydraulic conductivity (*K*) estimates can be obtained through several permeameter techniques. The constant head permeameter test is usually used for materials with moderate to high *K*, while a falling head permeameter test is primarily utilized to analyze materials with low permeability. During previous work by Alexander (2009) and Zhao and Illman (2017), a total of 642 temperature-corrected falling head permeameter analyses were performed on repacked samples to estimate *K* based on an equation provided by Freeze & Cherry (1977):

$$K = \frac{aL}{At} ln \frac{H_0}{H_1} \quad (8)$$

where a is the cross-sectional area (cm^2) of the manometer, L is soil sample length (cm), A is the cross-sectional area (cm^2) of the sample, t is the averaged time (s) from multiple trials, H_0 is the total head (cm) at the beginning of the test, and H_1 is the total head (cm) at the end of the test.

The falling head permeameter test is suitable for materials with moderate to high *K*, while low *K* material could also be tested, but the test could take a very long time to complete. By conducting a falling head permeameter test, vertical *K* is preferentially determined. As reported by Klute & Dirksen (1986), *K* of repacked samples estimated in the laboratory can be artificially lower than those from intact samples. In addition, the extraction and repacking process may induce fractures and destroy the internal structures that are well-preserved in intact samples. Sudicky (1988) demonstrated that the potential error caused by using repacked samples in permeameter tests to be small compared to the heterogeneity of the lithology within a highly heterogeneous site. Moreover, it is very difficult to recover substantial intact core samples from highly permeable zones (Butler, 2005; Alexander et al., 2011). As a consequence, underprediction of *K* is possible for permeameter tests conducted with materials from highly permeable intervals.

3.1.3 Slug Tests

Slug tests are important for water-supply investigations, as it is relatively easy to conduct and can yield *K* values that are representative of the aquifer around the well screen. A total of 43 slug tests were conducted by Alexander (2009) and Xie (2015), and *Micron System* pressure transducers were utilized to collect data automatically. To conduct a slug test, the ambient hydraulic head was perturbed by adding a slug of water. The increased head causes water to move out of the well and into the aquifer resulting in the head recovering over time until it reaches the initial head. Since the arrival of the slug is non-instantaneous, additional translation

was conducted based on Butler (1997), which neglects the noisy early time part of the data (Alexander et al., 2011; Xie, 2015). Most traditional slug test models are not consistent with the geometry of CMT wells, and for a screen length that is eight times longer than the radius, it is appropriate to use the following version of the Hvorslev model (Hvorslev, 1951):

$$K = \frac{r^2 \ln\left({}^{L_e}/_{R}\right)}{2L_e t_{37}} \tag{9}$$

where r is the effective radius (cm) of the piezometer, R is the screen radius (cm), L_e is the well screen length (cm) and t_{37} is the time (s) it takes for the water to fall to 37% of the initial value.

The slug test is suitable for materials that have moderate to low values of *K*, while high *K* materials could also be tested and analyzed. Moreover, the sampled volume of the slug test is usually considered to be considerably smaller compared to a pumping test, and the estimated hydraulic parameters are only representative of materials around the test interval, and usually not between boreholes based on site heterogeneity (Butler, 1997).

3.2 Direct Push Methods

Strata often exhibit considerable continuity in the lateral direction, but much less in the vertical. According to experimental research, hydraulic conductivity varies in a similar manner (Sudicky, 1986; Hess et al., 1992). Conventional methods are usually not capable of providing reliable and sufficient information about local heterogeneity efficiently, as a result, various DP approaches have been developed to rapidly characterize vertical variations of *K* along at finer resolution within unconsolidated formations. Direct-push slug test (DPST) and direct-push permeameter (DPP) are extensions of slug and permeameter tests where several tests can be performed at multiple elevations of interest in a single DP location. However, DPST is very sensitive to skin effects formed during the advancement of the probe resulting in a zone of

compaction around the probe and is relatively time-consuming to obtain a detailed vertical variation of *K*. DPP has shown to be less influenced by a skin effect, while it can only provide volumetric *K* estimates (~ 0.4m) and may encounter anomalies in highly heterogeneous media (Lowry et al., 1999; Butler, 2005; McCall & Christy, 2020). DPIL is one promising technique for measuring *K* with a pressure transducer placed on the surface (Dietrich et al., 2008). HPT is similar to DPIL, but the transducer is installed just above the probe tip down-hole, thus pressure measurements are taken continuously during probe advancement. This adjustment eliminates the necessity for friction loss correction in the supply line compared to DPIL (McCall and Christy, 2020).

3.2.1 HPT methods

Geoprobe System, Inc (Salinas, Kansas) began the design of HPT in 2006 (Geoprobe 2006). It has quickly become one of the most novel and efficient methods for high-resolution (~ 1.5cm) site characterization of unconsolidated formations based on the percussion probing method with a hydraulic hammer. Some of the most important components of the HPT are the replaceable injection port, downhole pressure transducer and electrical conductivity (EC) array. The injection screen is placed around 0.4 m above the probe tip, which avoids the influence of potential pressure build-up near the conical probe tip during the advancement, as it is commonly reported in cone penetrometer test (CPT) studies (Robertson et al., 1992). A robust downhole pressure sensor is utilized to measure the total pressure required to inject water into the formation. Additionally, the down-hole positioning of the sensor makes it possible to assess the ambient hydrostatic pressure at depth with several dissipation tests that are usually performed in transmissive zones. An EC array is located at the lower part of the probe, which can

simultaneously provide a bulk formation EC log as the probe is advanced into the formation (McCall and Christy, 2020).

During a HPT survey, an injection test is continuously performed by injecting water via the screen. Typically, water is injected at a rate of 200 to 300 mL/min, but in reality, flow rates below 100ml/min with the HPT system are not recommended as the flowmeter becomes unstable and reported flow rates are not accurate. Flow rates, HPT pressure and EC data are continuously recorded along with depth during advancement and data are displayed on a laptop computer screen through an acquisition software called DI Viewer (Geoprobe, 2015). A combination of pressure and EC log is useful for constructing site stratigraphy models based on some basic rubrics: (1) high pressure indicates low permeability formation and vice versa; (2) high EC implies the existence of clay material and therefore, decreased permeability. However, pressure or EC data alone is very complicated to be interpreted and may lead to misleading results since not all materials containing fine grains possess high EC. Ionic pollutants, chemicals used for remediation, and road salt may result in an artificially increased EC. Calcite cementing can lead to high pressure (McCall et al., 2014; McCall et al., 2017).

Various (site-dependent) formulae can be utilized to transfer the collected data to K measurements. For example, the measured HPT ratio of injection rate over corrected pressure can be explicitly converted to K estimates through an empirical formula developed by McCall and Christy (2010):

$$K = 21.14\, ln\left(\frac{Q}{P_c}\right) - 41.71 \tag{10}$$

where Q is the injection rate (mL/min) and P_c is the corrected water pressure (psi) which can be calculated by:

$$P_c = P_t - (P_{hdro} + P_{atm}) \tag{11}$$

where P_t is the total pressure (psi) measured by the pressure transducer, P_{hdro} is the hydrostatic water pressure (psi) measured through dissipation tests that are usually conducted in high permeable zones. P_{atm} is the ambient atmospheric pressure (psi).

The equation (10) was generated by fitting co-located HPT-ratio with slug test estimated *K* in an alluvial aquifer. The model was then evaluated by paring the HPT-ratio and slug test estimations from five other sites underlain by unconsolidated sediments (McCall & Christy, 2020). The model fit to data is significantly beneficial as the effect of dynamic movement, pressure generated by advancement and skin effects are all built into the equation. The biggest challenge for HPT is the relatively narrow range of estimate from $3.5E-07$ to $2.6E-04$ m/s, which ranges around three orders of magnitude. Data near the higher bound of estimates could be much higher and vice versa (McCall & Christy, 2020). Finding or fabricating a more sensitive pressure transducer that can withstand the force of hammering can potentially increase the higher range of estimates. However, Sun et al. (2022) have found that this limited range primarily presented a difficulty in estimating *K* for low permeable materials such as clay.

Borden et al. (2021) indicated that while the probe is rapidly advanced through aquitard units, the high pressure formed around the probe tip may not fully dissipate before the injection port passes through the tip location. In addition, the high stress around the probe tip may induce local shear failure. As the probe is advanced, the plasticized soil is pushed to the side, resulting in a less permeable zone (i.e., skin effect) surrounding the HPT injection port and therefore, leading to a higher pressure reading, and lower *K* estimates. As a result, a new relationship was developed based on numerical simulations of the physical flow process:

$$K = E(0.1235VD^2 + 0.119Q)P_c^{-1.017} \quad (12)$$

Where E is an empirically-derived efficiency factor to account for potential effects of soil displacement, V is probe rate of penetration (cm/s), and D is probe diameter (cm). The model was then assessed by paring 23 HPT survey K results using equation (12) assuming E = 1 to adjacent slug test K estimations along screen intervals at four different field sites. Continuous lower estimates of K were reported compared to slug test results, which is demonstrated to be the reason of skin effect, and the best E was then determined to be 2.02.

The efficiency factor from Equation (12) is established by averaging considerable fluctuations in centimetre-resolution results (HPT K measurements) over a substantial screened volume of the slug test, which is questionable. In addition, considering varying degrees of heterogeneity and conditions at different sites, empirical formulae developed from other sites may result in biased K estimates for other locations. To overcome these issues, a site-specific power law model for the NCRS without upscaling point-scale measurements based on inverse modeling was built by Zhao & Illman (2022b):

$$\begin{cases} logK = 1.41 \log\left(\frac{Q}{P_c}\right) + 4.25, \ \frac{Q}{P_c} < 5.5E-06 \ \ m^2/s \\ K = 6.9 \times 10^{-4} \ m/s \ , \frac{Q}{P_c} \geq 5.5E-06 \ \ m^2/s \end{cases} \quad (13)$$

The K estimates from the power law model have shown better corresponds with permeameter test results compared to those generated from the equation (10) at the NCRS.

3.3 Inverse modeling methods

3.3.1 Calibration of the geological model

One effective way for capturing the spatial variation of hydraulic parameters is to develop stratigraphic and zonation models in which the layers are calibrated to pumping tests or

other data through trial-and-error or automated methods (Martin & Frind, 1998; Jones et al., 2008; Berg & Illman, 2015; Zhao & Illman, 2017, 2018). Through laboratory synthetic sandbox experiments, Zhao et al. (2016) and Luo et al. (2017) discovered that calibrations of inaccurate zonation can result in unrealistic *K* and *Ss* estimates. As a consequence, the accuracy of estimated hydraulic parameters is strongly reliant on the ability to establish an accurate geological model.

The joint calibration of a 19-layer geological model built for the NCRS using transient drawdown data from eight pumping tests (PW1-1. PW1-4, PW1-6, PW1-7, PW2-3, PW3-3, PW4-3, and PW5-3) was conducted by Zhao & Illman (2018). The calibration was performed by coupling HGS (Aquanty, 2019) with the parameter estimation code, PEST (Doherty, 2004) while treating elements in each layer to be homogeneous and isotropic for individual layers. The simulations of groundwater flow during inverse modeling were conducted on the same 3-D domain that was used in this study for forward groundwater modeling and will be described in the next section.

3.3.2 Highly parameterized transient hydraulic tomography (THT)

A large number of synthetic, laboratory experiments and field studies have shown that HT is one of the most promising inverse methods for analyzing multiple pumping tests data to estimate hydraulic parameters sequentially or simultaneously (Yeh & Liu, 2000; Bohling et al., 2002; Zhu & Yeh, 2005; Xiang et al., 2009; Bohling & Butler, 2010; Liu et al., 2021; Tong et al., 2021; Jiang et al., 2022; Luo et al., 2022; Ning et al., 2022). Compared with traditional methods, HT can map hydrostratigraphic spatial variations between boreholes and reveal the most salient heterogeneous features with a level of detail that has never been attainable.

The Simultaneous Successive Linear Estimator code (SimSLE) developed by Xiang et al. (2009) was utilized by Zhao & Illman (2018) to inversely model transient drawdown data from the above-mentioned eight pumping tests, while treating each element to be homogeneous and isotropic. The inverse model results utilized for this study started with K and S_s values from the calibration of the geological model using the same grid.

3.3.3 Averaged THT geological model

The elemental K and Ss results from the highly parameterized THT analysis are then averaged for each layer based on the geological model to further assess the performance of HT and to compare with the results from the calibration of geological model using the parameter estimation code, PEST (Doherty, 2004).

Chapter 4. Description of Groundwater Model and Experimental Design

According to Liu et al. (2007) and Berg and Illman (2011a), the best way to assess *K* estimates from HT, is the independent prediction of pumping tests or other drawdown-inducing events that have not been used during model calibration. As a consequence, A 3-D forward numerical model with the same dimensions as the 19-layer geology model was built using HGS (Aquanty, 2019). The model was discretized into 31,713 rectangular finite elements of varying sizes and 34,816 nodes to simulate groundwater flow for model validation. From the central well cluster area to the model boundary, the element size gradually increases, with blocks expanding from 0.5m × 0.5m × 0.5m to 5m × 5m × 0.5m. In this study, the unsaturated zone is not considered, and the water table is designated as the model's upper boundary. The water table is modelled as a flat surface since the change in water level is less than the height of the elements at the top. The Catfish Creek till is identified as a hydraulic barrier and is served as the lower boundary of the model (Alexander et al., 2011). Both the top and bottom model boundaries were treated as impermeable boundaries, while the remaining four boundaries were treated as constant head boundaries. The computational mesh is provided as Figure S2 in the Supplementary Information section.

After obtaining *K* estimates from various approaches, qualitative and quantitative analyses were conducted from adjacent well and DP locations along with site stratigraphy. The estimated *K* were then populated into each layer of the geological model which was discretized in the same manner as the groundwater model to build the property fields. Descriptive statistical analysis based on box-and-whisker plots, qualitative and quantitative analysis of the layered *K* were then conducted to better compare and contrast the abundant data sets. Nine cases were created for forward groundwater simulations, specifically, case 1a is grain size analysis using

three models, case 1b is the permeameter test, case 1c is the slug test, case 2a is HPT using McCall and Christy (2010)'s model, case 2b is HPT using the Borden et al. (2021)'s model, case 2c is HPT using the power law model (Zhao and Illman, 2022b), case 3a is the calibration of geological model using the parameter estimation code, PEST (Doherty, 2004), case 3b is the calibration of the geological model using THT by smoothing the highly parameterized THT analysis results, while treating elements in each layer to be homogeneous, and case 3c is the highly parameterized THT analysis. Since there is no information about the true K field and most of the conventional and HPT methods are not capable of providing S_s estimates, the ability to forecast drawdowns under steady state is the first metric utilized for evaluation. As a result, forward simulations of seven independent pumping tests (PW1-3. PW1-5, PW3-1, PW3-4, PW5-1, PW5-4, and PW5-5) that were not used for calibration were conducted, and the results are summarized and presented in scatterplots. Then, transient forward simulations were performed for simulations of transient drawdown data from the same pumping tests. Specifically, the heterogeneous S_s generated from case 3a (PEST-calibrated geological model) are utilized in case 1a to case 2c for transient forward simulations.

Chapter 5. Statistical Analysis of Hydraulic Conductivity Collected from Various Site Characterization Approaches

5.1 Direct comparison of *K* estimates

Figure 1 shows that CMT1 is spatially close to HPT3, while CMT3 is close to both HPT6 and HPT6-2. Therefore, *K* estimates at CMT1 and CMT3 from permeameter tests, grain size analysis using three models, slug tests as well as elemental highly parameterized THT analysis could be compared with HPT results obtained at adjacent DP locations qualitatively and quantitatively.

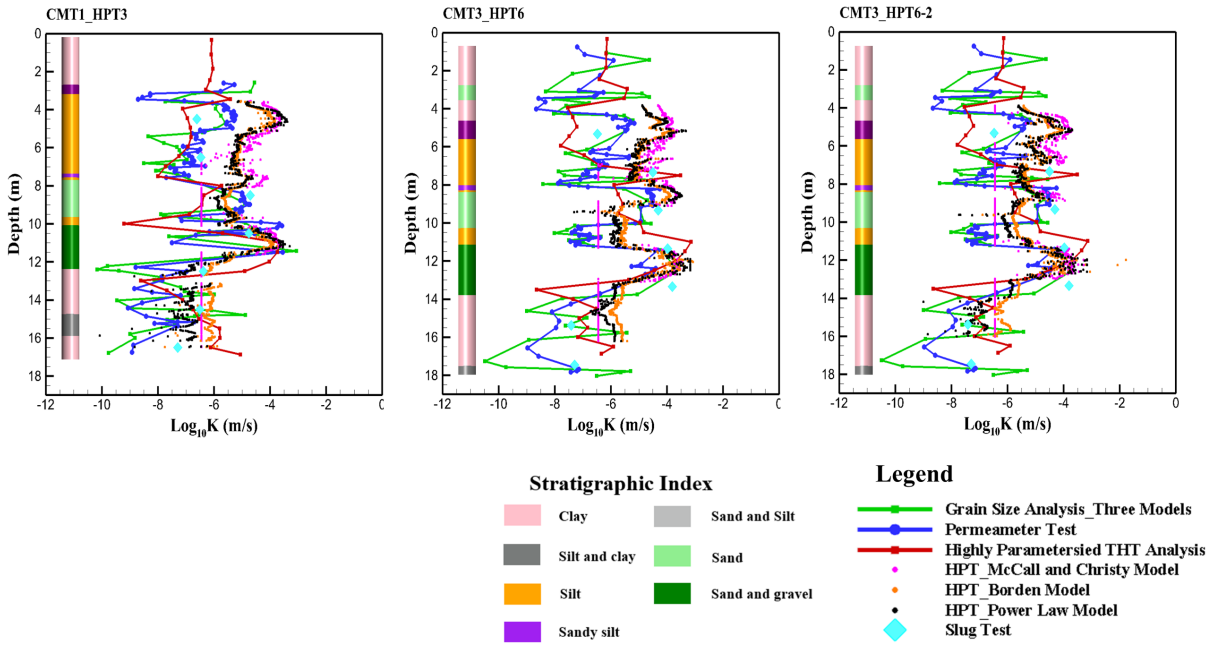


Figure 3. Vertical profiles of $\log_{10}K$ (m/s) estimates from grain size analyses using three models, permeameter tests, slug tests, HPT with three formulae, as well as the highly parameterized THT analysis along site stratigraphy at borehole, DP, and elemental locations.

In Figure 3, stratigraphy from core analysis, K estimations from various methods are provided as $\log_{10}K$ (m/s) along two CMT wells and at three DP locations with HPT surveys. Results show that K measurements are highly variable (around 7 orders of magnitude) across the two CMT wells indicating the highly heterogeneous nature of K at the site. Figure 3 reveals that K varies by several orders of magnitude within 1 m, thus interlayer heterogeneity could be captured by most of the methods, while only small-scale measurements such as grain size analysis, permeameter test, HPT method and the highly parametrized THT analysis reveal the salient intralayer heterogeneity which is usually hard to be detected through core logging. Additionally, the alternating aquifer-aquitard system is reflected in the datasets.

In terms of traditional methods, point-scale measurements of K from grain size analysis and permeameter test follow a similar trend. Slug test results also follow the general trend, but the measured K values are generally larger than the grain size analysis and permeameter test estimates, especially at highly permeable zone. Such behaviour is referred to as the scale effect, which is primarily caused by different sample volumes from various approaches. Therefore, the method that samples large volumes may encounter highly permeable zones not discovered by other methods that sample smaller volumes (Vesselinov et al., 2001; Illman and Neuman, 2003; Martinez-Landa and Carrera, 2005; Illman and Tartakovsky, 2006).

However, this phenomenon does not apply to HPT methods which sample extremely small intervals for high-resolution site characterization methods. HPT results at three DP locations generally follow the trend of K from permeameter tests, while the K estimates are around 1 to 2 orders of magnitude larger than those estimated by permeameter tests, especially from 4 m to 8 m where local geology is primary low permeable materials such as clay and silt (Fig. 3). This is mainly due to the limited range of estimates, which results in K that are, at best,

order-of-magnitude estimates for the *K* of the geological media. Using various site-dependent formulae yields similar results at this upper depth range. Significant differences are observed in the middle and lower portions of the site.

According to CMT1_HPT3 (Fig. 3), at around 8 m – 10 m, together with the core log, local geology is primarily highly permeable materials such as sand, thus *K* estimates from permeameter tests, grain size analyses and slug tests yielded relatively higher estimates of *K* than for silt materials located above and below, while HPT using McCall and Christy (2010)'s model only yields a fixed estimate of the lower bound ($3.5E-07$ m/s). The Borden et al. (2021) and the power law model (Zhao and Illman, 2022b) both provide more reasonable estimates that are higher than those generated from McCall and Christy (2010)'s relationship.

According to CMT3_HPT6 and CMT3_HPT6-2 (Fig. 3), at depth ranges from 9 m – 11 m, a transition zone from sand to silt is discovered by core logs, and permeameter tests and grain size analyses measurements both capture this variation. However, none of the HPT methods successfully detect the change, while McCall and Christy (2010)'s model only gives a fixed lower bound. The Catfish Creek till located at depths below 12 m for CMT1 and below 14 m for CMT3 is detected by a significant drop in *K* estimates from traditional methods. McCall and Christy (2010)'s model yields a fixed lower bound, while Borden et al. (2021)'s model generates even higher estimates of *K*. Surprisingly, the power law model (Zhao and Illman, 2022b) delineates the low permeable materials at HPT3 and HPT6-2 and yields estimates that are close to grain size analyses and permeameter tests. The elemental highly parameterized THT analysis results are also plotted for comparison (Fig. 3). From 0 m to 3 m, *K* estimates are quite smooth because there is no monitoring data available for inversion. Below 3 m, compared with other methods, *K* estimates from THT captures the small-scale interlayer heterogeneity. For example,

the transition zone at 3 m and 10 m from CMT1, are not captured by the other methods. However, THT results at CMT1 from 12 m to 16 m indicate a substantial increase in *K*, which does not conform to the characterized site geology. Thus, the stratigraphy information should be used with caution in practice.

Table 1. Descriptive statistics of *K* from various site characterization approaches at the NCRS.

Method	n	Minimum *K* (m/s)	Maximum *K* (m/s)	Geomean of *K* (m/s)	Range of $\log_{10}$ *K*	Variance of $\log_{10}$ *K*
Grain Size Analyses_Three Models	270	3.07E-11	2.50E-03	1.19E-07	7.91	2.63
Permeameter Tests	642	1.15E-10	4.63E-03	3.03E-07	7.60	1.55
Slug Tests	43	1.21E-08	1.68E-04	2.65E-06	4.14	1.47
HPT McCall and Christy Model	7660	3.53E-07	2.65E-04	2.85E-06	2.88	1.28
HPT_Borden Model	7660	1.13E-08	2.69E-04	5.78E-06	4.38	0.38
HPT_Power Law Model	7660	8.50E-11	6.90E-04	3.84E-06	6.91	0.86
Highly Parameterised THT Analysis	31713	4.20E-11	2.90E-03	5.79E-07	7.84	1.47

Table 1 summarizes the *K* statistics from various approaches across the site including minimum, maximum, geometric mean of *K* in m/s as well as the range of *K* in logarithm with a base of 10 and the population variance of $log_{10}K$. Which can be calculated by:

$$\sigma^2 = \frac{\sum(X-u)^2}{N} \tag{14}$$

where X is each individual value, u is the population mean and N is the number of values in the population.

Compared to conventional methods, the HPT method yields a significantly larger number of *K* due to high-resolution profiling intervals. The THT analysis has the largest number of *K*

measurements due to significant parameterization. The geometric mean of K (K_G) from HPT methods are higher than those generated through traditional methods. In addition, K_G increase from the grain size analyses to permeameter tests and to slug tests due to the scale effect. Moreover, McCall and Christy (2010)'s model has the smallest range of $\log_{10}K$ (m/s), while Borden et al. (2021)'s model extends the range especially at the lower end. In contrast, the power law model (Zhao and Illman, 2022b) extends the range in both the higher and lower ends and yields the largest range of $\log_{10}K$ (m/s) among the three formulae to interpret HPT data. Slug tests yield a relatively small range of K estimates, while grain size analyses with three models, permeameter tests, and THT analysis all yield larger range of estimates.

In terms of the variance of $\log_{10}K$, the value for slug tests is relatively high given the smallest number of measurements. Grain size analysis yields the highest $\log_{10}K$ variance probably because of the use of three models to target various soil types, while each formula has its assumption behind it. Remarkably, both Borden et al. (2021) and the power law model (Zhao and Illman, 2022b) yield a smaller $\log_{10}K$ variance given an increased range of estimates compared with the variance estimate based on McCall and Christy (2010)'s relationship. It is also noteworthy that the highly parameterized THT analysis yields 31,713 estimates of K, but results in a relatively small variance of 1.47.

5.2 Descriptive statistical analysis of zonation populated *K*

K measurements from various approaches are then used to populate the 19-layer geological model by taking the geometric mean of all data points located in each layer. Measurements from similar soil material are attributed to layers that have no sample data available. As only 43 K estimates in 11 out of 19 layers of the geological model are available from slug tests, it is most difficult to populate the model. As a result, K_G from layers 4, 8, 16 and

18 (clay) is assigned to layers 1 (clay); K_G from layers 2, 7 and 14 (silt) is assigned to layer 10 (silt), K_G from layer 13 (sandy silt) is assigned to layers 6 and 9 (sandy silt), K_G from layer 11 (sand) is assigned to layer 3 (sand), K_G from layer 17 (clay & silt) is assigned to layer 19 (clay & silt), and K_G from layers 3 and 11 (sand) and 2 and 10 (silt) is assigned to layer 5 (sand & silt). Similar but simpler steps are also performed for the other methods if there are layers that do not contain any K estimates. Additionally, the maximum, upper quartile, median, K_G, lower quartile and minimum are also calculated for each layer, and details are summarized in Tables S1 to S7 (in the Supplementary Information section) and plotted as box-and-whisker plots in Figure 4. The lower and higher range of K estimates from McCall and Christy (2010)'s model and the higher range of the power law model (Zhao and Illman, 2022b) are indicated beside the box plot.

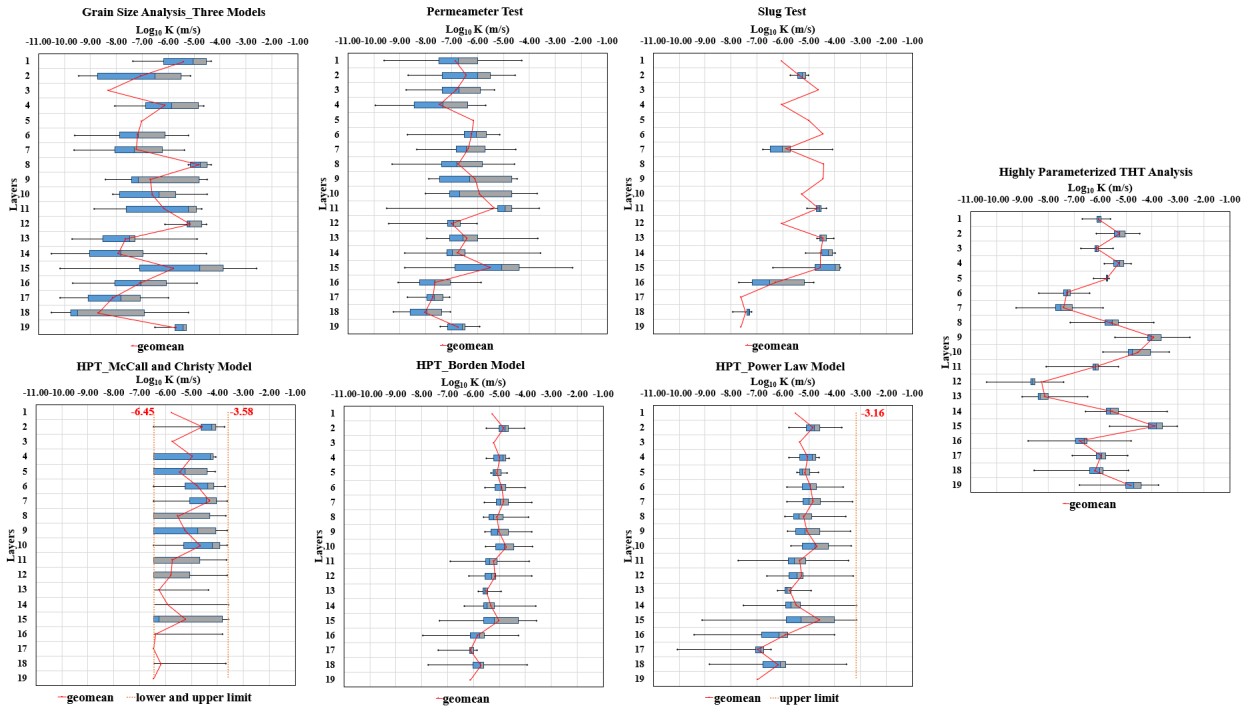


Figure 4. Box-and-whisker plots of K estimates from various site characterization methods for 19 layers of the geological model.

Examination of Figure 4 reveals that most of the box plots are either positively skewed or negatively skewed, and there is rarely a normal distribution. In addition, the interquartile range (IQR) of K estimates from the permeameter tests and grain size analyses over 19 layers is generally larger than those from other methods, suggesting larger spreads of K estimates in each layer. The IQR for the highly parameterized THT analysis is consistently smaller, which indicates less dispersion of the data set throughout. The IQR from McCall and Christy (2010)'s model is larger than the Borden et al. (2021) and power law (Zhao and Illman, 2022b) models. In addition, the K_G from Borden et al. (2021)'s model is less variable than the McCall and Christy (2010) and power law (Zhao and Illman, 2022b) models.

Next, the K_G from various approaches together with calibration of geological model results using PEST through 19 layers are plotted together as shown in Figure 5, and statistics are summarized in Table 2. Figure 5 reveals that it is very hard to accurately characterize a heterogeneous site such as the NCRS as $\log_{10}K$ (m/s) could range about 4 orders of magnitude within a single geological unit using various site characterization approaches. Overall, the slug test yields higher K estimates than from grain size analyses and permeameter tests. HPT K estimates using three different models yield similar K estimates, while the estimates are generally higher than those generated from conventional methods (i.e., grain size analysis and permeameter test). The K values estimated from the PEST-calibrated geological model are close to those generated from the averaged THT geological model. Table 2 shows that the PEST-calibrated geological model has the largest range of K estimates followed by the averaged THT geological model and grain size analyses using three models. It is surprising to find out that after populating K estimates, Borden et al. (2021)'s model yields the smallest range of K estimates, while the power law model (Zhao and Illman, 2022b) has a slightly larger range than McCall and

Christy (2010)'s model. In addition, K_G values from permeameter tests and grain size analyses are relatively smaller than those generated from PEST and averaged THT geological models, while the slug test and three HPT formulae all yield relatively larger K_G estimates.

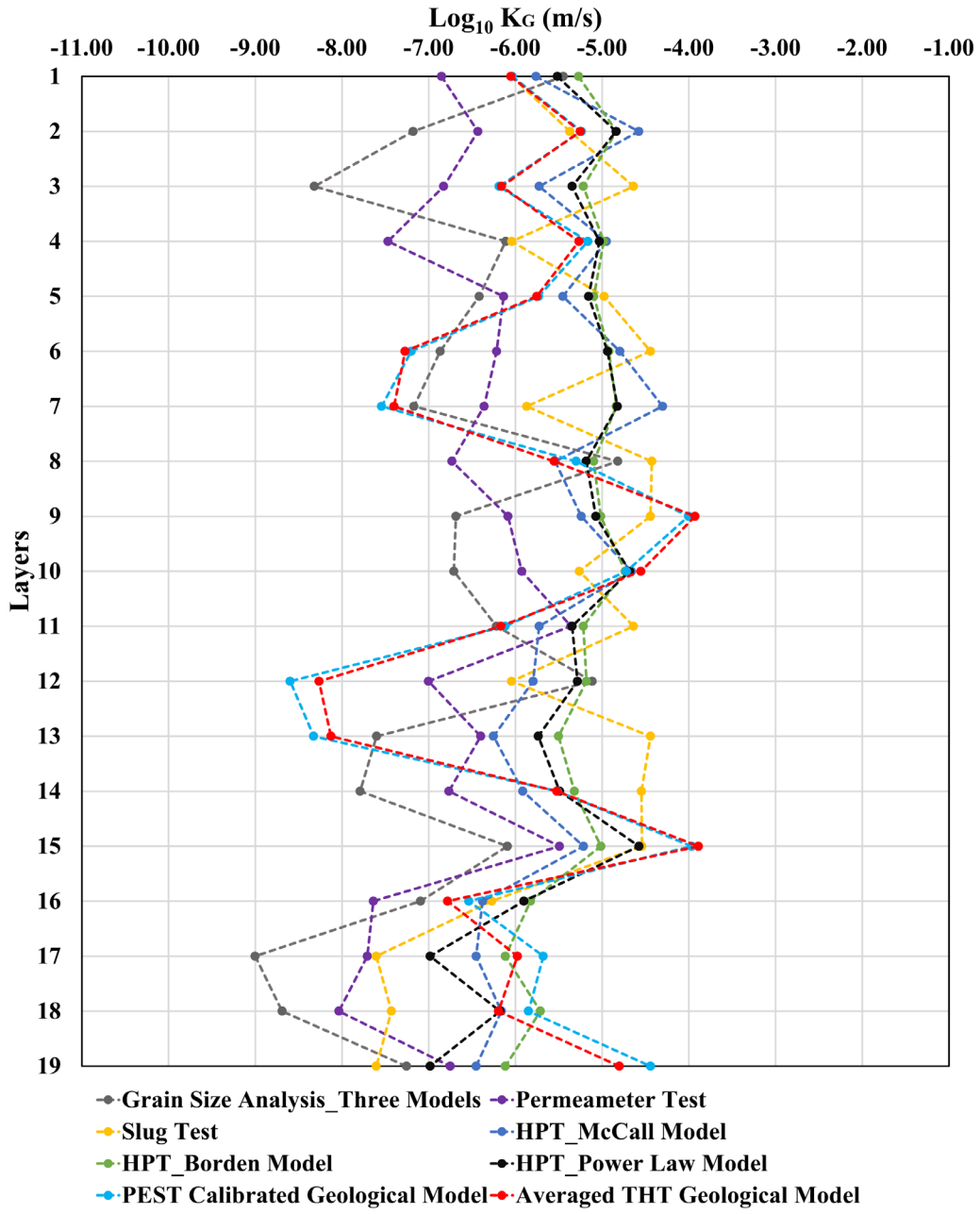


Figure 5. K_G estimates from various site characterization approaches for 19 layers of the geological model.

Table 2. Summary of K_G estimates for the 19-layer geological model from various site characterization approaches.

Method	Minimum (m/s)	Maximum (m/s)	Geomean of K (m/s)	Range of $\log_{10}$ K
Grain Size Analyses_Three Models	9.98E-10	1.52E-05	1.34E-07	4.18
Permeameter Tests	9.23E-09	4.36E-06	2.28E-07	2.67
Slug Tests	2.49E-08	3.74E-05	3.11E-06	3.18
HPT_McCall and Christy Model	3.53E-07	4.98E-05	2.85E-06	2.15
HPT_Borden Model	7.67E-07	1.85E-05	5.45E-06	1.38
HPT_Power Law Model	1.04E-07	2.66E-05	3.76E-06	2.41
PEST Calibrated Geological Model	2.53E-09	1.07E-04	1.25E-06	4.63
Averaged THT Geological Model	5.44E-09	1.29E-04	1.14E-06	4.37

Chapter 6. Evaluation and Validation of Hydraulic Conductivity from Various Subsurface Characterization Methods Using a Groundwater Model

6.1 Comparison of *K* distributions

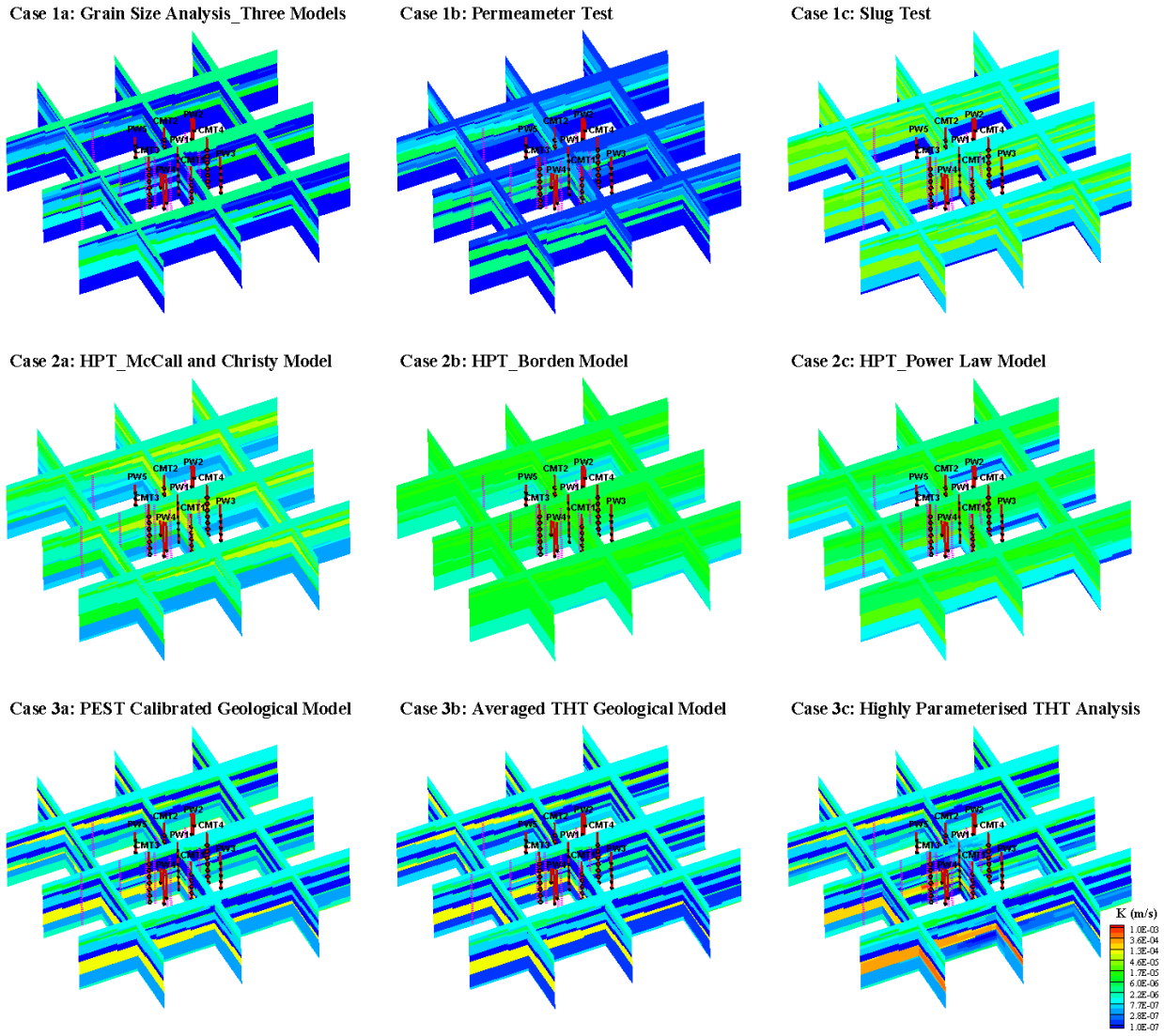


Figure 6. *K* distributions at the NCRS from various site characterization approaches. Positions of CMT and PW wells (red lines) along with their screened intervals (black colour) as well as HPT loggings (dashed pink lines) are shown on each subfigure.

The estimated *K* distributions from traditional, HPT, and inverse modeling are presented as fenced diagrams in Figure 6. As previously noted, the primary characteristic of the site is the alternating aquifer-aquitard system comprised of two high-*K* units separated by a discontinuous low-*K* layer. The lower aquifer is made up of sandy gravel and the higher aquifer is made up of sand to sandy silt. Above and below the two aquifers are low-permeable materials such as silt and clay.

The *K* distributions from HPT using three different models (Cases 2a – 2c) are generally biased towards the higher *K* end. Specifically, results from McCall and Christy (2010)'s model (Case 2a) capture the lower aquitard. However, *K* estimates tend to be larger than those generated from conventional methods. In addition, only the lower aquifer is revealed, while the clay layer 4 and silt layers 6, 7 and 10 located in the upper aquitard, based on Table S4, have higher *K* estimates than the most permeable aquifer unit layer 15, which does not conform to known geology. Results from Borden et al. (2021)'s model (Case 2b) only captures the lower aquitard, while *K* estimates are generally larger, thus everything above the lower aquitard is hard to be distinguished. The *K* distributions from the power law model (Zhao and Illman, 2022b) (Case 2c) only captures the lower aquifer and the lowest aquitard, while the *K* values for units above the lower aquifer are generally less variable and the aquitard layers have generally larger *K* estimates. The less variable values from high-resolution HPT methods are mainly due to the limited range of estimates and the upscaling of centimetre-scale measurements into much larger grid elements (for each layer). This type of averaging is meaningful and could generate results with reasonable fidelity because building forward and inverse groundwater models for flow prediction and remediation decision-making has become one of the most efficient ways for site characterization while most of the grids are not designed in millimetres or, even centimetre scale,

especially for large-scale water-quality investigations to save computational burden (Alexander et al., 2011; Berg & Illman, 2011b, 2013, 2015; Brauchler et al., 2013; Luo et al., 2022; Tong et al., 2021; Zhao & Illman, 2017, 2018).

Examination of three inversion-derived property fields reveals *K* variations more accurately. The *K* values from the PEST calibrated geological model (Case 3a) capture the expected variation from one layer to the next, specifically, the double-aquifer system, and the aquitard in between is clearly implied. Aquitards above and below the aquifer system are also specified, however, the *K* estimate for the upper aquitard is relatively larger than it generated from permeameter test, which is because of the sparse monitoring data at the uppermost model domain. The unreasonably large 95% confidence intervals, according to Zhao and Illman (2018), could attribute to the merging of layers and fixing the layer geometry during model calibration. The averaged THT geological model (Case 3b) has a similar *K* distribution compared to the PEST calibrated geological model (as shown in Figure 5), and the *K* estimate for the lowest aquifer agrees more with those obtained by traditional methods (i.e., grain size analyses and permeameter). The highly parameterized THT analysis (Case 3c) yields a similar property field, while the estimated *K* for the lower aquifer layer 15 is higher compared with Cases 3a and 3b.

6.2 Results from model validation

The performance of each *K* distribution obtained by various methods is then evaluated by predicting independent pumping tests that are not used for model calibration using HGS (Aquanty, 2019) with the grid structures described earlier. As previously noted, a total of 15 pumping tests (PW1-1, PW1-3, PW1-4, PW1-5, PW1-6, PW1-7, PW2-3, PW3-1, PW3-3, PW3-4, PW4-3, PW5-1, PW5-3, PW5-4, and PW5-5) were conducted at the NCRS. Of those 15 tests, eight tests were utilized for case 3 model calibrations (PW1-1. PW1-4, PW1-6, PW1-7, PW2-3,

PW3-3, PW4-3, and PW5-3) by Zhao & Illman (2018). Therefore, for this study, seven tests not used in model calibration by Zhao and Illman (2018) are chosen for model validation.

Since most of the conventional (i.e., permeameter test and grain size analysis) and all HPT methods cannot provide S_s estimates, steady-state simulation is the first metric to analyze various approaches. Only late-time pressure heads from ports that reach steady or quasi-steady state are chosen, which results in 153 head data. To better evaluate the correspondence between the simulated and observed drawdown values, quantitative analyses are first performed by comparing the coefficient of determination (R^2), mean absolute error (L_1) and mean square error (L_2), which are provided as:

$$R^2 = \left\{ \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(\hat{X}_i - \overline{\hat{X}_\iota}\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \times \frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_i - \overline{\hat{X}_\iota}\right)^2}} \right\}^2 \tag{15}$$

$$L_1 = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - \hat{X}_i\right| \tag{16}$$

$$L_2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \hat{X}_i\right)^2 \tag{17}$$

where n is the total number of data, i indicates the data number, X_i is the simulated drawdown, $\hat{X}_i$ is the observed drawdown, $\overline{X}$ is the mean of simulated drawdown, $\overline{\hat{X}_\iota}$ is the mean of observed drawdown.

Statistics calculated from each method through seven simulations are summarized in Tables S8 to S10 (in the Supplementary Information section). Cells in Tables S8 to S10 are colour-coded to enhance the comparison, specifically, the green colour represents the maximum value of R^2, the minimum value is assigned with red colour and the 60-percentile value has a

colour yellow. While for the L_1 and L_2 norms, the red colour is assigned the maximum value, and the minimum value is represented by green. The arithmetic mean of R^2, L_1 and L_2 norms is calculated to rank various approaches. Examination of Tables S8 to S10 reveals that the highly parameterized THT analysis (Case 3c) performs the best yielding the smallest disagreement between simulated and measured drawdowns and the highest R^2, followed by the averaged THT geological model (Case 3b) and PEST calibrated geological model (Case 3a) methods. In addition, the discrepancy between simulated and observed drawdown values is smallest for the power law model (Zhao and Illman, 2022b) among the three HPT formulae. Three conventional methods, especially the grain size analysis (Case 1a) and permeameter test (Case 1b) rank at the lower end. The simulation results are also assessed by plotting scatterplots, as shown in Figure 7. To evaluate the performance, a linear model is fitted including the slope, intercept and R^2, a 45° line is also included in each plot to indicate a perfect match. The slope and intercept details calculated from each method through seven simulations are summarized in Table S11 for interested readers.

To enhance our comparison and evaluation, transient simulations are also performed. Since most of the selected traditional methods and HPT methods cannot yield S_s estimates, the heterogeneous S_s estimated from case 3a (PEST calibrated geological model) are assigned to case 1a to case 2c. Three points are selected from the early, intermediate, and late time of each drawdown curve, which results in a total of 388 head data. It is worth noting that less drawdowns are selected from PW3-1 and PW5-1 because significant Noordbergum effects are observed from majority of the monitoring ports and the two pumping tests were only monitored through CMT wells. The Noordbergum effect is also called poroelastic effect which will produce drawdowns that deviate substantially from those predicted from traditional groundwater theories and may

exhibit itself as reverse water level fluctuations (Verruijt, 1969; Rodrigues, 1983; Berg and Illman, 2015). However, based on Berg et al. (2011), the late time drawdowns after the Noordbergum effects being fully dissipated can be utilized for transient analysis. As a results, only late times are selected from those particular monitoring ports.

Quantitative and qualitative analyses are conducted, the R^2, L_1, L_2, slope and intercept of the linear model are summarized in Tables S12 to S15, while corresponding scatterplots are presented in Figure 8. Specifically, the L_1 norms are presented for transient simulations as a 3-D histogram with varying heights as shown in Figure 9, which will be discussed in the next section. Meanwhile, observed and simulated drawdowns from various approaches are plotted in Figure 10 for the simulation of a pumping test at PW1-3 and the rest are presented in Figures S3 to S8 in the Supplementary Information section. Selected transient pressure heads are shown as solid, black circles on the observed drawdown curves, while forward simulation results of various cases are provided with curves of different colors. Specifically, the red straight line represents case 1a, red long dashed line represents case 1b, red short dashed line represents case 1c; case 2a is shown as green straight line, case 2b is shown as green long dashed line, case 3c is shown as green short dashed line; the blue straight line indicates case 3a, blue long dashed line indicates case 3b and the blue short dashed line indicates case 3c.

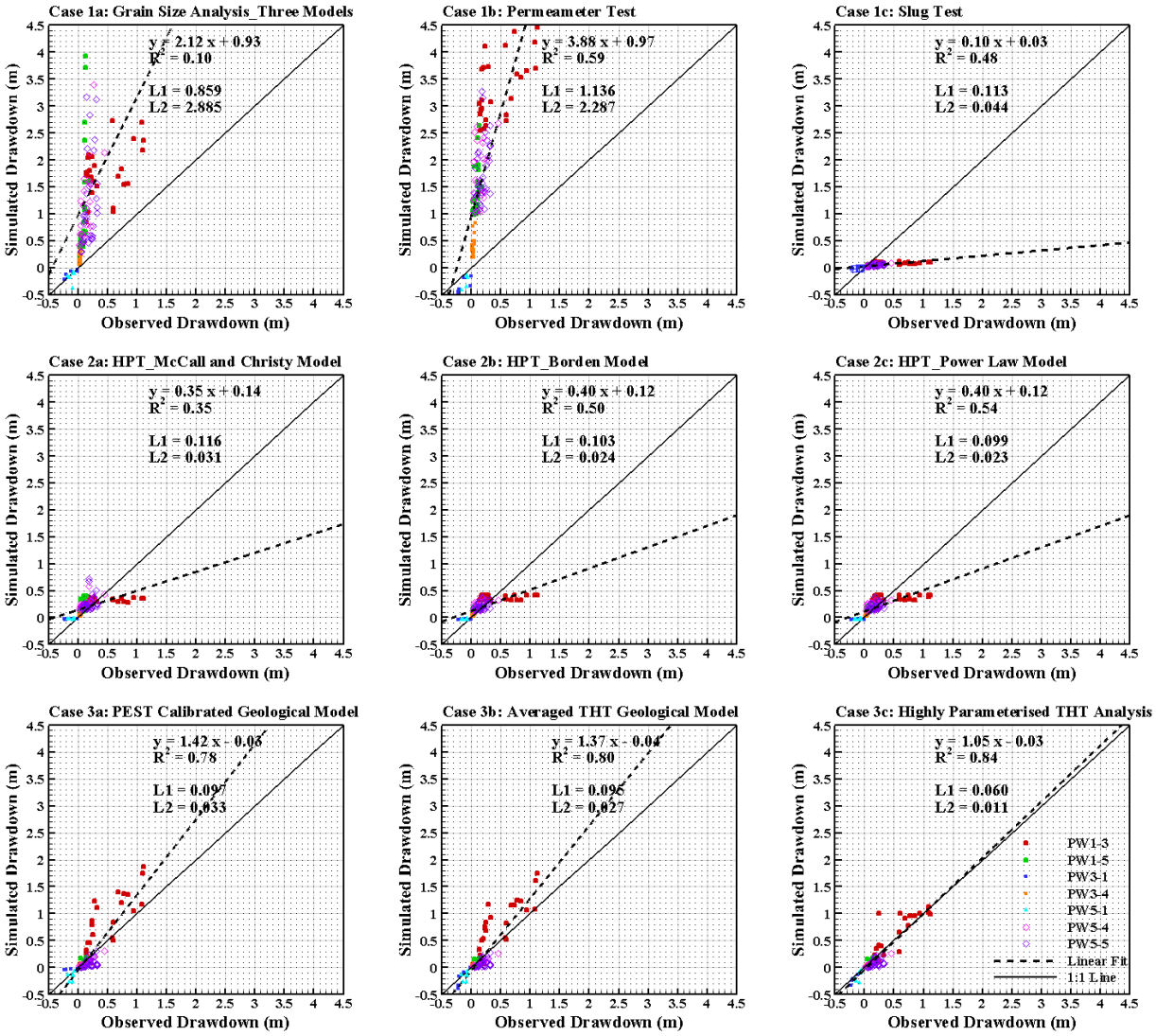


Figure 7. Scatterplots of observed versus simulated drawdowns from various *K* characterization approaches for model validation under steady state conditions.

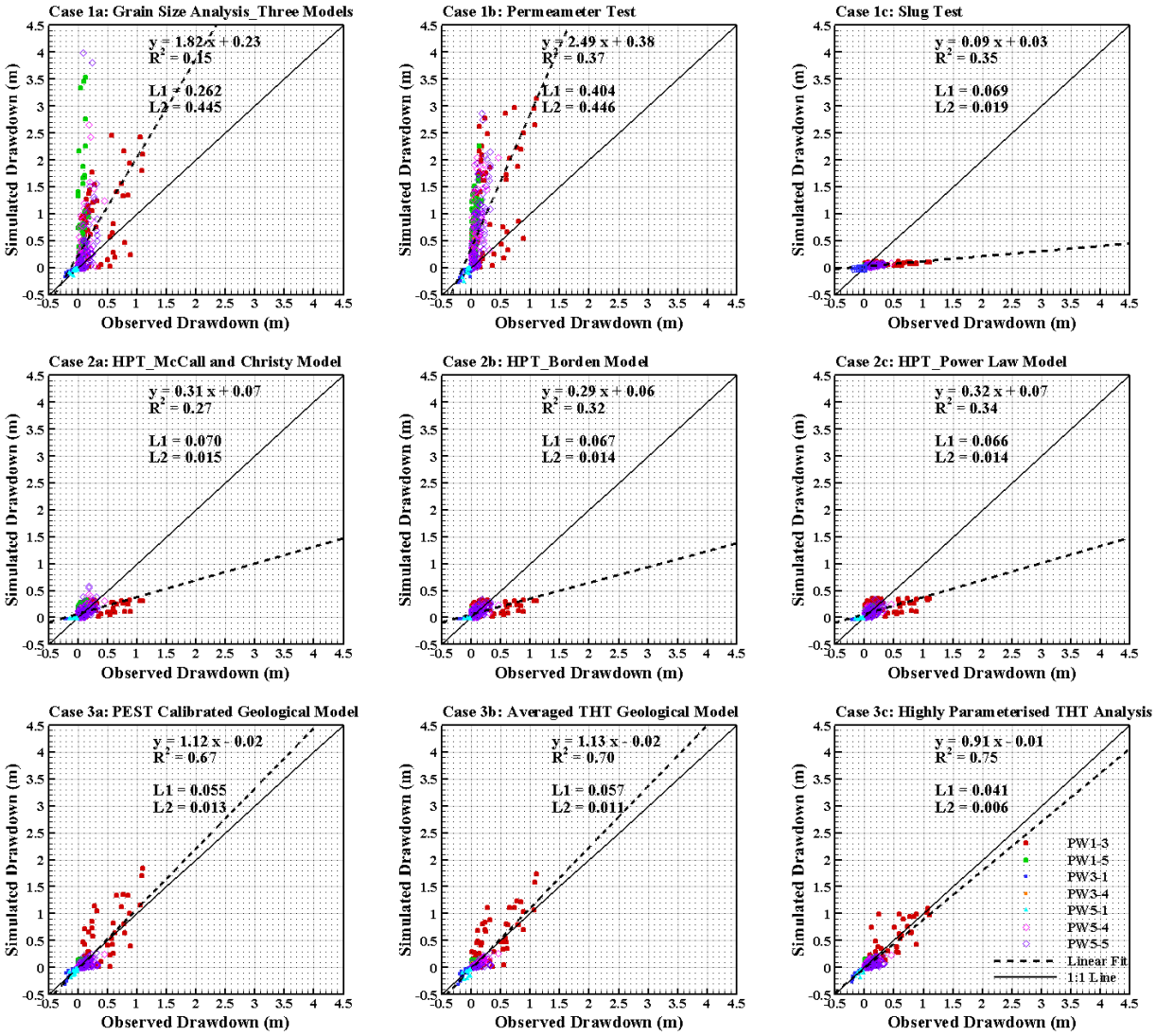


Figure 8. Scatterplots of observed versus simulated drawdowns from various *K* characterization approaches for model validation under transient conditions.

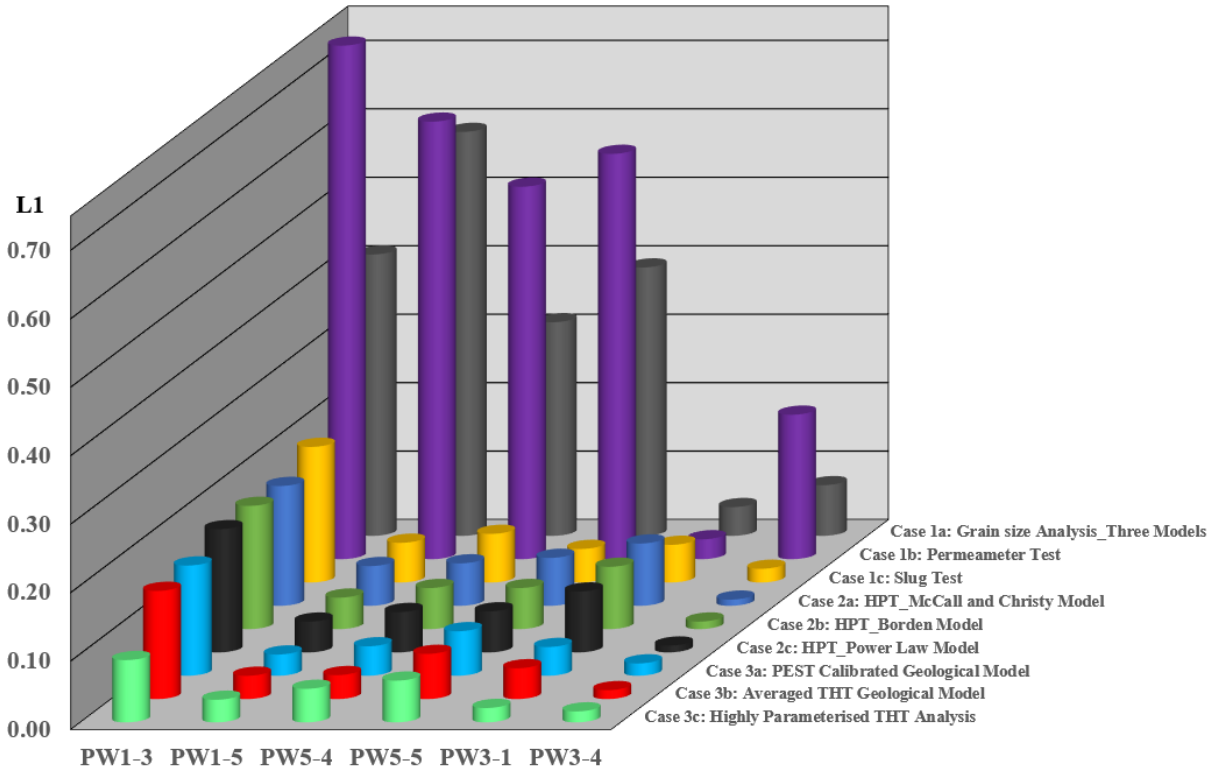
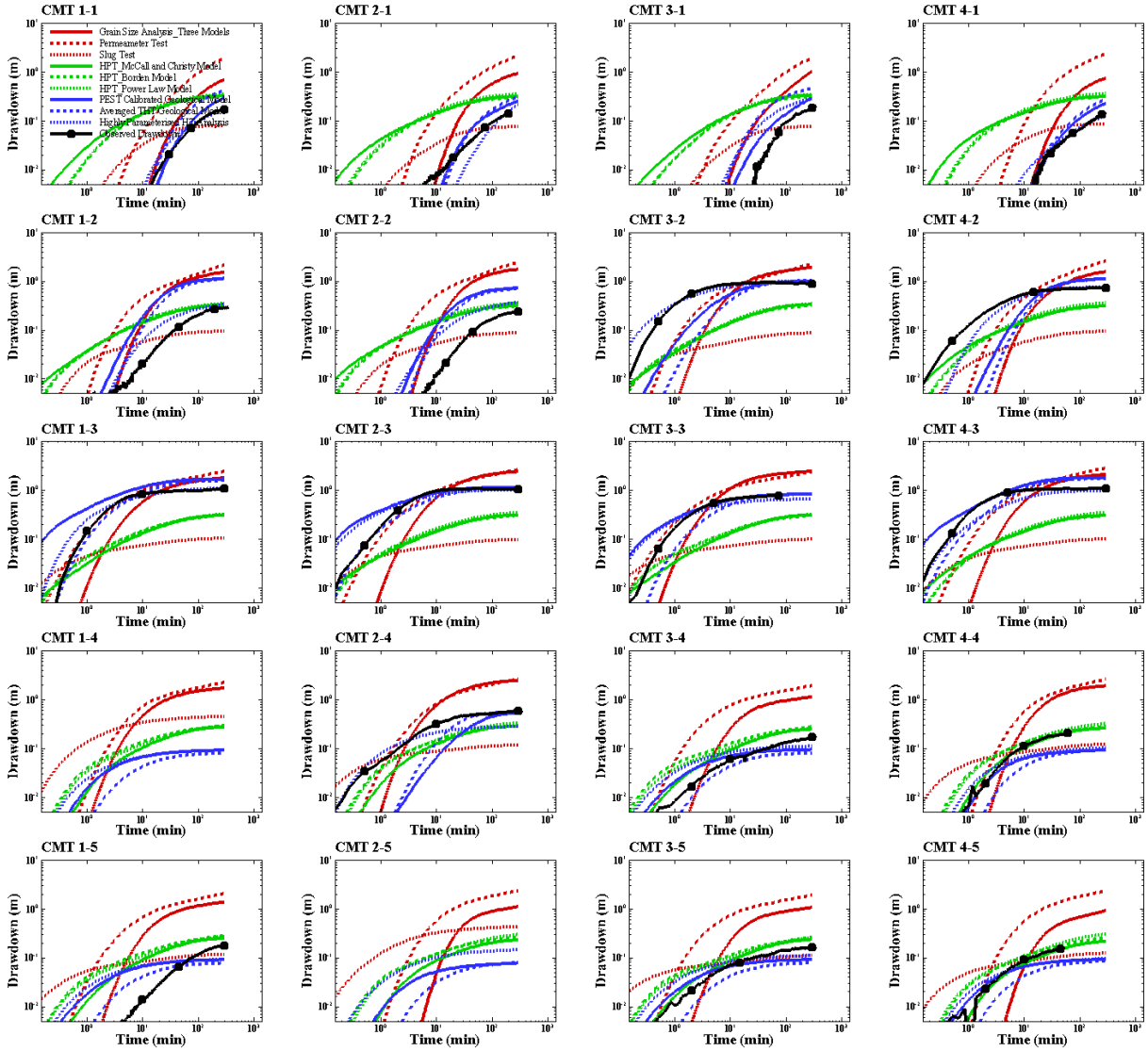


Figure 9. L_1 norms of observed versus simulated drawdowns from various *K* characterization approaches through the forward simulations of seven independent pumping tests under transient conditions.

a)


CMT 1-1
CMT 2-1
CMT 3-1
CMT 4-1
CMT 1-2
CMT 2-2
CMT 3-2
CMT 4-2
CMT 1-3
CMT 2-3
CMT 3-3
CMT 4-3
CMT 1-4
CMT 2-4
CMT 3-4
CMT 4-4
CMT 1-5
CMT 2-5
CMT 3-5
CMT 4-5
Drawdown (m)
Time (min)
Grain Size Analysis_Three Models
Permeameter Test
Slug Test
HPT_McCall and Christy Model
HPT_Borden Model
HPT_Power Law Model

b)

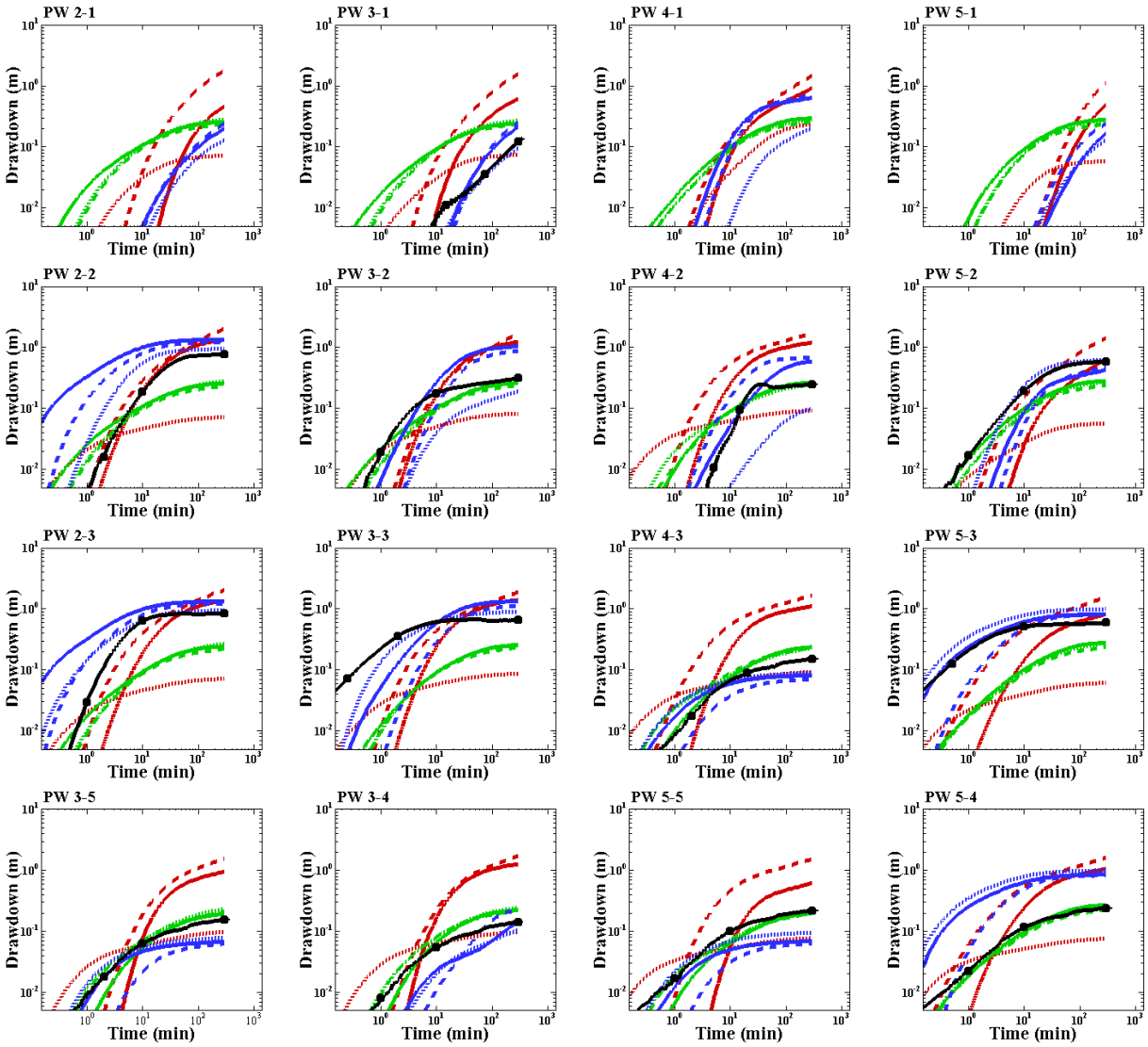


Figure 10. Simulated drawdown curves from forward simulations with various *K* characterization approaches at a) CMT and b) PW wells from PW1-3. Observed drawdown curves are plotted as black curves and the data selected for computation of model performance metrics and scatterplots are solid black circles.

6.3 Discussion

Examination of Figures 7 and 8 reveals that steady-state and transient simulation results are quite comparable. Specifically, both groundwater models built with traditional (Cases 1a - 1c) and HPT (Cases 2a - 2b) *K* estimates yield biased drawdown predictions. The PEST calibration of geological model (Case 3a) and averaged THT Geological Model (Case 3b) both yield a good prediction of drawdowns, while the highly parametrized THT analysis (Case 3c) produces an excellent match.

In terms of traditional methods, both *K* estimates from permeameter tests and grain size analyses overpredict drawdowns, while *K* estimates from slug tests underpredict drawdowns under both steady-state and transient-state conditions. Usually, materials such as silt and clay that have lower *K,* but higher S_s tend to provide higher drawdowns during pumping tests, while high permeable materials such as sand or sand and gravel are likely to yield lower drawdowns. According to Tables 1, 2, and Figure 5, overall, permeameter tests and grain size analyses provide smaller *K* estimates than slug tests. This phenomenon can be explained by the scale effect, which is primarily due to various sample volumes from different methods. Therefore, the method that samples larger volumes, like the slug test, may be impacted by highly permeable zones not discovered by other methods that sample smaller volumes such as the permeameter tests and grain size analyses. Nevertheless, none of the *K* estimations from traditional methods represent the true *K* field and cannot be utilized to accurately simulate groundwater flow under both steady-state and transient conditions. According to Figure 10 and Figures S3 to S8, conventional methods either overpredict or underpredict drawdowns to some extent.

In terms of the grain size analyses, the underestimation of K could be attributed to two potential reasons: (1) the majority of the relationships developed are for high permeable

materials, but for this study, we utilize three models that are appropriate for various soil types at the NCRS, however, models do not perform as expected. Specifically, K estimates from sand layers are smaller than those generated from clay layers; and (2) significant sample loss in the high permeable zone (Alexander et al., 2011). Using a split spoon to sample saturated sand, the probability of losing fines in the recovery process is high. Therefore, those finer fractions cannot even make it to the sieves to be further analyzed.

The overprection of drawdowns from the permeameter test could because this method targets primarily vertical estimates of *K*. Zhao and Illman (2017) through automatic model calibration of seven individual pumping tests (PW1-1, PW1-4, PW1-6, PW2-3, PW3-3, PW4-3, and PW5-3) with steady-state drawdown responses and Zhao & Illman (2018) through effective parameter calibration of eight pumping tests (PW1-1, PW1-4, PW1-6, PW1-7, PW2-3, PW3-3, PW4-3 and PW5-3) with transient pressure heads both found that the vertical hydraulic conductivity in the z-direction (K_z) is about two orders of magnitude smaller than the horizontal hydraulic conductivities in the x- (K_x) and y- (K_y) directions for the NCRS. As a result, it is not suitable to utilize approaches that focus on the vertical parameters to characterize such a highly heterogeneous glaciofluvial site. Even though the repacked samples instead of intact cores are utilized, we believe the nature of original samples is still preserved at some extent otherwise it is useless to use repacked samples for any permeameter analysis. In addition, Klute & Dirksen (1986) demonstrated that *K* of repacked samples estimated in the laboratory can be artificially lower than those from intact samples. Moreover, it is very difficult to recover substantial intact core samples from highly permeable zones (Butler, 2005; Alexander et al., 2011). As a consequence, underprediction of K is possible for permeameter tests conducted with materials from highly permeable intervals.

It is surprising to note that the slug tests consistently underpredict the drawdowns as this method is widely used for water-supply and water-quality investigations (Butler, 1997, 2005; Cardiff et al., 2011). This can be explained by several reasons: (1) the relatively sparse data points available at the site. A total of 43 slug tests were performed and analyzed, and the *K* estimations are then populated into the 19-layer geological model. However, due to the lack of data points, only 11 layers have been assigned with *K*. 2 out of 3 sandy silt layers do not have any data points available and have been assigned with unreasonably large estimates of *K*. As a consequence, a denser slug test well setting is needed to better analyze the performance of slug tests for the characterization of the highly heterogeneous site while this is time-consuming and usually not possible with a limited budget. (2) Hvorslev (1951) method is a slug test solution developed for confined aquifers while the highly heterogeneous aquifer-aquitard system in this study is semi-confined. (3) The adding of "slug" is conducted by injecting a fixed volume of water or by performing short duration pumping tests (Alexander et al., 2011; Xie, 2015). Since the processes are non-instantaneous, error could be introduced into the collected data even though additional translation was taken.

HPT *K* estimates persistently yield biased low predictions of drawdowns under both steady-state and transient conditions. Validation results by using three different models are slightly different and improved. Specifically, based on Figures 7 and 8, R^2 increase, while both L_1 and L_2 norms decrease from case 2a to 2c. As previously mentioned, the power law model (Zhao and Illman, 2022b) is a site-specific relationship developed for the NCRS, therefore, building a site-specific model to interpret HPT data is helpful in terms of site characterization. However, the advancement is not significant, based on Figure 10 and Figures S3 to S8, case 2a to case 2c yield similar head predictions that all underpredict observed drawdowns. According to

Tables 1 and 2, the K_G estimated from three HPT methods are relatively large than those generated from conventional methods. The scale effect cannot explain the behaviour of high-resolution HPT methods. The limited range of *K* estimates presents a substantial challenge to precisely characterize low-permeable materials such as silt and clay. While the loss of Darcian flow into fine-grained materials during injection is the primary cause for the inability of the rapid logging method to provide accurate *K* estimates for these low permeable zones. The loss of Darcian flow in low *K* materials may express itself as bypass of water along the probe and rods (preferential flow paths). As a result, instead of building a site-specific formula, it would be necessary to modify the logging equipment and method to achieve an extended range for both higher and lower *K* subsurface formations (McCall & Christy, 2020).

Inverse modeling approaches with various parametrizations all yield results with great optimism at the NCRS. The calibration and geological model (Case 3a) and averaged THT geological model (Case 3b) both yield a good drawdown match with the measured data, while the highly parameterized THT analysis (Case 3c) with prior geological information yields excellent forward simulation results under both steady and transient conditions. Based on Tables S8 to S10 and Tables S12 to S14 (in the Supplementary Information section), the highly parametrized THT analysis consistently (Case 3c) yields the best R^2, L_1 and L_2 norms under both steady and transient states followed by the calibration of geological model (Case 3a) and averaged THT geological model methods, which indicates that, even though the property *K* fields (refer to Figure 6, Case 3a to 3c) reveal similar overall characteristics, the elemental scale differences in *K* could lead to noticeable differences between simulated drawdowns at various observation intervals. As a result, the highly parametrized inverse modeling approach that can accurately map both interlayer and intralayer heterogeneities with regularization should be

advocated for high-resolution site characterization. Based on Figure 9, it is clear that with the advancement of generations of site characterization methodologies and theories, the ability to accurately map the site heterogeneity is increased significantly.

Ss plays an important role in transient state simulations, however, most of the conventional and HPT methods cannot yield reliable and sufficient *Ss* estimates for numerical modeling. As a result, the importance of including accurate *Ss* for groundwater simulation is analyzed in the study by creating two cases: (1) using an effective *Ss* from Zhao and Illman. (2018) case 1a, which couples HGS (Aquanty, 2019) with the parameter estimation code, PEST (Doherty, 2005), while treats the aquifer-aquitard system to be homogeneous and isotropic; (2) using the heterogeneous *Ss* generated from the PEST calibrated geological model (Case 3a). Transient simulations are performed along seven pumping tests previously used through case 1a to case 2c. L_1 and L_2 are summarized in Table S16 and S17, where yellow color represents the results from heterogeneous *Ss,* and blue color represents the results from homogeneous *Ss*. If the discrepancy between simulated and observed drawdowns from the heterogeneous case is smaller than the homogeneous case, a green check will be marked, otherwise, a red cross will be given. Examination of Table S16 and S17 reveals that, in most cases, accurate and heterogeneous *Ss* yield better simulation results than a homogeneous *Ss*. As a result, the accurate delineation of heterogeneous *Ss* from inverse modeling is remarkable and should be advocated to practitioners and hydrogeologists.

Chapter 7. Summary and Conclusions

The accurate delineation of subsurface heterogeneity in K and S_s is important in building an accurate groundwater model for robust predictions of groundwater flow and remediation decision-making. Traditional methods such as slug and pumping tests are commonly used for water-supply investigations, where it may be sufficient to use a single estimate of averaged K and S_s over a reasonable volume, while such approximations can be of limited value for contaminant transport investigations. Over the last two decades, several DP-based field methods such as DPIL and HPT have been developed to assess small-scale spatial variations of K in heterogeneous unconsolidated formations. The robust machine with advanced theory has made it one of the most efficient approaches for site investigation and characterization compared to conventional methods, while information between each DP location is hard to be interpreted. Inverse modeling such as calibration of geological model and HT has been developed and analyzed significantly in the last decade and a half, numerous promising findings from synthetic, laboratory and field experiments have proved that HT is an effective and robust approach to map site heterogeneity between boreholes. Qualitative and quantitative analyses were conducted by various researchers to compare different approaches but there is no consensus about which method can generate parameter estimations that are representative of a site. In this study, we utilize a 19-layer geological model and a forward groundwater model to evaluate the performance of three generations of site characterization approaches for K including: (1) grain size analyses; (2) permeameter tests; (3) slug tests; (4) HPT with three different models (McCall and Christy, 2020; Borden et al., 2021; and Zhao and Illman, 2022b); (5) calibration of a geological model; and the (6) highly parameterized THT analysis in terms of their ability to

predict drawdowns under both steady-state and transient conditions. This study leads to the following findings and conclusions:

1. Despite the time and effort to conduct 270 grain size analyses, 642 permeameter tests, and 43 slug tests, all of the traditional methods yield biased *K* estimates that lead to poor predictions of drawdowns from pumping tests. The scale effect is observed as the measured *K* increases with the sampled volume. A majority of the grain size analysis solutions were developed based on high permeable materials, which presents a challenge for the applicability in heterogeneous settings. More importantly, the low sample recovery in the high permeable zone will lead to biased low *K* estimates. Permeameter tests tend to yield lower *K* estimates due to the influence of using repacked samples and different lab conditions that could lead to differences from in-situ conditions. In addition, it is very difficult to recover considerable intact core samples from high permeable zones. As a result, underprediction of *K* is possible for permeameter tests conducted with materials from highly permeable intervals. In addition, the current density of CMT wells cannot provide sufficient slug test data for robust forward groundwater simulations, while this method has a relatively larger sample volume than the grain size analysis and permeameter test, thus, a higher chance to include high permeable zones that are not seen by point-scale approaches, thus, lead to relatively higher estimates of *K*. In addition, considerable care must be taken when adding/removing the "slug". Moreover, when selecting formulas, the assumptions and inherent mechanisms must be considered very carefully. Nevertheless, none of the traditional methods can provide adequate information between boreholes without utilizing additional geostatistical interpolation tools such as kriging.

2. HPT is one of the most novel DP approaches, which has the merit of providing substantial measurements of vertical *K* to investigate spatial variation efficiently compared to conventional methods. The interpretation from pressure and EC logs between each survey can be employed for site conceptualization. Building site-specific models based on HPT data is helpful, however, *K* estimations from the HPT method are primarily constraint by the limited range of estimates, which presents a considerable challenge to characterize low permeable materials such as silt and clay. The forward simulation of centimeter-scale *K* estimates from three HPT relationships generates a biased low prediction of drawdowns. Given HPT's significant advantage in the geohydrologic characterization of unconsolidated deposits, it is necessary to advance the logging apparatus and theories to attain an extended range of estimates for both higher and lower *K* geological media.
3. Majority of the traditional and HPT methods cannot provide precise and enough *Ss* for numerical modeling while including accurate and heterogeneous *Ss* is crucial in terms of groundwater simulations. As a consequence, inversion approaches that can delineate accurate *Ss* with various model parametrization should be advocated.
4. Inverse modeling approaches such as calibration of geological model yield good validation results, while the highly parameterized THT analysis consistently yields excellent drawdown predictions under both steady-state and transient conditions. The highly parameterized estimated *K* field revealed the most salient features of interlayer and intralayer heterogeneity. The inclusion of accurate geological zonation and long-term pumping responses from both aquifer and aquitard layers makes HT a clear choice for accurate site characterization compared to traditional and DP methods. The precise

parametrized tomogram is representative of local hydrostratigraphic features. While the accurate prediction of drawdowns from pumping tests is promising, further studies are needed to see whether these K distributions are useful for contaminant transport predictions.

Based on the study results, we recommend the following workflow when characterizing an heterogeneous site which may have plume contamination:

1. Conduct various HPT tests. The generated HPT pressure and EC logs are helpful for site conceptualization. Specifically, a fuzzy clustering analysis of HPT pressure logs has been conducted by Zhao and Illman (2022a), the layer boundaries and thicknesses information was used to build geological models and has shown promising performance in groundwater modeling. In addition, analyzing cross-sectional view of pressure and EC logs is useful in detecting contaminant and its migration path (McCall & Christy, 2020).
2. Install and develop a number of pumping and monitoring wells with screen ports isolated by packers. Then perform multiple pumping/injection tests with longer durations to stress both the aquifer and aquitard units. If contaminants are detected in step 1, considerable care must be taken when pumping water out.
3. Calibrate geological model and HT. The generated geological model from step 1 should calibrate with collected pressure heads, and the inversion of HT then starts with the heterogeneous K and Ss values generated from the calibration of geological model. The inversion with desired degree of parametrization should be conducted by calibrating all pressure heads simultaneously with the code created by Xiang et al. (2009). And the generated property tomograms should be a reasonable representation of the subsurface geological media, which captures the majority of the inter- and intra-layer heterogeneity.

References

Aquanty, 2019. HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport.

ARANZ Geo. Limited., 2015. Leapfrog Hydro 2.2.3. 3D Geological Modeling Software.

Alexander M. (2009). Evaluation of traditional hydrogeologic characterization approaches in a highly heterogeneous glaciofluvial aquifer/aquitard system [M.Sc. thesis]. *University of Waterloo*.

Alexander, M., Berg, S. J., & Illman, W. A. (2011). Field study of hydrogeologic characterization methods in a heterogeneous aquifer. *Ground Water*, 49(3), 365–382. https://doi.org/10.1111/j.1745-6584.2010.00729.x

Barr, D. W. (2001). Coefficient of permeability determined by measurable parameters. *Ground Water*, 39(3), 356–361. https://doi.org/10.1111/j.1745-6584.2001.tb02318.x

Beckie, R., & Harvey, C. F. (2002). What does a slug test measure: an investigation of instrument response and the effects of heterogeneity. *Water Resources Research*, 38(12), 26-1-26–14. https://doi.org/10.1029/2001wr001072

Berg, S. J., Hsieh, P. A., & Illman, W. A. (2011). Estimating hydraulic parameters when poroelastic effects are significant. *Ground Water*, 49(6), 815–829. https://doi.org/10.1111/j.1745-6584.2010.00781.x

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

Berg, S. J., & 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., & 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., & 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

Berg, S. J., Illman, W. A., & Mok, C. M. W. (2015). Joint estimation of hydraulic and poroelastic parameters from a pumping test. *Ground Water*, 53(5), 759–770. https://doi.org/10.1111/gwat.12271

Bohling, G. C., & Butler, J. J. (2010). Inherent limitations of hydraulic tomography. *Ground Water*, 48(6), 809–824. https://doi.org/10.1111/j.1745-6584.2010.00757.x

Bohling, G. C., Zhan, X., Butler, J. J., & Zheng, L. (2002). Steady shape analysis of tomographic pumping tests for characterization of aquifer heterogeneities. *Water Resources Research*, 38(12), 60-1-60–15. https://doi.org/10.1029/2001wr001176

Borden, R. C., Cha, K. Y., & Liu, G. (2021). A physically based approach for estimating hydraulic conductivity from HPT pressure and flowrate. *Ground Water*, 59(2), 266–272. https://doi.org/10.1111/gwat.13039

Brauchler, R., Hu, R., Hu, L., Jiménez, S., Bayer, P., Dietrich, P., & Ptak, T. (2013). Rapid field application of hydraulic tomography for resolving aquifer heterogeneity in unconsolidated sediments. *Water Resources Research*, 49(4), 2013–2024. https://doi.org/10.1002/wrcr.20181

Butler, J. J. (2002). A simple correction for slug tests in small-diameter wells. *Ground Water*, 40(3), 303–308. https://doi.org/10.1111/J.1745-6584.2002.TB02658.X

Butler, J. J. (2005). Hydrogeological methods for estimation of spatial variations in hydraulic conductivity. In S. S. Rubin Yoram and Hubbard (Ed.), *Hydrogeophysics*, 23-58. Springer Netherlands. https://doi.org/10.1007/1-4020-3102-5_2

Butler, J.J. Jr. (1997). The design, performance, and analysis of slug tests. Boca Raton, FL: Lewis Publishers.

Butler, J. J., Dietrich, P., Wittig, V., & Christy, T. (2007). Characterizing hydraulic conductivity with the direct-push permeameter. *Ground Water*, 45(4), 409–419. https://doi.org/10.1111/j.1745-6584.2007.00300.x

Butler, J. J., Jr, Garnett, E. J., & Healey, J. M. (2003). Analysis of slug tests in formations of high hydraulic conductivity. *Ground Water*, 41(5), 620–630. https://doi.org/10.1111/j.1745-6584.2003.tb02400.x

Cardiff, M., Barrash, W., Thoma, M., & Malama, B. (2011). Information content of slug tests for estimating hydraulic properties in realistic, high-conductivity aquifer scenarios. *Journal of Hydrology*, 403(1–2), 66–82. https://doi.org/10.1016/j.jhydrol.2011.03.044

Chapuis, R. P., Dallaire, V., Marcotte, D., Chouteau, M., Acevedo, N., & Gagnon, F. (2005). Evaluating the hydraulic conductivity at three different scales within an unconfined sand aquifer at Lachenaie, Quebec. *Canadian Geotechnical Journal*, 42(4), 1212–1220. https://doi.org/10.1139/t05-045

Cheong, J. Y., Hamm, S. Y., Kim, H. S., Ko, E. J., Yang, K., & Lee, J. H. (2008). Estimating hydraulic conductivity using grain-size analyses, aquifer tests, and numerical modeling in a riverside alluvial system in South Korea. *Hydrogeology Journal*, 16(6), 1129–1143. https://doi.org/10.1007/s10040-008-0303-4

Cooper, H. H., and Jacob, C. E. (1946), A generalized graphical method for evaluating formation constants and summarizing well-field history, Eos Trans. AGU, 27(4), 526– 534. doi:10.1029/TR027i004p00526.

de Marsily, G., Delay, F., Gonçalvès, J., Renard, P., Teles, V., & Violette, S. (2005). Dealing with spatial heterogeneity. *Hydrogeology Journal*, 13(1), 161-183. https://doi.org/10.1007/s10040-004-0432-3

Devlin, J. F. (2015). HydrogeoSieveXL: un outil développé sous Excel pour estimer la conductivité hydraulique à parrit de l'analyse de la granulométrie. *Hydrogeology Journal*, 23(4), 837–844. https://doi.org/10.1007/s10040-015-1255-0

Dietrich, P., Butler, J. J., & Faiß, K. (2008). A rapid method for hydraulic profiling in unconsolidated formations. *Ground Water*, 46(2), 323–328. https://doi.org/10.1111/j.1745-6584.2007.00377.x

Doherty, J., 2005. PEST: Model-Independent Parameter Estimation User Manual. Watermark Numer. Comput, Brisbane, Australia

Freeze, R. A., & Cherry, J. A. (1977). Groundwater. Prentice-Hall.

Freyberg, D. L. (1986). A natural gradient experiment on solute transport in a sand aquifer: 2. Spatial moments and the advection and dispersion of nonreactive tracers. *Water Resources Research*, 22(13), 2031–2046. https://doi.org/10.1029/WR022i013p02031

Geoprobe. 2006. Geoprobe® Hydraulic Profiling Tool (HPT) System: Standard Operating Procedure, Technical Bulletin No. MK3137 (Revised 2015). Salina, Kansas: Kejr Inc./Geoprobe Systems.

Geoprobe. (2015). Geoprobe ® Hydraulic Profiling Tool (HPT) System Standard Operating Procedure.

Hazen, A. (1911). Discussion: Dams on sand foundations. Transactions, *American Society of Civil Engineers*, 73(11).

Hess, K. M., Wolf, S. H., & Celia, M. A. (1992). Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 3. Hydraulic conductivity variability and calculated macrodispersivities. *Water Resources Research*, 28(8), 2011–2027. https://doi.org/10.1029/92WR00668

Hinsby, K., Bjerg, P. L., Andersen, L. J., Skov, B., & Clausen, E. v. (1992). A mini slug test method for determination of a local hydraulic conductivity of an unconfined sandy aquifer. *Journal of Hydrology*, 136(1–4), 87–106. https://doi.org/10.1016/0022-1694(92)90006-H

Hussain, F., & Nabi, G. (2016). Empirical formulae evaluation for hydraulic conductivity determination based on grain size analysis. *Pyrex Journal of Research in Environmental Studies*, 3(3). http://www.pyrexjournals.org/pjres

Hvorslev, M.J. 1951.Time Lag and Soil Permeability in Ground-Water Observations, Bull. No. 36. Vicksburg, Mississippi: Waterways Experiment Station, Corps of Engineers, U.S.Army, 1–50.

Illman, W. A., & Neuman, S. P. (2003). Steady-state analysis of cross-hole pneumatic injection tests in unsaturated fractured tuff. *Journal of Hydrology*, 281(1–2), 36–54. https://doi.org/10.1016/S0022-1694(03)00199-9

Illman, W. A., & Tartakovsky, D. M. (2006). Asymptotic analysis of cross-hole hydraulic tests in fractured granite. *Ground Water*, 44(4), 555–563. https://doi.org/10.1111/j.1745-6584.2006.00201.x

Ishaku, J. M., Gadzama, E. W., & Kaigama, U. (2011). Evaluation of empirical formulae for the determination of hydraulic conductivity based on grain-size analysis. *Journal of Geology and Mining Research*, 3(4).

Jiang, L., Sun, R., Xiao, W., Liang, X., & Jim Yeh, T. C. (2022). Spatial correlation analysis between hydraulic conductivity and specific storage in a heterogeneous sandbox by hydraulic tomography. *Journal of Hydrology*, 610. https://doi.org/10.1016/j.jhydrol.2022.127921

Jim Yeh, T. -C, Mas-Pla, J., Williams, T. M., & McCarthy, J. F. (1995). Observation and Three-Dimensional Simulation of Chloride Plumes in a Sandy Aquifer Under Forced-Gradient

Conditions. *Water Resources Research*, 31(9), 2141–2157. https://doi.org/10.1029/95WR01947

Jones, J. P., Sudicky, E. A., & McLaren, R. G. (2008). Application of a fully-integrated surface-subsurface flow model at the watershed-scale: A case study. *Water Resources Research*, 44(3). https://doi.org/10.1029/2006WR005603

Karrow, P. F. (1979). Quaternary geology of the University of Waterloo campus. Department of Earth Sciences, *University of Waterloo*, Waterloo, ON.

Karrow, P. F. (1993). Quaternary geology, Stratford-Conestogo area. Ontario Ministry of Northern Development and Mines, 283.

Klute, A., & Dirksen, C. (1986). Hydraulic conductivity and diffusivity: laboratory methods. Methods of Soil Analysis, 687-734. John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.2136/sssabookser5.1.2ed.c28

Kozeny, J. (1953). Das wasser im boden. grundwasserbewegung. *Hydraulik*, 380-445. Springer.

Krüger, E. (1918). Die grundwasserbewegung. *Internationale Mitteilungen Für Bodenkunde*, 8(5–6), 105–122.

Krumbein, W. C., & Monk, G. D. (1943). Permeability as a function of the size parameters of unconsolidated sand. *Transactions of the AIME*, 151(01), 153–163.

Leven, C., & Dietrich, P. (2006). What information can we get from pumping tests?-comparing pumping test configurations using sensitivity coefficients. *Journal of Hydrology*, 319(1–4), 199–215. https://doi.org/10.1016/J.JHYDROL.2005.06.030

Liu, F., Yeh, T. C. J., Song, X., Wang, Y. L., Wen, J. C., Hao, Y., & Wang, W. (2021). Temporal sampling and role of flux measurements for subsurface heterogeneous characterization in groundwater basins using hydraulic tomography. *Hydrological Processes*, 35(8). https://doi.org/10.1002/hyp.14299

Liu, G., Borden, R. C., & Butler, J. J. (2019). Simulation assessment of direct push injection logging for high-resolution aquifer characterization. *Ground Water*, 57(4), 562–574. https://doi.org/10.1111/gwat.12826

Liu, G., Butler, J. J., Reboulet, E., & Knobbe, S. (2012). Hydraulic conductivity profiling with direct push methods. *Ground Water*, 17(1), 19–29. https://doi.org/10.1007/s00767-011-0182-9

Liu, X., Illman, W. A., Craig, A. J., Zhu, J., & Yeh, T. C. J. (2007). Laboratory sandbox validation of transient hydraulic tomography. *Water Resources Research*, 43(5). https://doi.org/10.1029/2006WR005144

Lowry, W., Mason, N., Chipman, V., Kisiel, K., & Stockton, J. (1999). In-situ permeability measurements with direct push techniques: phase II topical report. *United States*. https://doi.org/10.2172/773872

Luo, N., Illman, W. A., & Zha, Y. (2022). Large-scale three-dimensional hydraulic tomography analyses of long-term municipal wellfield operations. *Journal of Hydrology*, 610. https://doi.org/10.1016/j.jhydrol.2022.127911

Luo, N., Zhao, Z., Illman, W. A., & Berg, S. J. (2017). Comparative study of transient hydraulic tomography with varying parameterizations and zonations: Laboratory sandbox

investigation. *Journal of Hydrology*, 554, 758–779. https://doi.org/10.1016/j.jhydrol.2017.09.045

Martin, P. J., & Frind, E. O. (1998). Modeling a complex multi-aquifer system: the waterloo moraine. *Ground Water*, 36(4), 679–690. https://doi.org/10.1111/j.1745-6584.1998.tb02843.x

Martinez-Landa, L., & Carrera, J. (2005). An analysis of hydraulic conductivity scale effects in granite (full-scale engineered barrier experiment (FEBEX), Grimsel, Switzerland). *Water Resources Research*, 41(3), 1–13. https://doi.org/10.1029/2004WR003458

McCall, Wesley and Thomas Christy, 2010. Development of a Hydraulic Conductivity-Estimate for the Hydraulic Profiling Tool (HPT) Abstract and Presentation, The 2010 North American Environmental Field Conference& Exposition. The Nielsen Environmental Field School, Las Cruces, NM. January.

McCall, W., & Christy, T. M. (2020). The hydraulic profiling tool for hydrogeologic investigation of unconsolidated formations. *Groundwater Monitoring & Remediation*, 40(3), 89–103. https://doi.org/https://doi.org/10.1111/gwmr.12399

McCall, W., Christy, T. M., & Evald, M. K. (2017). Applying the HPT-GWS for hydrostratigraphy, Water Quality and Aquifer Recharge Investigations. *Groundwater Monitoring and Remediation*, 37(1), 78–91. https://doi.org/10.1111/gwmr.12193

Mccall, W., Christy, T. M., Pipp, D., Terkelsen, M., Christensen, A., Weber, K., & Engelsen, P. (2014). Field application of the combined membrane-interface probe and hydraulic profiling tool (MiHpt). *Groundwater Monitoring and Remediation*, 34(2), 85–95. https://doi.org/10.1111/gwmr.12051

Ning, Z., Luo, N., Inaba, K., Nakashima, T., Shimizu, T., & Illman, W. A. (2022). Three-dimensional hydraulic tomography analyses to investigate commingling issues of reproducibility, data density, and geological prior models. *Journal of Hydrology*, 128785. https://doi.org/10.1016/j.jhydrol.2022.128785

Puckett, W. E., Dane, J. H., & Hajek, B. F. (1985). Physical and mineralogical data to determine soil hydraulic properties. *Soil Science Society of America Journal*, 49(4), 831–836. https://doi.org/https://doi.org/10.2136/sssaj1985.03615995004900040008x

Xie, Q. (2015). Slug tests analysis with different analytical models at a highly heterogeneous field site [B.Sc. thesis]. *University of Waterloo.*

Rehfeldt, K. R., Boggs, J. M., & Gelhar, L. W. (1992). Field study of dispersion in a heterogeneous aquifer: geostatistical analysis of hydraulic conductivity. *Water Resources Research*, 28(12), 3309–3324. https://doi.org/10.1029/92WR01758

Robertson, P. K., Sully, J. P., Woeller, D. J., Lunne, T., Powell, J. J. M., & Gillespie, D. G. (1992). Estimating coefficient of consolidation from piezocone tests. *Canadian Geotechnical Journal*, 29. https://doi.org/10.1139/t92-061

Rodrigues, J. D. (1983). The noordbergum effect and characterization of aquitards at the rio maior mining project. *Ground Water*, 21(2), 200–207. https://doi.org/10.1111/j.1745-6584.1983.tb00714.x

Rosas, J., Lopez, O., Missimer, T. M., Coulibaly, K. M., Dehwah, A. H. A., Sesler, K., Lujan, L. R., & Mantilla, D. (2014). Determination of hydraulic conductivity from grain-size distribution for different depositional environments. *Ground Water*, 52(3), 399–413. https://doi.org/10.1111/gwat.12078

Sebol, L. A. (2000). Determination of groundwater age using CFCs in three shallow aquifers in Southern Ontario. *University of Waterloo*.

Song, J., Chen, X., Cheng, C., Wang, D., Lackey, S., & Xu, Z. (2009). Feasibility of grain-size analysis methods for determination of vertical hydraulic conductivity of streambeds. *Journal of Hydrology*, 375(3–4), 428–437. https://doi.org/10.1016/j.jhydrol.2009.06.043

Stienstra, P., & van Deen, J. K. (1994). Field data collection techniques—Unconventional sounding and sampling methods. *Engineering Geology of Quaternary Sediments*, 41–55.

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/WR022i013p02069

Sudicky, E. (1988). Reply. *Water Resources Research*, 24, no.6: 895–896.

Sudicky, E. A., & Huyakorn, P. S. (1991). Contaminant migration in imperfectly known heterogeneous groundwater systems. *Reviews of Geophysics*, 29(S1), 240–253. https://doi.org/10.1002/rog.1991.29.s1.240

Sudicky, E. A., Illman, W. A., Goltz, I. K., Adams, J. J., & McLaren, R. G. (2010). Heterogeneity in hydraulic conductivity and its role on the macroscale transport of a solute plume: From measurements to a practical application of stochastic flow and transport theory. *Water Resources Research*, 46(1). https://doi.org/10.1029/2008wr007558

Sun D, Luo N, Vandenhoff A, Wang C, Zhao Z, Rudolph D L, & Illman W A. (2022). Evaluation of the hydraulic profiling tool (HPT) at a highly heterogeneous field site underlain by glaciofluvial deposits.

Terzaghi, K. (1925). Principles of soil mechanics: Eng. *News-Record*, 95(742), 1068.

Theis, C.V. (1935) The relation between the lowering of piezometric surface and the rate of the duration of discharge of well using groundwater storage. Eos, *Transactions American Geophysical Union*, 16, 519-524. https://doi.org/10.1029/TR016i002p00519

Tong, X., Illman, W. A., Berg, S. J., & Luo, N. (2021). Hydraulic tomography analysis of municipal-well operation data with geology-based groundwater models. *Hydrogeology Journal*, 29(5), 1979–1997. https://doi.org/10.1007/S10040-021-02320-4/FIGURES/10

Verruijt, A. (1969). Elastic storage of aquifers. *Flow through Porous Media*, 1, 331–376.

Vesselinov, V. v., Neuman, S. P., & Illman, W. A. (2001). Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff 2. Equivalent parameters, high-resolution stochastic imaging and scale effects. *Water Resources Research*, 37(12), 3019–3041. https://doi.org/10.1029/2000WR000135

Vienken, T., & Dietrich, P. (2011). Field evaluation of methods for determining hydraulic conductivity from grain size data. *Journal of Hydrology*, 400(1–2), 58–71. https://doi.org/10.1016/j.jhydrol.2011.01.022

Vukovic, M. and Soro, A. (1992) Determination of Hydraulic Conductivity of Porous Media from Grain-Size Composition. *Water Resources Publications*, LLC Highlands Ranch, Clorado.

White, I. (1988). Comment on "A natural gradient experiment on solute transport in a sand aquifer: Spatial variability of hydraulic conductivity and its role in the dispersion process" by E. A. Sudicky. *Water Resources Research*, 24 (6), 892-894. https://doi.org/10.1029/WR024i006p00892

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

Yeh, T. C. J., & 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

Zhao, Z., Berg, S. J., Illman, W. A., & Qi, Y. (2022). Improving predictions of solute transport in a laboratory sandbox aquifer through high-resolution characterization with hydraulic tomography. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2022.128673

Zhao, Z., & Illman, W. A. (2017). On the importance of geological data for three-dimensional steady-state hydraulic tomography analysis at a highly heterogeneous aquifer-aquitard system. *Journal of Hydrology*, 544, 640–657. https://doi.org/10.1016/j.jhydrol.2016.12.004

Zhao, Z., & 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. (2022a). Integrating hydraulic profiling tool pressure logs and hydraulic tomography for improved high-resolution characterization of subsurface heterogeneity. *Journal of Hydrology*, 610. https://doi.org/10.1016/j.jhydrol.2022.127971

Zhao, Z., & Illman, W. A. (2022b). Improved high-resolution characterization of hydraulic conductivity through inverse modeling of HPT profiles and steady-state hydraulic tomography: Field and synthetic studies. *Journal of Hydrology*, 612. https://doi.org/10.1016/j.jhydrol.2022.128124

Zhao, Z., Illman, W. A., & 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

Zhu, J., & 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

Appendix A. Additional Tables

Grain Size Analysis_Three Models											
	layers	# of data points	Model used	Geomean of K (m/s)	Goemean of K (m/min)	Log10 (K) (m/s)	min	lower	median	upper	max
Grain Size Analysis_ Three Models (Total of 270 data points)	1-clay	17	Puckett et al (1985)	4.03E-06	2.42E-04	-5.39	-7.37	-6.16	-5.05	-4.54	-4.36
	2-silt	16	Barr (2001)	8.99E-08	5.39E-06	-7.05	-9.45	-8.72	-6.52	-5.50	-5.14
	3-sand	1	Hazen (1911)	4.80E-09	2.88E-07	-8.32	-8.32	-8.32	-8.32	-8.32	-8.32
	4-clay	8	Puckett et al (1985)	7.73E-07	4.64E-05	-6.11	-8.05	-6.87	-5.87	-4.83	-4.62
	5-sand & silt	0	NA	9.32E-08	5.59E-06	-7.03	-7.03	-7.03	-7.03	-7.03	-7.03
	6-sandy silt	11	Barr (2001)	6.49E-08	3.89E-06	-7.19	-9.61	-7.86	-7.15	-6.13	-5.21
	7-silt	58	Barr (2001)	6.03E-08	3.62E-06	-7.22	-9.63	-8.06	-7.31	-6.23	-5.38
	8-clay	4	Puckett et al (1985)	1.52E-05	9.09E-04	-4.82	-5.22	-5.15	-4.76	-4.51	-4.36
	9-sandy silt	3	Barr (2001)	2.06E-07	1.24E-05	-6.69	-8.42	-7.42	-7.14	-4.80	-4.50
	10-silt	7	Barr (2001)	2.42E-07	1.45E-05	-6.62	-8.14	-7.88	-6.35	-5.70	-4.50
	11-sand	18	Hazen (1911)	7.23E-07	4.34E-05	-6.14	-8.85	-7.62	-5.21	-4.92	-4.70
	12-clay	10	Puckett et al (1985)	7.15E-06	4.29E-04	-5.15	-6.12	-5.28	-5.17	-4.72	-4.54
	13-sandy silt	9	Barr (2001)	2.31E-08	1.39E-06	-7.64	-9.70	-8.52	-7.49	-7.28	-4.88
	14-silt	29	Barr (2001)	1.21E-08	7.23E-07	-7.92	-10.51	-9.04	-7.82	-6.98	-4.53
	15-sand and gravel	14	Hazen (1911)	1.64E-06	9.82E-05	-5.79	-10.17	-7.10	-4.78	-3.87	-2.60
	16-clay	41	Puckett et al (1985)	8.18E-08	4.91E-06	-7.09	-9.68	-8.05	-7.06	-6.08	-4.88
	17-clay & silt	8	Puckett et al (1985)	7.38E-09	4.43E-07	-8.13	-10.15	-9.08	-7.82	-7.07	-6.00
	18-clay	12	Puckett et al (1985)	2.04E-09	1.23E-07	-8.69	-10.49	-9.75	-9.51	-6.93	-5.22
	19-clay & silt	4	Puckett et al (1985)	2.05E-06	1.23E-04	-5.69	-6.51	-5.74	-5.44	-5.30	-5.30

Table S1. Populated grain size analysis using three models for the 19-layer geological model.

Permeameter Test										
	layers	# of data points	Geomean of K (m/s)	Goemean of K (m/min)	Log10 (K) (m/s)	min	lower	median	upper	max
Permeameter Test (Total of 642 Data Points)	1-clay	67	1.40E-07	8.42E-06	-6.85	-9.60	-7.50	-6.74	-5.98	-4.30
	2-silt	24	3.70E-07	2.22E-05	-6.43	-8.67	-7.35	-5.99	-5.51	-4.55
	3-sand	7	1.49E-07	8.94E-06	-6.83	-8.74	-7.37	-6.71	-5.89	-5.35
	4-clay	23	3.39E-08	2.03E-06	-7.47	-9.94	-8.43	-7.33	-6.39	-5.70
	5-sand & silt	0	7.31E-07	4.38E-05	-6.14	-6.14	-6.14	-6.14	-6.14	-6.14
	6-sandy silt	21	6.07E-07	7.11E-07	-6.22	-8.71	-6.52	-6.05	-5.67	-5.16
	7-silt	157	4.36E-07	2.62E-05	-6.36	-8.33	-6.82	-6.43	-5.72	-4.52
	8-clay	19	1.85E-07	1.11E-05	-6.73	-9.30	-7.38	-6.80	-5.83	-4.57
	9-sandy silt	13	8.21E-07	4.93E-05	-6.09	-7.88	-7.47	-6.30	-4.67	-4.48
	10-silt	19	1.19E-06	7.12E-05	-5.93	-8.00	-7.07	-6.68	-4.68	-3.70
	11-sand	69	4.36E-06	2.62E-04	-5.36	-9.51	-5.23	-4.95	-4.68	-3.63
	12-clay	20	9.93E-08	5.96E-06	-7.00	-9.43	-7.16	-6.93	-6.67	-6.01
	13-sandy silt	20	3.98E-07	2.39E-05	-6.40	-7.95	-7.08	-6.57	-5.99	-3.67
	14-silt	65	1.71E-07	1.03E-05	-6.77	-8.80	-7.19	-6.96	-6.47	-3.57
	15-sand and gravel	36	3.22E-06	1.93E-04	-5.49	-8.79	-6.88	-5.06	-4.39	-2.33
	16-clay	48	2.30E-08	1.38E-06	-7.64	-9.06	-8.23	-7.65	-7.06	-5.86
	17-clay & silt	10	1.96E-08	1.17E-06	-7.71	-8.70	-7.95	-7.66	-7.33	-7.07
	18-clay	17	9.23E-09	5.54E-07	-8.03	-9.24	-8.58	-7.95	-7.40	-7.06
	19-clay & silt	7	1.78E-07	1.07E-05	-6.75	-7.43	-7.16	-6.60	-6.49	-5.92

Table S2. Populated permeameter test for the 19-layer geological model.

Slug Test										
	layers	# of data points	Geomean of K (m/s)	Goemean of K (m/min)	Log10 (K) (m/s)	min	lower	median	upper	max
Slug Test (Total of 43 Data Points)	1-clay	0	9.05E-07	5.43E-05	-6.04	-6.04	-6.04	-6.04	-6.04	-6.04
	2-silt	2	4.24E-06	2.54E-04	-5.37	-5.72	-5.42	-5.25	-5.12	-5.03
	3-sand	0	2.31E-05	1.38E-03	-4.64	-4.64	-4.64	-4.64	-4.64	-4.64
	4-clay	1	9.10E-07	5.46E-05	-6.04	-6.04	-6.04	-6.04	-6.04	-6.04
	5-sand & silt	0	1.05E-05	6.32E-04	-4.98	-4.98	-4.98	-4.98	-4.98	-4.98
	6-sandy silt	0	3.62E-05	2.17E-03	-4.44	-4.44	-4.44	-4.44	-4.44	-4.44
	7-silt	11	1.35E-06	8.10E-05	-5.87	-6.78	-6.47	-6.02	-5.70	-4.07
	8-clay	1	3.74E-05	2.25E-03	-4.43	-4.43	-4.43	-4.43	-4.43	-4.43
	9-sandy silt	0	3.62E-05	2.17E-03	-4.44	-4.44	-4.44	-4.44	-4.44	-4.44
	10-silt	0	5.46E-06	3.27E-04	-5.26	-5.26	-5.26	-5.26	-5.26	-5.26
	11-sand	5	2.31E-05	1.38E-03	-4.64	-5.06	-4.71	-4.59	-4.52	-4.31
	12-clay	0	9.05E-07	5.43E-05	-6.04	-6.04	-6.04	-6.04	-6.04	-6.04
	13-sandy silt	4	3.62E-05	2.17E-03	-4.44	-4.69	-4.57	-4.52	-4.33	-4.04
	14-silt	2	2.84E-05	1.70E-03	-4.55	-5.11	-4.50	-4.25	-4.10	-3.98
	15-sand and gravel	5	2.88E-05	1.73E-03	-4.54	-6.38	-4.76	-3.98	-3.81	-3.77
	16-clay	7	5.32E-07	3.19E-05	-6.27	-7.70	-7.18	-6.50	-5.18	-4.81
	17-clay & silt	1	2.49E-08	1.49E-06	-7.60	-7.60	-7.60	-7.60	-7.60	-7.60
	18-clay	4	3.71E-08	2.22E-06	-7.43	-7.92	-7.39	-7.30	-7.27	-7.22
	19-clay & silt	0	2.49E-08	1.49E-06	-7.60	-7.60	-7.60	-7.60	-7.60	-7.60

Table S3. Populated slug test for the 19-layer geological model.

HPT_McCall Model										
	layers	# of data points	Geomean of K (m/s)	Goemean of K (m/min)	Log10 (K) (m/s)	min	lower	median	upper	max
HPT_McCall Model (Total of 7660 Data Points)	1-clay	0	1.73E-06	1.04E-04	-5.76	-5.76	-5.76	-5.76	-5.76	-5.76
	2-silt	86	2.65E-05	1.59E-03	-4.58	-6.45	-4.63	-4.22	-4.06	-3.72
	3-sand	0	1.88E-06	1.13E-04	-5.73	-5.73	-5.73	-5.73	-5.73	-5.73
	4-clay	79	1.12E-05	6.72E-04	-4.95	-6.45	-6.45	-4.28	-4.17	-4.07
	5-sand & silt	28	3.53E-06	2.12E-04	-5.45	-6.45	-6.45	-5.25	-4.39	-4.09
	6-sandy silt	353	1.60E-05	9.61E-04	-4.80	-6.45	-5.25	-4.37	-4.13	-3.70
	7-silt	1446	4.98E-05	2.99E-03	-4.30	-6.45	-5.07	-4.41	-4.04	-3.62
	8-clay	254	2.87E-06	1.72E-04	-5.54	-6.45	-6.45	-6.45	-4.28	-3.67
	9-sandy silt	114	5.75E-06	3.45E-04	-5.24	-6.45	-6.45	-4.76	-4.07	-3.63
	10-silt	108	2.18E-05	1.31E-03	-4.66	-6.45	-5.29	-4.19	-3.90	-3.63
	11-sand	1097	1.88E-06	1.13E-04	-5.73	-6.45	-6.45	-6.45	-4.69	-3.65
	12-clay	311	1.60E-06	9.63E-05	-5.79	-6.45	-6.45	-6.45	-5.07	-3.61
	13-sandy silt	205	5.59E-07	3.35E-05	-6.25	-6.45	-6.45	-6.45	-6.45	-4.34
	14-silt	552	1.21E-06	7.28E-05	-5.92	-6.45	-6.45	-6.45	-6.45	-3.58
	15-sand and gravel	1186	6.11E-06	3.66E-04	-5.21	-6.45	-6.45	-6.25	-3.80	-3.58
	16-clay	1470	4.18E-07	2.51E-05	-6.38	-6.45	-6.45	-6.45	-6.45	-3.80
	17-clay & silt	90	3.53E-07	2.12E-05	-6.45	-6.45	-6.45	-6.45	-6.45	-6.45
	18-clay	281	6.89E-07	4.13E-05	-6.16	-6.45	-6.45	-6.45	-6.45	-3.67
	19-clay & silt	0	3.53E-07	2.12E-05	-6.45	-6.45	-6.45	-6.45	-6.45	-6.45

Table S4. Populated HPT method using McCall (2010)'s model for the 19-layer geological model.

HPT_Borden Model										
	layers	# of data points	Geomean of K (m/s)	Goemean of K (m/min)	Log10 (K) (m/s)	min	lower	median	upper	max
HPT_Borden Model (Total of 7660 Data Points)	1-clay	0	5.35E-06	3.21E-04	-5.27	-5.27	-5.27	-5.27	-5.27	-5.27
	2-silt	86	1.46E-05	8.77E-04	-4.84	-5.49	-5.03	-4.79	-4.65	-4.04
	3-sand	0	6.06E-06	3.64E-04	-5.22	-5.22	-5.22	-5.22	-5.22	-5.22
	4-clay	79	1.04E-05	6.25E-04	-4.98	-5.51	-5.23	-4.85	-4.76	-4.64
	5-sand & silt	28	8.02E-06	4.81E-04	-5.10	-5.32	-5.25	-5.16	-4.94	-4.71
	6-sandy silt	353	1.20E-05	7.18E-04	-4.92	-5.55	-5.16	-4.93	-4.76	-4.00
	7-silt	1446	1.46E-05	8.76E-04	-4.84	-5.59	-5.11	-4.95	-4.65	-3.75
	8-clay	254	8.02E-06	4.81E-04	-5.10	-5.62	-5.39	-5.23	-4.87	-3.88
	9-sandy silt	114	9.66E-06	5.80E-04	-5.02	-5.55	-5.33	-5.08	-4.65	-3.76
	10-silt	108	1.85E-05	1.11E-03	-4.73	-5.53	-5.14	-4.81	-4.45	-3.73
	11-sand	1097	6.06E-06	3.64E-04	-5.22	-6.89	-5.54	-5.37	-5.08	-3.86
	12-clay	311	6.60E-06	3.96E-04	-5.18	-6.18	-5.55	-5.31	-5.13	-3.75
	13-sandy silt	205	3.14E-06	1.89E-04	-5.50	-5.81	-5.62	-5.57	-5.44	-4.93
	14-silt	552	4.80E-06	2.88E-04	-5.32	-6.35	-5.61	-5.47	-5.21	-3.60
	15-sand and gravel	1186	9.68E-06	5.81E-04	-5.01	-7.31	-5.60	-5.19	-4.26	-3.57
	16-clay	1470	1.49E-06	8.94E-05	-5.83	-7.95	-6.11	-5.76	-5.58	-4.26
	17-clay & silt	90	7.67E-07	4.60E-05	-6.12	-7.37	-6.16	-6.09	-5.99	-5.85
	18-clay	281	1.94E-06	1.16E-04	-5.71	-7.75	-6.02	-5.75	-5.61	-3.94
	19-clay & silt	0	7.67E-07	4.60E-05	-6.12	-6.12	-6.12	-6.12	-6.12	-6.12

Table S5. Populated HPT method using Borden et al. (2021)'s model for the 19-layer geological model.

HPT_Power Law Model										
	layers	# of data points	Geomean of K (m/s)	Geomean of K (m/min)	Log10 (K) (m/s)	min	lower	median	upper	max
HPT_Power Law Model (Total of 7660 Data Points)	1-clay	0	3.05E-06	1.83E-04	-5.52	-5.52	-5.52	-5.52	-5.52	-5.52
	2-silt	86	1.45E-05	8.69E-04	-4.84	-5.77	-5.10	-4.79	-4.59	-3.72
	3-sand	0	4.51E-06	2.71E-04	-5.35	-5.35	-5.35	-5.35	-5.35	-5.35
	4-clay	79	9.31E-06	5.59E-04	-5.03	-5.76	-5.35	-4.85	-4.74	-4.59
	5-sand & silt	28	6.98E-06	4.19E-04	-5.16	-5.45	-5.36	-5.25	-4.95	-4.64
	6-sandy silt	353	1.15E-05	6.93E-04	-4.94	-5.83	-5.25	-4.94	-4.70	-3.67
	7-silt	1446	1.50E-05	9.02E-04	-4.82	-5.85	-5.22	-4.98	-4.55	-3.32
	8-clay	254	6.56E-06	3.94E-04	-5.18	-5.92	-5.59	-5.37	-4.87	-3.56
	9-sandy silt	114	8.45E-06	5.07E-04	-5.07	-5.82	-5.50	-5.15	-4.58	-3.39
	10-silt	108	2.08E-05	1.25E-03	-4.68	-5.70	-5.25	-4.75	-4.25	-3.36
	11-sand	1097	4.51E-06	2.71E-04	-5.35	-7.72	-5.80	-5.55	-5.13	-3.46
	12-clay	311	5.19E-06	3.12E-04	-5.28	-6.62	-5.77	-5.45	-5.22	-3.28
	13-sandy silt	205	1.84E-06	1.10E-04	-5.74	-6.21	-5.91	-5.84	-5.69	-4.92
	14-silt	552	3.24E-06	1.94E-04	-5.49	-7.51	-5.90	-5.69	-5.32	-3.16
	15-sand and gravel	1186	2.66E-05	1.60E-03	-4.57	-9.10	-5.86	-5.29	-4.00	-3.16
	16-clay	1470	1.25E-06	7.53E-05	-5.90	-9.42	-6.81	-6.14	-5.83	-4.00
	17-clay & silt	90	1.04E-07	6.24E-06	-6.98	-10.07	-7.06	-6.85	-6.74	-6.46
	18-clay	281	6.65E-07	3.99E-05	-6.18	-8.83	-6.77	-6.09	-5.88	-3.54
	19-clay & silt	0	1.04E-07	6.24E-06	-6.98	-6.98	-6.98	-6.98	-6.98	-6.98

Table S6. Populated HPT method using the power law model for the 19-layer geological model.

Highly Parameterised THT Analysis																	
	layers	# of data points	Geomean of K (m/s)	Geomean of K (m/min)	Geomean of Ss (1/m)	Log10 (K) (m/s)	min	lower	median	upper	max	Log10 (Ss) (1/m)	min	lower	median	upper	max
Highly Parameterised THT Analysis (Total of 31713 Elements)	1-clay	4415	8.87E-07	5.32E-05	1.13E-04	-6.05	-6.69	-6.12	-6.04	-5.98	-5.60	-3.95	-4.26	-4.03	-3.98	-3.86	-3.54
	2-silt	1249	5.61E-06	3.36E-04	8.94E-05	-5.25	-6.16	-5.45	-5.23	-5.05	-4.48	-4.05	-4.39	-4.14	-4.02	-4.00	-3.23
	3-sand	593	6.94E-07	4.16E-05	1.02E-04	-6.16	-6.75	-6.21	-6.14	-6.08	-5.51	-3.99	-4.42	-4.02	-4.00	-3.95	-3.70
	4-clay	589	5.41E-06	3.24E-04	1.17E-04	-5.27	-5.84	-5.48	-5.24	-5.08	-4.80	-3.93	-4.36	-3.99	-3.96	-3.87	-3.49
	5-sand & silt	137	1.77E-06	1.06E-04	1.03E-04	-5.75	-6.26	-5.77	-5.73	-5.70	-5.67	-3.99	-4.09	-4.00	-3.99	-3.98	-3.92
	6-sandy silt	722	5.35E-08	3.21E-06	1.06E-04	-7.27	-8.36	-7.41	-7.26	-7.16	-6.40	-3.97	-4.47	-4.06	-4.00	-3.94	-2.99
	7-silt	3669	3.99E-08	2.39E-06	3.80E-05	-7.40	-9.24	-7.71	-7.52	-7.07	-5.90	-4.42	-5.64	-4.70	-4.38	-4.07	-2.90
	8-clay	791	2.80E-06	1.68E-04	3.61E-05	-5.55	-7.16	-5.80	-5.53	-5.30	-3.94	-4.44	-5.75	-4.88	-4.32	-4.00	-2.93
	9-sandy silt	442	1.18E-04	7.08E-03	2.91E-05	-3.93	-5.43	-4.17	-4.04	-3.66	-2.54	-4.54	-5.51	-4.80	-4.46	-4.22	-3.75
	10-silt	588	2.81E-05	1.68E-03	3.67E-05	-4.55	-5.88	-4.92	-4.75	-4.06	-3.33	-4.44	-5.74	-4.84	-4.29	-4.01	-3.78
	11-sand	2508	6.79E-07	4.08E-05	4.15E-05	-6.17	-8.07	-6.29	-6.16	-6.07	-5.29	-4.38	-5.82	-4.88	-4.07	-3.97	-3.38
	12-clay	1804	5.44E-09	3.26E-07	5.76E-05	-8.26	-10.38	-8.66	-8.60	-8.52	-7.40	-4.24	-5.83	-4.39	-4.01	-3.99	-3.57
	13-sandy silt	693	7.48E-09	4.49E-07	2.19E-05	-8.13	-9.01	-8.38	-8.32	-8.02	-6.47	-4.66	-5.89	-5.29	-4.50	-4.00	-3.92
	14-silt	1850	3.00E-06	1.80E-04	2.07E-05	-5.52	-6.55	-5.76	-5.62	-5.31	-3.42	-4.68	-5.79	-5.15	-4.77	-4.04	-3.88
	15-sand and gravel	2342	1.29E-04	7.72E-03	2.61E-05	-3.89	-5.64	-4.14	-3.83	-3.59	-3.03	-4.58	-5.35	-4.99	-4.60	-4.14	-3.99
	16-clay	5092	1.65E-07	9.91E-06	1.09E-04	-6.78	-8.79	-6.94	-6.62	-6.50	-4.81	-3.96	-5.10	-4.07	-4.00	-3.85	-1.89
	17-clay & silt	1119	1.06E-06	6.35E-05	8.93E-04	-5.98	-7.08	-6.15	-5.95	-5.80	-4.94	-3.05	-4.14	-3.96	-3.62	-3.28	-2.21
	18-clay	2642	6.39E-07	3.84E-05	1.66E-04	-6.19	-8.55	-6.41	-6.02	-5.88	-4.92	-3.78	-4.32	-4.01	-3.94	-3.58	-2.05
	19-clay & silt	468	1.58E-05	9.46E-04	1.64E-04	-4.80	-6.78	-5.02	-4.72	-4.42	-3.76	-3.78	-4.18	-3.98	-3.85	-3.63	-3.19

Table S7. Populated highly parameterized THT analysis (THT calibrated geological model) for the 19-layer geological model.

Approaches	PW1-3	PW1-5	PW5-4	PW5-5	PW3-1	PW3-4	PW5-1	Average	Rank
Case 1a: Grain size Analysis_Three Models	0.17	0.14	0.30	0.03	0.80	0.32	0.01	0.25	**7**
Case 1b: Permeameter Test	0.35	0.18	0.12	0.14	0.56	0.23	0.68	0.32	**5**
Case 1c: Slug Test	0.001	0.38	0.31	0.30	0.94	0.26	0.37	0.37	**4**
Case 2a: HPT_McCall and Christy Model	0.07	0.29	0.38	0.07	0.42	0.31	0.51	0.29	**6**
Case 2b: HPT_Borden Model	0.06	0.21	0.17	0.15	0.40	0.17	0.56	0.25	**7**
Case 2c: HPT_Power Law Model	0.21	0.23	0.22	0.15	0.41	0.18	0.61	0.29	**6**
Case 3a: PEST Calibrated Geological Model	0.68	0.41	0.87	0.11	0.81	0.01	0.08	0.42	**3**
Case 3b: Averaged THT Geological Model	0.68	0.39	0.87	0.10	0.82	0.01	0.11	0.43	**2**
Case 3c: Highly Parameterised THT Analysis	0.74	0.45	0.86	0.20	0.94	0.19	0.37	0.54	**1**

Table S8. R^2 of observed versus simulated drawdowns from seven pumping tests under steady-state condition for model validation.

Approaches	PW1-3	PW1-5	PW5-4	PW5-5	PW3-1	PW3-4	PW5-1	Average	Rank
Case 1a: Grain size Analysis_Three Models	1.314	1.547	1.176	1.733	0.038	0.132	0.075	0.859	**8**
Case 1b: Permeameter Test	2.873	1.178	1.572	1.591	0.201	0.342	0.196	1.136	**9**
Case 1c: Slug Test	0.375	0.046	0.093	0.138	0.024	0.021	0.092	0.113	**6**
Case 2a: HPT_McCall and Christy Model	0.279	0.105	0.109	0.112	0.114	0.013	0.079	0.116	**7**
Case 2b: HPT_Borden Model	0.284	0.063	0.098	0.072	0.112	0.012	0.078	0.103	**5**
Case 2c: HPT_Power Law Model	0.271	0.056	0.091	0.071	0.111	0.013	0.077	0.099	**4**
Case 3a: PEST Calibrated Geological Model	0.341	0.034	0.045	0.104	0.063	0.015	0.080	0.097	**3**
Case 3b: Averaged THT Geological Model	0.306	0.038	0.055	0.112	0.061	0.017	0.077	0.095	**2**
Case 3c: Highly Parameterised THT Analysis	0.139	0.036	0.043	0.104	0.024	0.013	0.064	0.060	**1**

Table S9. L_1 norm of observed versus simulated drawdowns from seven pumping tests under steady-state condition for model validation.

Approaches	PW1-3	PW1-5	PW5-4	PW5-5	PW3-1	PW3-4	PW5-1	Average	Rank
Case 1a: Grain size Analysis_Three Models	1.933	6.792	2.300	9.127	0.002	0.023	0.016	2.885	**9**
Case 1b: Permeameter Test	8.470	1.745	2.733	2.819	0.047	0.145	0.052	2.287	**8**
Case 1c: Slug Test	0.253	0.003	0.016	0.025	0.002	0.001	0.010	0.044	**7**
Case 2a: HPT_McCall and Christy Model	0.127	0.019	0.018	0.025	0.019	0.0003	0.007	0.031	**5**
Case 2b: HPT_Borden Model	0.116	0.006	0.015	0.009	0.018	0.0002	0.007	0.024	**3**
Case 2c: HPT_Power Law Model	0.112	0.005	0.012	0.008	0.017	0.0002	0.007	0.023	**2**
Case 3a: PEST Calibrated Geological Model	0.192	0.002	0.003	0.018	0.008	0.0004	0.010	0.033	**6**
Case 3b: Averaged THT Geological Model	0.147	0.002	0.005	0.020	0.007	0.0005	0.009	0.027	**4**
Case 3c: Highly Parameterised THT Analysis	0.043	0.002	0.004	0.016	0.002	0.0002	0.009	0.011	**1**

Table S10. L_2 norm of observed versus simulated drawdowns from seven pumping tests under steady-state condition for model validation.

Approaches		PW1-3	PW1-5	PW5-4	PW5-5	PW3-1	PW3-4	PW5-1
Case 1a: Grain size Analysis_Three Models	Slope	0.57	21.20	5.79	4.91	1.07	10.14	0.29
	Intercept	1.51	-0.09	0.51	1.01	0.05	-0.30	-0.14
Case 1b: Permeameter Test	Slope	1.02	4.53	1.94	2.62	2.39	28.45	3.17
	Intercept	2.86	0.98	1.44	1.29	0.10	-0.98	0.03
Case 1c: Slug Test	Slope	-0.002	0.17	0.09	0.11	1.25	0.35	0.12
	Intercept	0.09	0.02	0.03	0.03	0.01	0.00	0.00
Case 2a: HPT_McCall and Christy Model	Slope	0.02	1.22	0.65	0.44	0.11	2.56	0.10
	Intercept	0.32	0.08	0.15	0.18	-0.01	-0.08	-0.02
Case 2b: HPT_Borden Model	Slope	0.03	0.33	0.28	0.34	0.16	1.75	0.15
	Intercept	0.36	0.13	0.18	0.16	0.00	-0.04	-0.01
Case 2c: HPT_Power Law Model	Slope	0.06	0.41	0.29	0.31	0.17	1.92	0.17
	Intercept	0.33	0.12	0.17	0.16	0.00	-0.05	-0.01
Case 3a: PEST Calibrated Geological Model	Slope	1.33	0.95	0.72	0.32	1.99	1.45	0.52
	Intercept	0.14	-0.02	-0.001	0.03	0.15	-0.04	-0.12
Case 3b: Averaged THT Geological Model	Slope	1.21	0.82	0.63	0.27	1.98	1.06	0.62
	Intercept	0.15	-0.02	0.00	0.03	0.15	-0.02	-0.11
Case 3c: Highly Parameterised THT Analysis	Slope	1.01	0.95	0.61	0.35	1.29	2.09	1.38
	Intercept	0.04	-0.02	0.01	0.02	0.03	-0.07	-0.02

Table S11. Statistics of the linear model fit from scatterplots of simulated versus observed drawdowns during model validation under steady-state condition.

Approaches	PW1-3	PW1-5	PW5-4	PW5-5	PW3-1	PW3-4	PW5-1	Average	Rank
Case 1a: Grain size Analysis_Three Models	0.41	0.22	0.27	0.12	0.80	0.26	0.45	0.36	**4**
Case 1b: Permeameter Test	0.32	0.34	0.26	0.46	0.64	0.44	0.71	0.45	**2**
Case 1c: Slug Test	0.11	0.25	0.33	0.56	0.59	0.40	0.09	0.33	**6**
Case 2a: HPT_McCall and Christy Model	0.18	0.31	0.37	0.27	0.66	0.40	0.23	0.35	**5**
Case 2b: HPT_Borden Model	0.16	0.26	0.23	0.40	0.59	0.30	0.20	0.30	**8**
Case 2c: HPT_Power Law Model	0.18	0.25	0.27	0.38	0.60	0.32	0.23	0.32	**7**
Case 3a: PEST Calibrated Geological Model	0.65	0.33	0.57	0.25	0.78	0.19	0.33	0.44	**3**
Case 3b: Averaged THT Geological Model	0.69	0.28	0.83	0.25	0.79	0.13	0.17	0.45	**2**
Case 3c: Highly Parameterised THT Analysis	0.75	0.34	0.40	0.35	0.92	0.39	0.40	0.51	**1**

Table S12. R^2 of observed versus simulated drawdowns from seven pumping tests under transient-state condition for model validation.

Approaches	PW1-3	PW1-5	PW5-4	PW5-5	PW3-1	PW3-4	PW5-1	Average	Rank
Case 1a: Grain size Analysis_Three Models	0.411	0.590	0.312	0.392	0.032	0.071	0.027	0.262	**8**
Case 1b: Permeameter Test	0.791	0.639	0.544	0.592	0.037	0.187	0.036	0.404	**9**
Case 1c: Slug Test	0.190	0.037	0.071	0.049	0.072	0.013	0.048	0.069	**6**
Case 2a: HPT_McCall and Christy Model	0.175	0.058	0.062	0.070	0.069	0.013	0.042	0.070	**7**
Case 2b: HPT_Borden Model	0.180	0.046	0.060	0.060	0.068	0.010	0.044	0.067	**5**
Case 2c: HPT_Power Law Model	0.180	0.045	0.059	0.060	0.067	0.010	0.042	0.066	**4**
Case 3a: PEST Calibrated Geological Model	0.161	0.031	0.043	0.065	0.032	0.016	0.036	0.055	**2**
Case 3b: Averaged THT Geological Model	0.158	0.034	0.035	0.066	0.034	0.013	0.058	0.057	**3**
Case 3c: Highly Parameterised THT Analysis	0.091	0.033	0.049	0.061	0.017	0.009	0.028	0.041	**1**

Table S13. L_1 norm of observed versus simulated drawdowns from seven pumping tests under transient-state condition for model validation.

Approaches	PW1-3	PW1-5	PW5-4	PW5-5	PW3-1	PW3-4	PW5-1	Average	Rank
Case 1a: Grain size Analysis_Three Models	0.353	1.211	0.297	1.244	0.002	0.008	0.001	0.445	**7**
Case 1b: Permeameter Test	1.183	0.601	0.567	0.713	0.003	0.054	0.003	0.446	**8**
Case 1c: Slug Test	0.105	0.002	0.008	0.004	0.010	0.0002	0.004	0.019	**6**
Case 2a: HPT_McCall and Christy Model	0.073	0.006	0.006	0.009	0.009	0.0002	0.003	0.015	**5**
Case 2b: HPT_Borden Model	0.073	0.003	0.006	0.005	0.009	0.0002	0.003	0.014	**4**
Case 2c: HPT_Power Law Model	0.071	0.003	0.005	0.005	0.008	0.0001	0.003	0.014	**4**
Case 3a: PEST Calibrated Geological Model	0.069	0.002	0.004	0.009	0.002	0.0003	0.002	0.013	**3**
Case 3b: Averaged THT Geological Model	0.058	0.002	0.002	0.009	0.002	0.0003	0.006	0.011	**2**
Case 3c: Highly Parameterised THT Analysis	0.024	0.002	0.004	0.008	0.001	0.0001	0.002	0.006	**1**

Table S14. L_2 norm of observed versus simulated drawdowns from seven pumping tests under transient-state condition for model validation.

Approaches		PW1-3	PW1-5	PW5-4	PW5-5	PW3-1	PW3-4	PW5-1
Case 1a: Grain size Analysis_Three Models	Slope	1.45	11.00	3.27	4.35	0.76	3.12	0.47
	Intercept	0.20	-0.03	0.07	-0.02	0.01	0.00	-0.05
Case 1b: Permeameter Test	Slope	1.83	6.59	3.66	5.21	1.24	10.90	0.45
	Intercept	0.54	0.30	0.29	0.12	0.03	-0.14	-0.03
Case 1c: Slug Test	Slope	0.21	0.71	0.52	0.64	0.40	0.83	1.77
	Intercept	0.18	0.07	0.09	0.06	-0.02	0.02	-0.04
Case 2a: HPT_McCall and Christy Model	Slope	0.16	0.97	0.66	0.64	0.07	0.88	4.10
	Intercept	0.12	0.06	0.05	0.05	-0.01	0.00	0.03
Case 2b: HPT_Borden Model	Slope	0.15	0.61	0.41	0.59	0.09	0.88	2.79
	Intercept	0.12	0.06	0.05	0.04	-0.01	0.00	-0.01
Case 2c: HPT_Power Law Model	Slope	0.17	0.58	0.44	0.55	0.10	0.90	2.67
	Intercept	0.13	0.06	0.06	0.04	-0.01	0.00	0.00
Case 3a: PEST Calibrated Geological Model	Slope	1.21	0.69	0.48	0.36	1.25	1.02	0.47
	Intercept	0.00	0.01	0.01	0.02	0.04	-0.02	-0.02
Case 3b: Averaged THT Geological Model	Slope	1.17	0.60	0.64	0.34	1.25	0.78	0.27
	Intercept	0.03	0.01	0.00	0.01	0.03	-0.01	-0.03
Case 3c: Highly Parameterised THT Analysis	Slope	0.95	0.73	0.44	0.39	1.10	1.25	0.51
	Intercept	0.01	0.01	0.03	0.01	0.02	-0.02	-0.02

Table S15. Statistics of the linear model fit from scatterplots of simulated versus observed drawdowns during model validation under transient state condition.

Approaches	PW1-3	PW1-5	PW5-4	PW5-5	PW3-1	PW3-4	PW5-1
Case 1a Grain size Analysis_Three Models	✓ 0.411	✓ 0.590	✓ 0.312	✓ 0.392	✓ 0.032	✓ 0.071	✓ 0.027
	0.716	1.221	0.601	0.778	0.053	0.175	0.107
Case 1b: Permeameter Test	✓ 0.791	✓ 0.639	✓ 0.544	✓ 0.592	✓ 0.037	✓ 0.187	✓ 0.036
	1.345	0.922	0.799	0.795	0.174	0.330	0.239
Case 1c: Slug Test	! 0.190	✗ 0.037	✗ 0.071	✓ 0.049	✓ 0.072	✓ 0.013	✓ 0.048
	0.190	0.036	0.058	0.076	0.097	0.022	0.049
Case 2a: HPT_McCall and Christy Model	✓ 0.175	✓ 0.058	✓ 0.062	✓ 0.070	✓ 0.069	✓ 0.013	✗ 0.042
	0.208	0.113	0.105	0.099	0.085	0.021	0.040
Case 2b: HPT_Borden Model	✓ 0.180	✓ 0.046	✓ 0.060	✓ 0.060	✓ 0.068	✓ 0.010	✗ 0.044
	0.228	0.089	0.097	0.082	0.081	0.016	0.039
Case 2c: HPT_Power Law Model	✓ 0.180	✓ 0.045	✓ 0.059	✓ 0.060	✓ 0.067	✓ 0.010	✓ 0.042
	0.222	0.081	0.091	0.077	0.081	0.016	0.049

Table S16. L1 norms of observed versus simulated drawdowns from the seven pumping tests under transient state condition. The yellow blocks represent the heterogeneous *Ss* case (*Ss* from case 3a), the blue bocks represent the homogeneous *Ss* case (*Ss* from Zhao and Illman. 2018, case 1a).

Approaches	PW1-3	PW1-5	PW5-4	PW5-5	PW3-1	PW3-4	PW5-1
Case 1a: Grain size Analysis_Three Models	✓ 0.353	✓ 1.211	✓ 0.297	✓ 1.244	✓ 0.002	✓ 0.008	✓ 0.001
	0.903	5.031	1.060	3.549	0.005	0.034	0.024
Case 1b: Permeameter Test	✓ 1.183	✓ 0.601	✓ 0.567	✓ 0.713	✓ 0.003	✓ 0.054	✓ 0.003
	2.937	1.139	1.036	1.114	0.035	0.126	0.072
Case 1c: Slug Test	✗ 0.105	! 0.002	✗ 0.008	✓ 0.004	✓ 0.010	✓0.0002	✓ 0.004
	0.102	0.002	0.007	0.011	0.013	0.0008	0.005
Case 2a: HPT_McCall and Christy Model	✓ 0.073	✓ 0.006	✓ 0.006	✓ 0.009	✓ 0.009	✓0.0002	! 0.003
	0.080	0.020	0.017	0.020	0.011	0.0008	0.003
Case 2b: HPT_Borden Model	✓ 0.073	✓ 0.003	✓ 0.006	✓ 0.005	✓ 0.009	✓0.0002	! 0.003
	0.075	0.010	0.014	0.012	0.010	0.0004	0.003
Case 2c: HPT_Power Law Model	✓ 0.071	✓ 0.003	✓ 0.005	✓ 0.005	✓ 0.008	✓0.0001	✓ 0.002
	0.073	0.009	0.013	0.010	0.010	0.0004	0.003

Table S17. L2 norms of observed versus simulated drawdowns from the seven pumping tests under transient state condition. The yellow blocks represent the heterogeneous *Ss* case (*Ss* from case 3a), the blue bocks represent the homogeneous *Ss* case (*Ss* from Zhao and Illman. 2018, case 1a).

Appendix B. Additional Figures

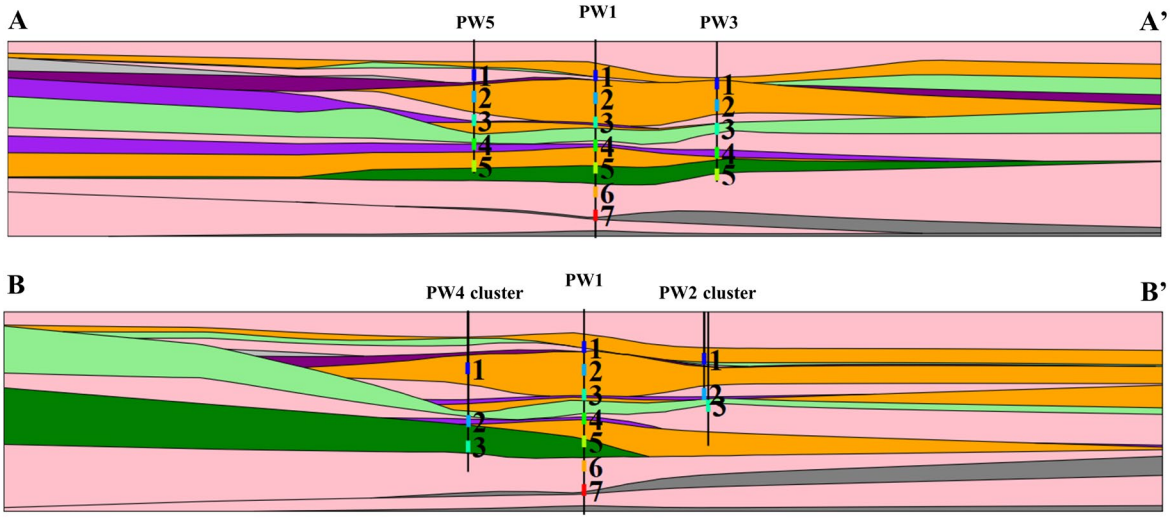


Figure S1. Cross-sectional view of the 19-layer geological zonation model with CMT and PW screened intervals shown in cross sections A-A' and B-B'.

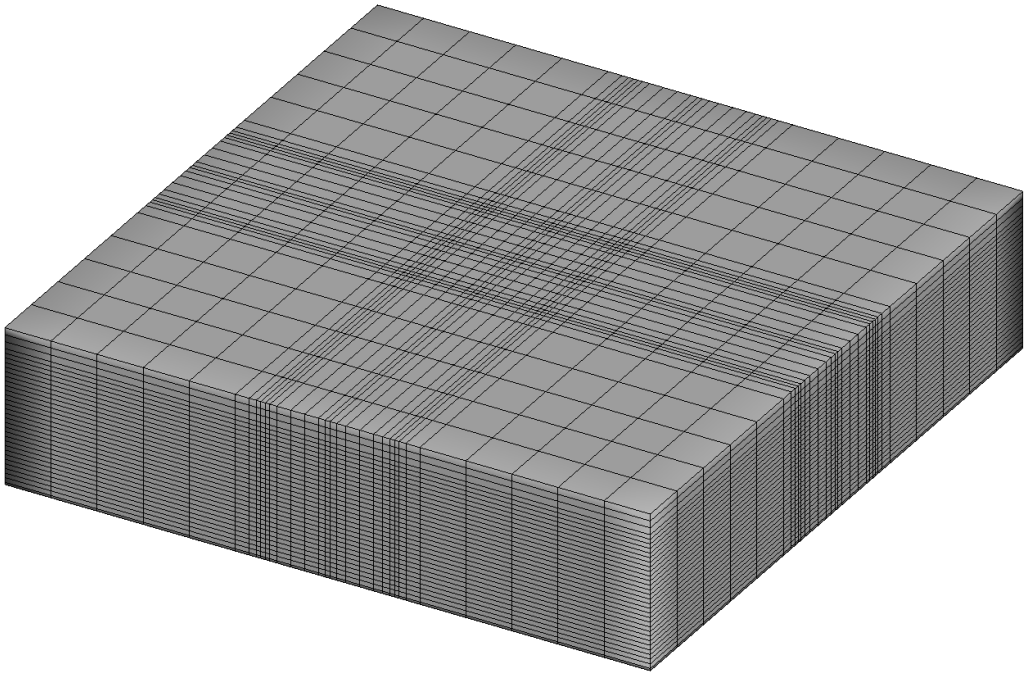
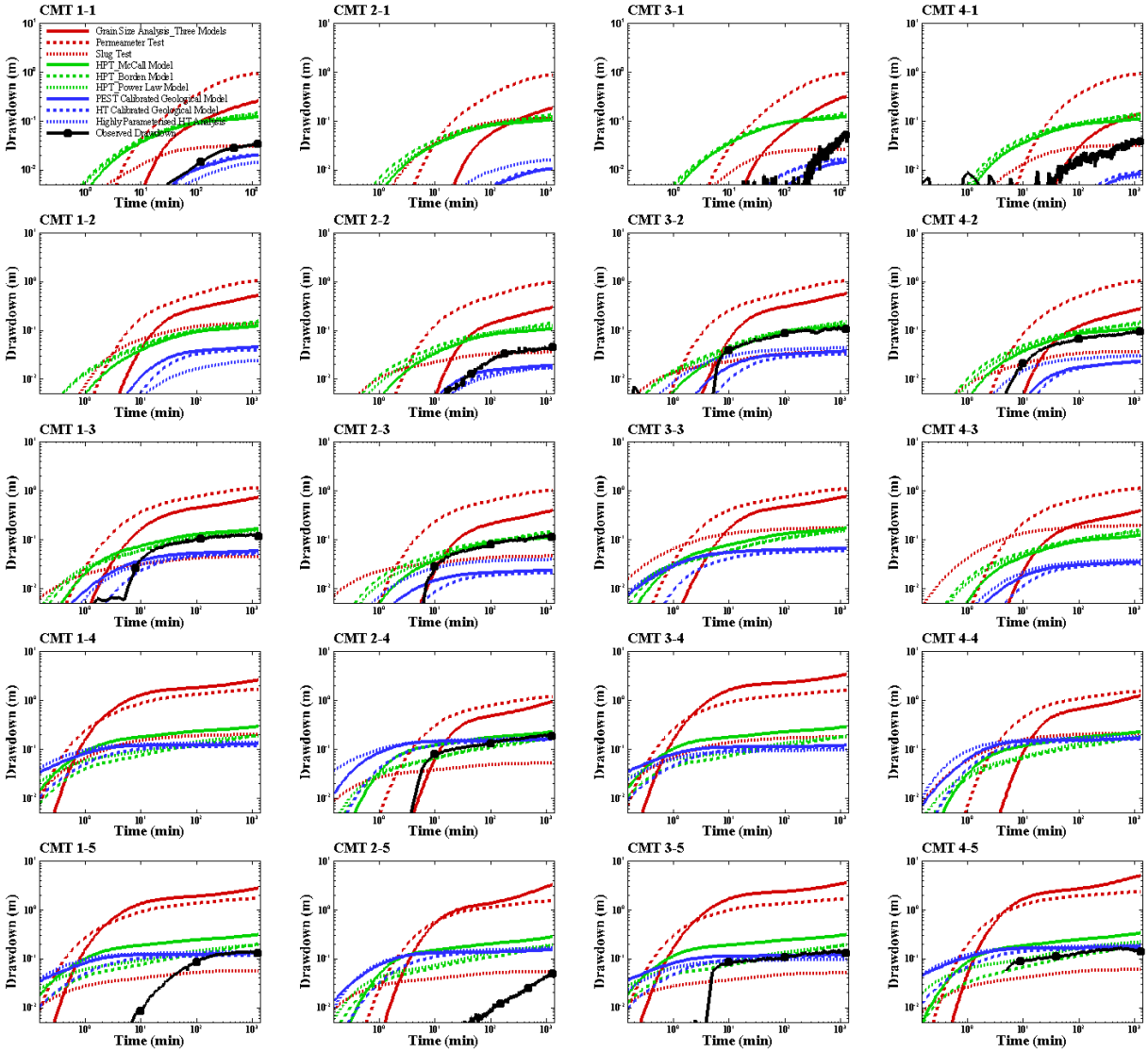


Figure S2. Mesh grid used for forward simulation.


CMT 1-1
CMT 2-1
CMT 3-1
CMT 4-1
CMT 1-2
CMT 2-2
CMT 3-2
CMT 4-2
CMT 1-3
CMT 2-3
CMT 3-3
CMT 4-3
CMT 1-4
CMT 2-4
CMT 3-4
CMT 4-4
CMT 1-5
CMT 2-5
CMT 3-5
CMT 4-5
Drawdown (m)
Time (min)
Grain Size Analysis_Three Models
Permeameter Test
Slug Test
HPT_McCall Model
HPT_Borden Model
HPT_Power Law Model
PEST Calibrated Geological Model
HT Calibrated Geological Model
Highly Parameterised HT Analysis
Observed Drawdown

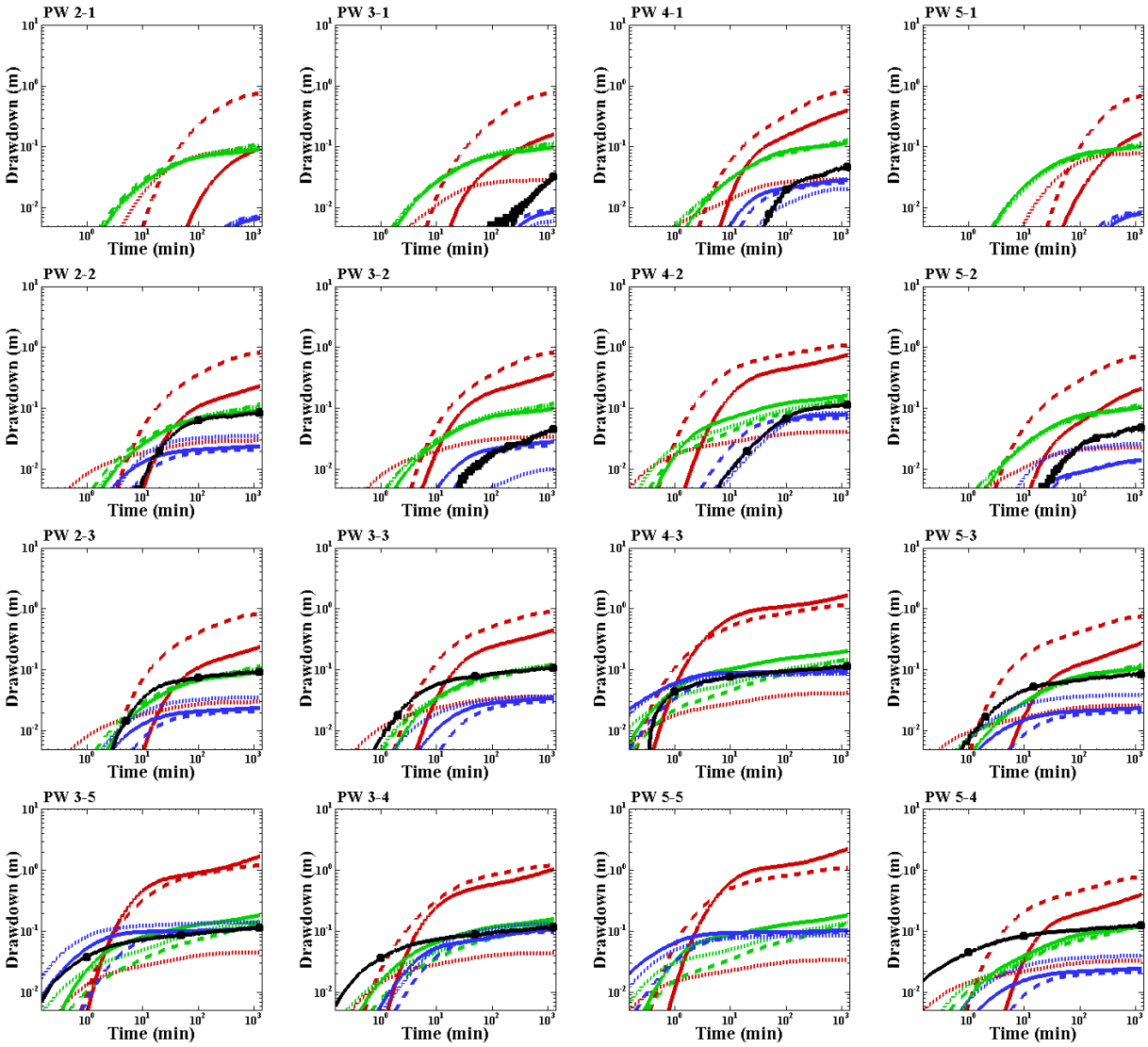


Figure S3. Validation drawdown curve from various approaches at CMT and PW wells from PW1-5.

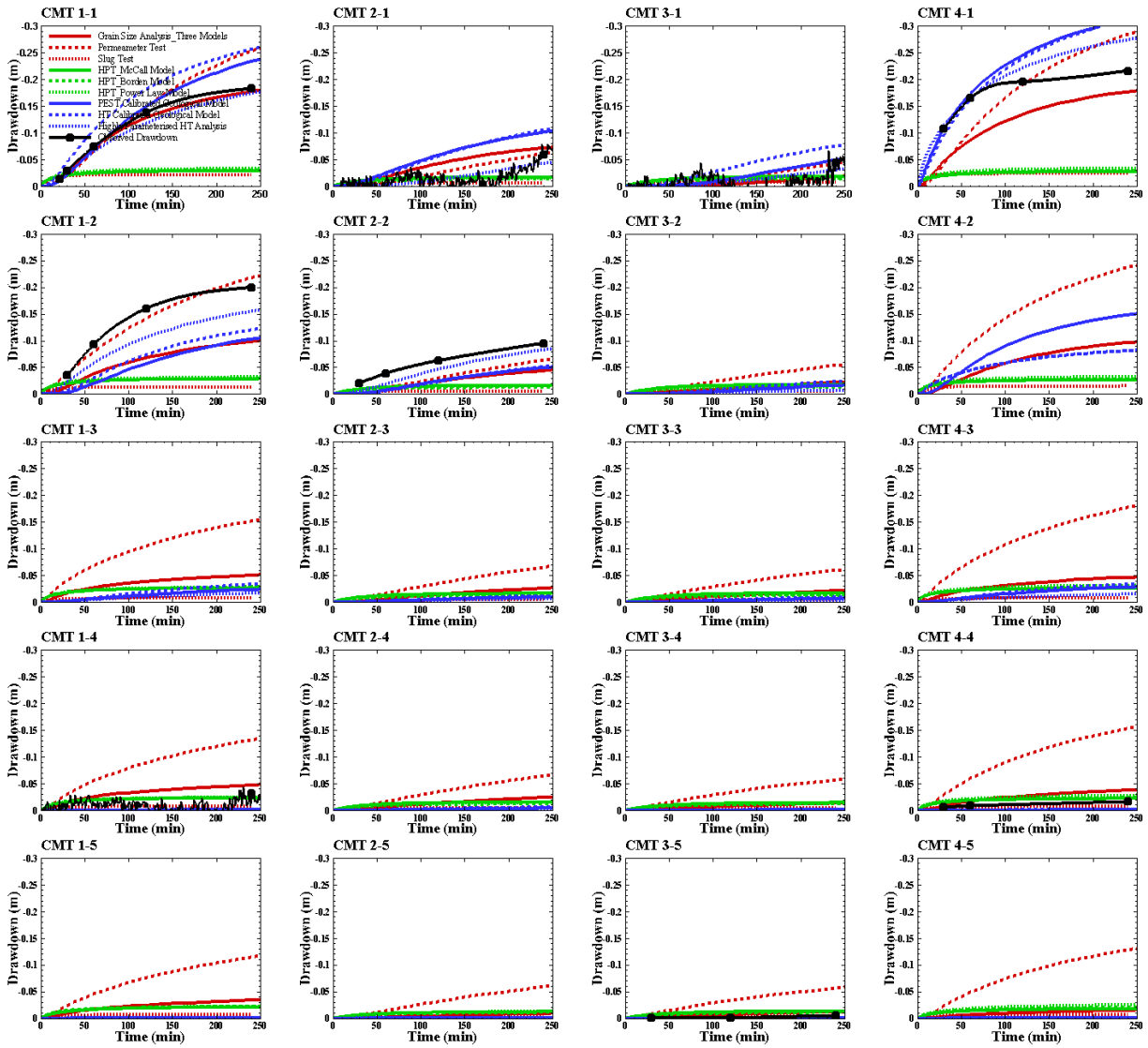
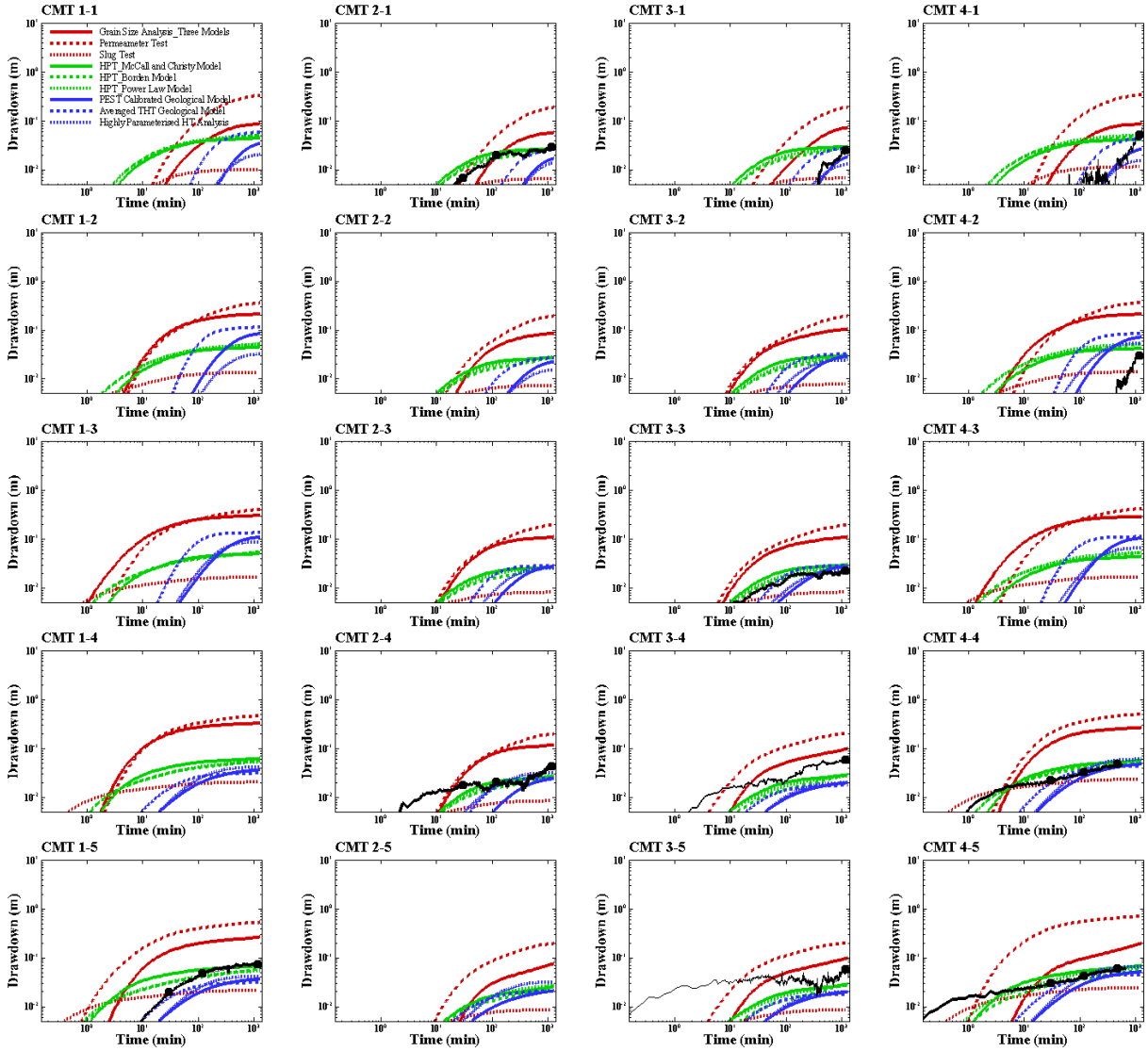


Figure S4. Validation drawdown curve from various approaches at CMT wells from PW3-1.


CMT 1-1
CMT 2-1
CMT 3-1
CMT 4-1
CMT 1-2
CMT 2-2
CMT 3-2
CMT 4-2
CMT 1-3
CMT 2-3
CMT 3-3
CMT 4-3
CMT 1-4
CMT 2-4
CMT 3-4
CMT 4-4
CMT 1-5
CMT 2-5
CMT 3-5
CMT 4-5
Drawdown (m)
Time (min)
Grain Size Analysis_Three Models
Permeameter Test
Slug Test
HPT_McCall and Christy Model
HPT_Borden Model
HPT_Power Law Model
PEST Calibrated Geological Model
Averaged THT Geological Model
Highly Parameterised HTT Analysis

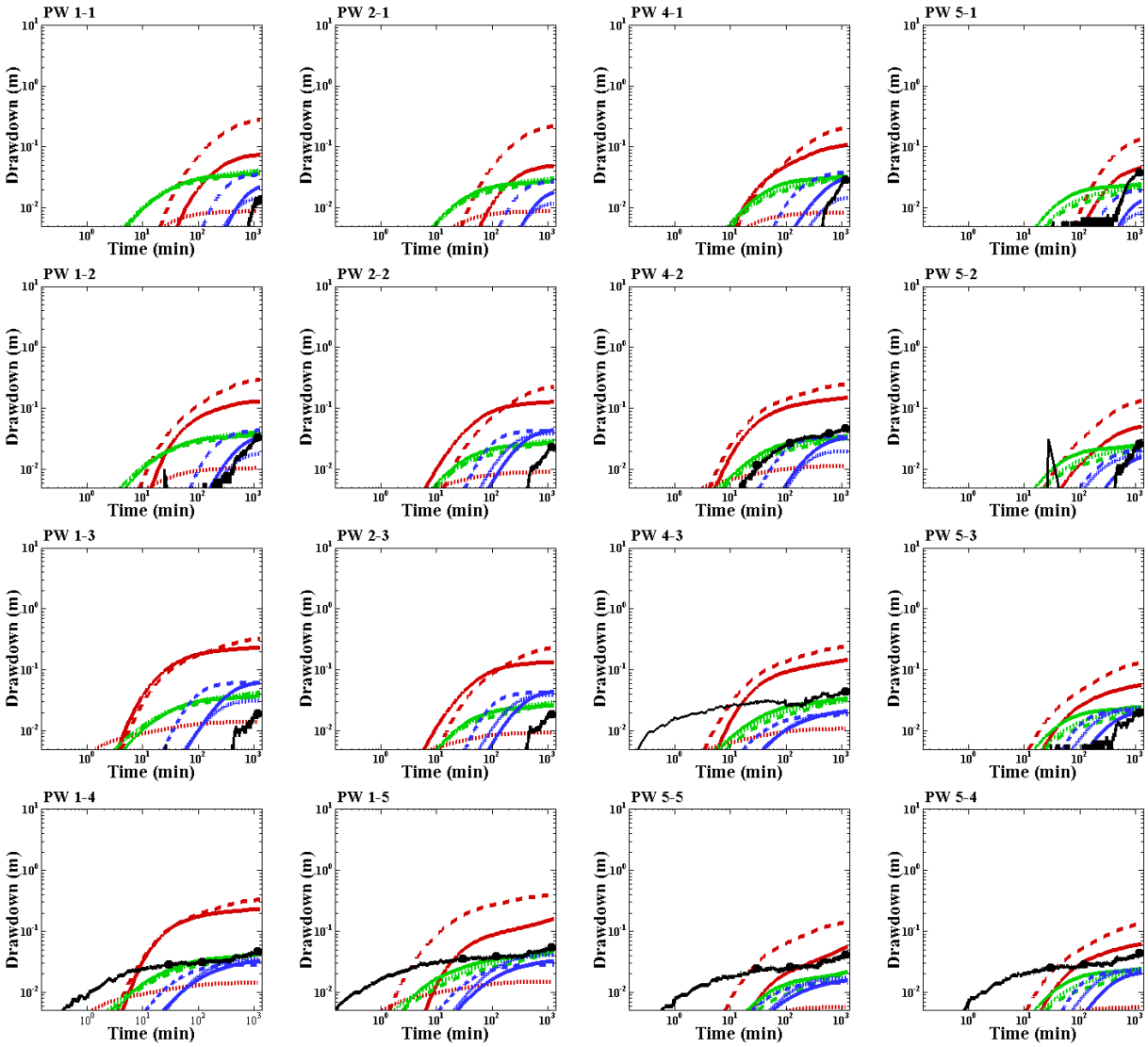


Figure S5. Validation drawdown curve from various approaches at CMT and PW wells from PW3-4.

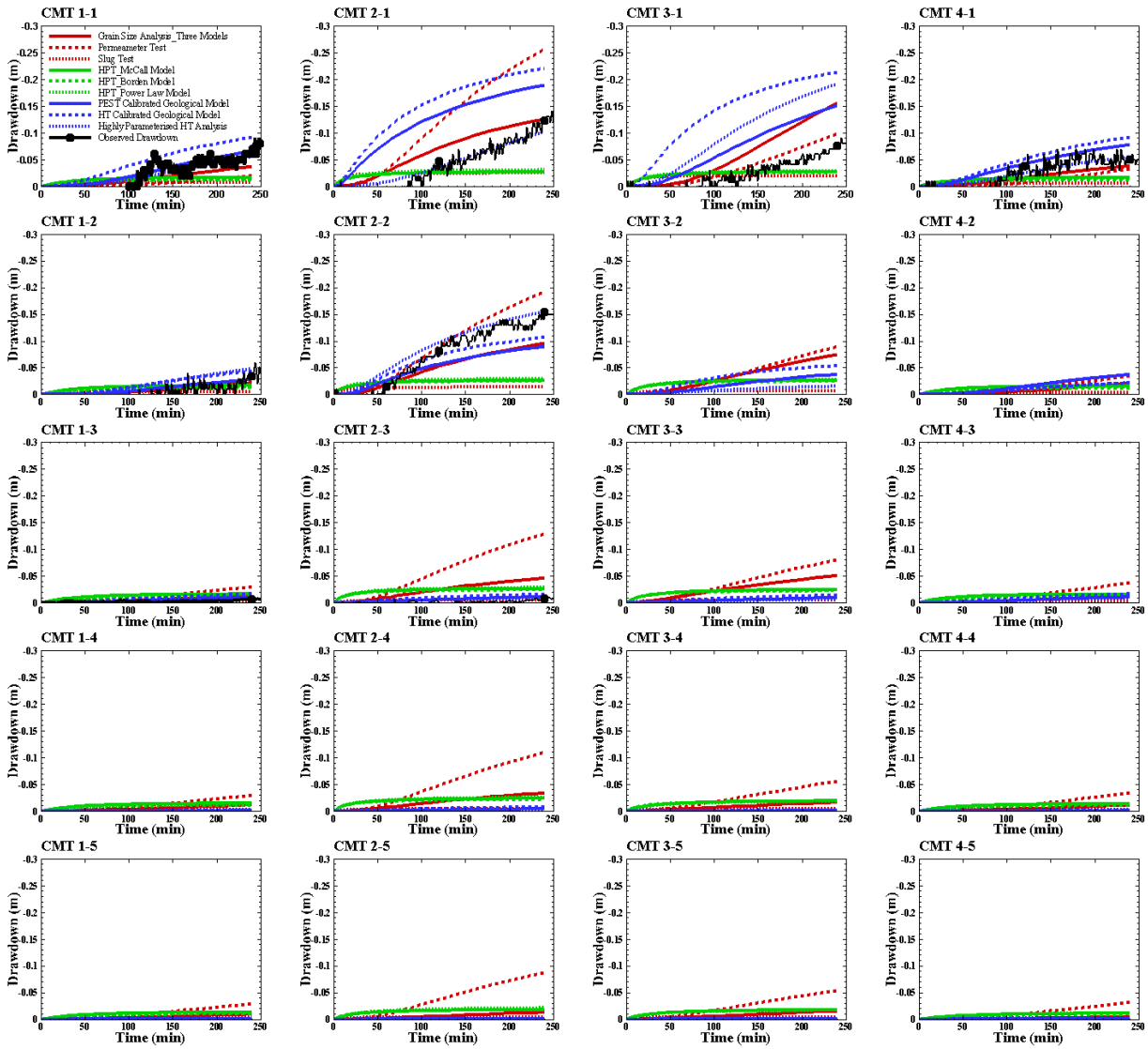
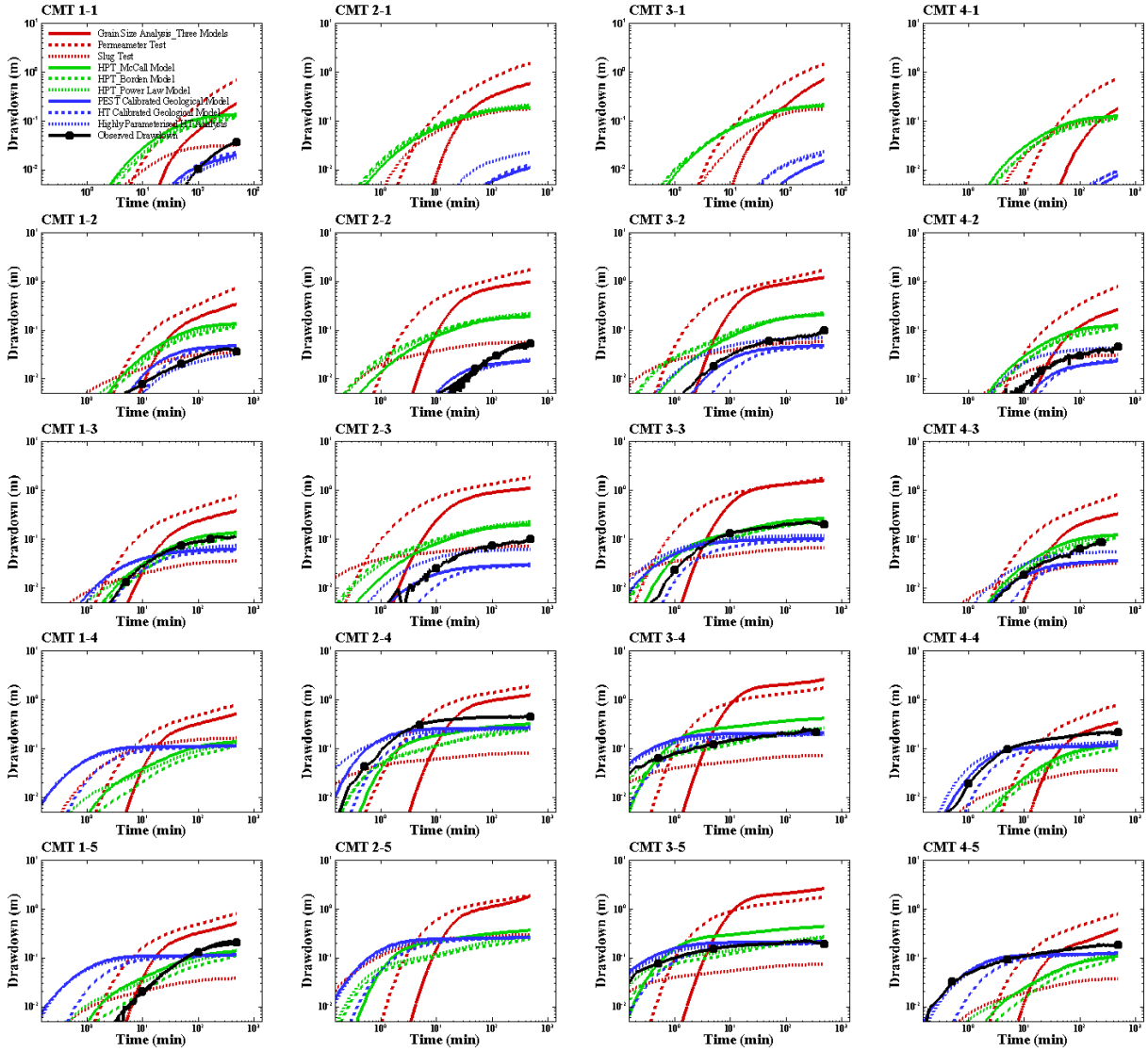


Figure S6. Validation drawdown curve from various approaches at CMT and PW wells from PW5-1.


CMT 1-1
CMT 2-1
CMT 3-1
CMT 4-1
CMT 1-2
CMT 2-2
CMT 3-2
CMT 4-2
CMT 1-3
CMT 2-3
CMT 3-3
CMT 4-3
CMT 1-4
CMT 2-4
CMT 3-4
CMT 4-4
CMT 1-5
CMT 2-5
CMT 3-5
CMT 4-5
Drawdown (m)
Time (min)
Grain Size Analysis_Three Models
Permeameter Test
Slug Test
HPT_McCall Model
HPT_Borden Model
HPT_Power Law Model
PEST Calibrated Geological Model
HT Calibrated Geological Model
Highly Parameterized Grid Cells
Observed Drawdown

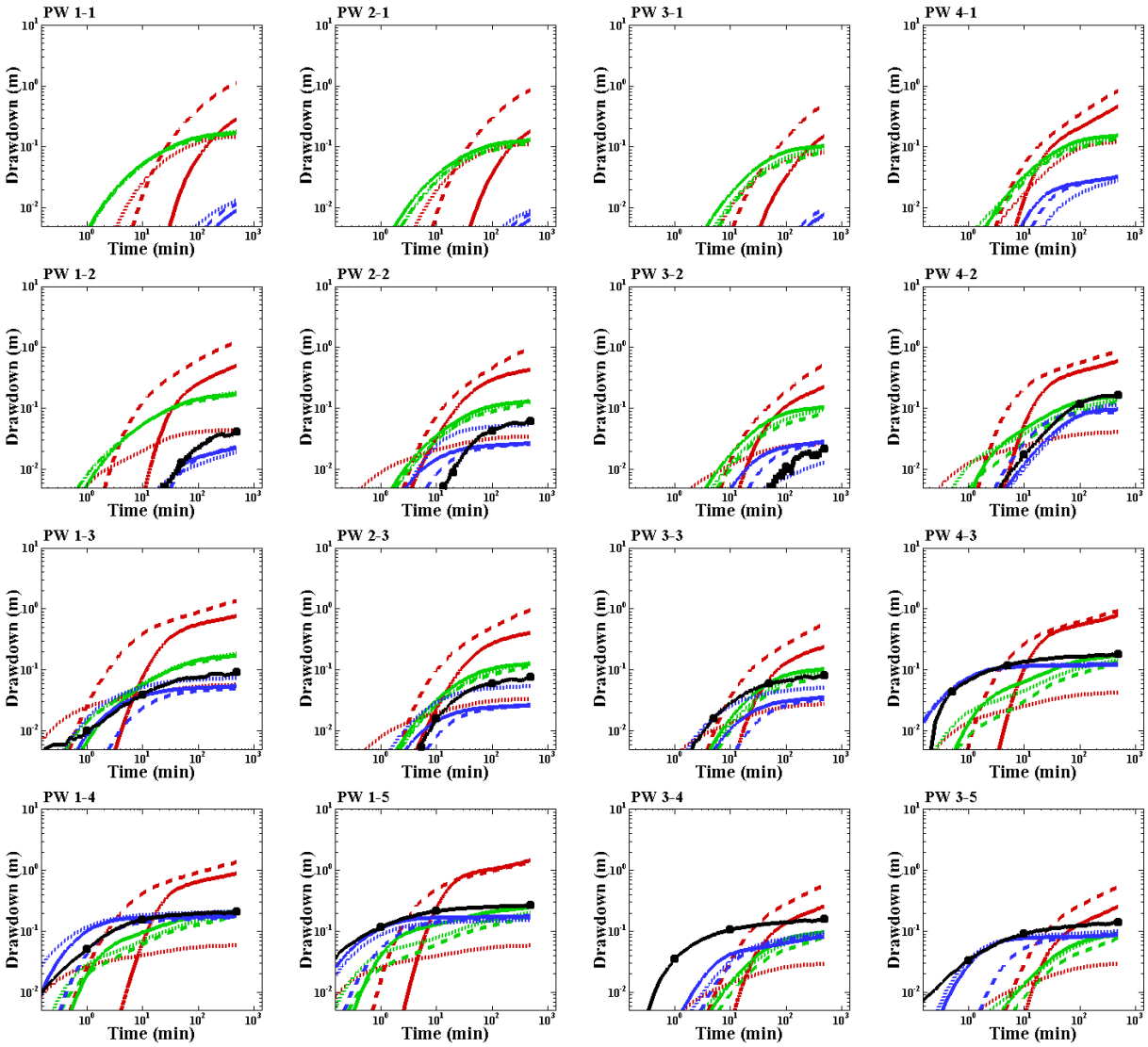
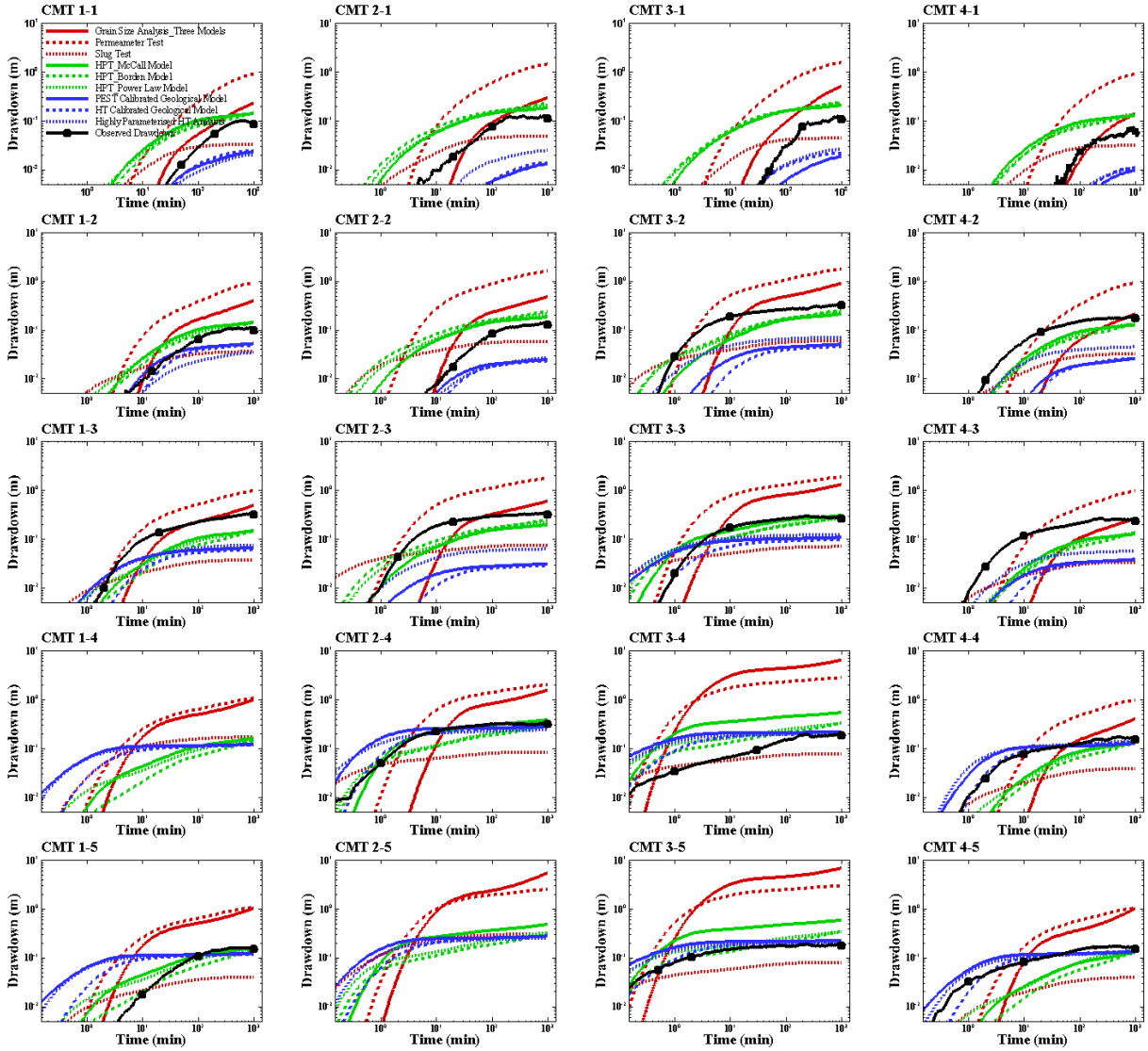


Figure S7. Validation drawdown curve from various approaches at CMT and PW wells from PW5-4.


CMT 1-1
CMT 2-1
CMT 3-1
CMT 4-1
CMT 1-2
CMT 2-2
CMT 3-2
CMT 4-2
CMT 1-3
CMT 2-3
CMT 3-3
CMT 4-3
CMT 1-4
CMT 2-4
CMT 3-4
CMT 4-4
CMT 1-5
CMT 2-5
CMT 3-5
CMT 4-5
Drawdown (m)
Time (min)
Grain Size Analysis_Three Models
Permeameter Test
Slug Test
HPT_McCall Model
HPT_Borden Model
HPT_Power Law Model
PEST Calibrated Geological Model
HT Calibrated Geological Model
Highly Parameterised HT Analysis
Observed Drawdown

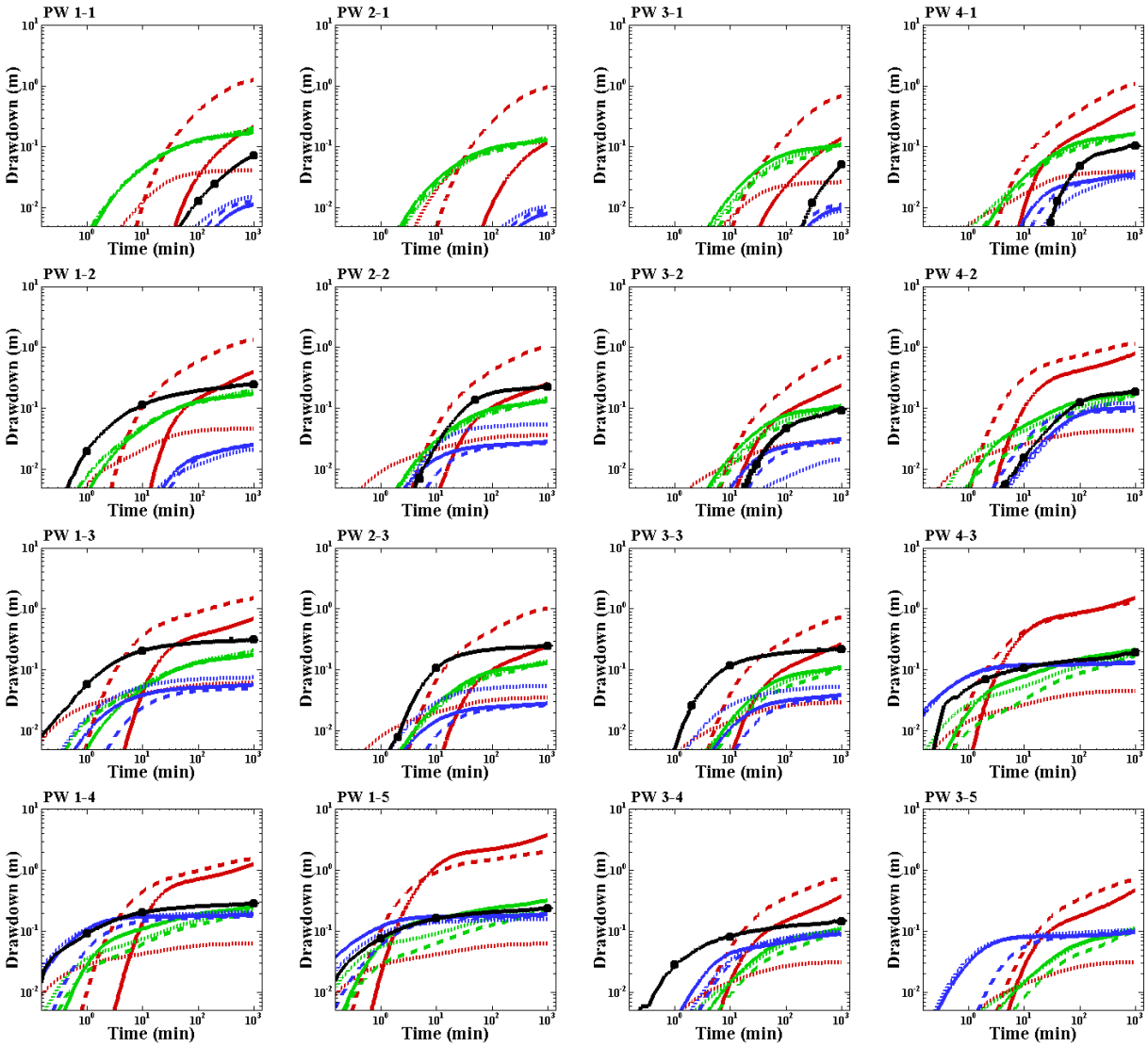


Figure S8. Validation drawdown curve from various approaches at CMT and PW wells from PW5-5.