Integration of Borehole Geophysical Logging with Hydraulic Tomography Analysis for Improved Groundwater Flow and Transport Predictions

by

Chenxi Wang

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Earth Sciences

Waterloo, Ontario, Canada, 2025

© Chenxi Wang 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:

Michael Cardiff

Professor of Geoscience, University of Wisconsin-Madison

Supervisor:

Walter A. Illman

Professor of Earth and Environmental Sciences, University of Waterloo

Internal Members:

Colby M. Steelman

Assistant Professor of Earth and Environmental Sciences, University of Waterloo

André Unger

Associate Professor of Earth and Environmental Sciences, University of Waterloo

Internal-external Number:

James R. Craig

Professor of Civil and Environmental Engineering, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the 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.

Statement of Contributions

The research objectives of my doctoral thesis entail the field experiments of geophysical logging surveys and heat tracer tests, as well as the numerical simulations of groundwater flow, solute/heat tracer transport, and inverse modelling. By collaborating with my esteemed colleagues, I completed five parts of the thesis work.

This doctoral thesis is presented as a collection of three published and two submitted manuscripts. My co-authors, including my supervisor and collaborators, provided guidance, resources, assistance in field work, and editorial input. My contribution to each manuscript, as well as the specific roles of my co-authors, are detailed below.

Study I: Hydraulic Tomography Estimates Improved by Zonal Information From the Clustering of Geophysical Survey Data

This research was a numerical sandbox experiment conducted by Chenxi Wang under the supervision of Dr. Walter A. Illman. Chenxi Wang designed this study with consultations from Dr. Walter A. Illman. Chenxi Wang completed the data analysis and drafted the manuscript. Dr. Walter A. Illman provided editorial input on manuscript drafts.

Citation (Chapter 3):

Wang, C., & Illman, W. A. (2023). Hydraulic Tomography Estimates Improved by Zonal Information From the Clustering of Geophysical Survey Data. *Water Resources Research*, *59*, e2023WR035191. https://doi.org/10.1029/2023WR035191

Study II: Comparison of NMR-Derived Hydraulic Conductivity with Various Hydraulic Testing Methods

This research was conducted at the University of Waterloo by Chenxi Wang under the supervision of Dr. Walter A. Illman. Chenxi Wang designed the study with consultations from Dr. Colby M. Steelman and Dr. Walter A. Illman. Chenxi Wang completed the data collection with assistance from Zeren Ning and guidance from Dr. Colby M. Steelman and Dr. David O. Walsh. Chenxi Wang completed the data analysis and drafted the manuscript. Dr. Colby M. Steelman, Zeren Ning, Dr. David O. Walsh, and Dr. Walter A. Illman provided editorial input on manuscript drafts.

Citation (Chapter 4):

Wang, C., Steelman, C. M., Ning, Z., Walsh, D. O., & Illman, W. A. (2025a). Comparison of NMR-Derived Hydraulic Conductivity with Various Hydraulic Testing Methods. *Groundwater*, *63*(5), 713–724. https://doi.org/10.1111/gwat.70016

Study III: Characterizing Spatial Heterogeneity of Hydraulic Conductivity Using Borehole NMR in a Complex Groundwater Flow System

This research was conducted at the University of Waterloo by Chenxi Wang under the supervision of Dr. Walter A. Illman. Chenxi Wang designed the study with consultations from Dr. Colby M. Steelman and Dr. Walter A. Illman. The NMR data used in this study were collected primarily by Chenxi Wang from Study II. Chenxi Wang completed the data analysis and drafted the manuscript. Dr. Colby M. Steelman and Dr. Walter A. Illman provided editorial input on manuscript drafts.

Citation (Chapter 5):

Wang, C., Steelman, C. M., & Illman, W. A. (2025b). Characterizing Spatial Heterogeneity of Hydraulic Conductivity using Borehole NMR in a Complex Groundwater Flow System. *Water Resources Research*, *61*, e2024WR039717. https://doi.org/10.1029/2024WR039717

Study IV: Integrating NMR Logging with Three-Dimensional Steady-State Hydraulic Tomography for Improved Mapping of Hydraulic Conductivity Distribution

This research was conducted at the University of Waterloo by Chenxi Wang under the supervision of Dr. Walter A. Illman. Chenxi Wang designed the study with consultations from Dr. Walter A. Illman. Chenxi Wang completed the data analysis and drafted the manuscript. Dr. Walter A. Illman provided editorial input on manuscript drafts.

Citation (Chapter 6):

Wang, C., & Illman, W. A. (2025). Integrating NMR Logging with Three-Dimensional Steady-State Hydraulic Tomography for Improved Mapping of Hydraulic Conductivity Distribution. Submitted to *Journal of Hydrology* (Under review).

Study V: Three-Dimensional Analysis of Heat Tracer Transport with Hydraulic Conductivity Distributions from Various Approaches in a Complex Aquifer System

This research was conducted at the University of Waterloo by Chenxi Wang under the supervision of Dr. Walter A. Illman. Chenxi Wang designed the study with consultations from Dr. Walter A. Illman. The heat tracer test data was collected by Chenxi Wang with assistance from Zeren Ning. Dr. Steve J. Berg provided suggestions on the numerical simulation of heat tracer transport. Chenxi Wang completed

the data analysis and drafted the manuscript. Zeren Ning, Dr. Ning Luo, Dr. Zhanfeng Zhao, Dr. Steve J. Berg, and Dr. Walter A. Illman provided editorial input on manuscript drafts.

Citation (Chapter 7):

Wang C., Ning Z., Luo N., Zhao, Z., Berg S. J., & Illman, W. A. (2025). Three-Dimensional Analysis of Heat Tracer Transport with Hydraulic Conductivity Distributions from Various Approaches in a Complex Aquifer System. Submitted to *Water Resources Research* (Under review).

As lead author of these five chapters, I was responsible for conceptualizing the study designs with input from my supervisor, carried out data collection and analysis, and drafted and submitted manuscripts as a corresponding author. My coauthors provided guidance during each step of the research and provided feedback on draft manuscripts.

Abstract

Hydraulic Tomography (HT) has been demonstrated to be a robust approach to characterize the subsurface heterogeneity, which employs the inversion of head data collected from multiple pumping tests to estimate the spatial distribution of hydraulic properties (e.g., hydraulic conductivity (K) and specific storage (S_s)). However, the resolution of K and S_s tomograms is reduced when the number of pumping tests and observation density gradually decrease. Another issue is that HT's ability to predict solute/heat tracer transport behavior has not been rigorously examined for complex aquifer systems. Previous studies showed that different types of data (e.g., geological, geophysical, and other hydraulic testing data) carrying non-redundant information can be integrated with HT analysis to improve the mapping of K heterogeneity. This thesis evaluates the effectiveness of integrating geophysical logging data with HT analysis for improved imaging of K distributions. Furthermore, a heat tracer test was conducted in a highly heterogeneous glaciofluvial deposit to investigate the feasibility of reproducing the spatial distribution of observed temperature responses based on a heat transport model with HT K estimates.

Five sequential studies are documented in this thesis to explore the integration of borehole geophysical logging with HT analysis for improved mapping of K and porosity heterogeneity, which enables enhanced predictions of groundwater flow and solute/heat tracer transport:

(1) Study I integrated two conventional geophysical logging surveys, including electrical conductivity (EC) and gamma ray (GR) logging, with HT analysis to yield 2D K fields in a numerical sandbox experiment. A new spatial conditioning term was proposed to better delineate the hydrostratigraphy from geophysical logging data, which was used to derive the initial guess of K fields for geostatistical inverse modelling of HT analysis. The HT K models with comparative initial guesses of K distributions were evaluated based on their predictive capabilities for groundwater flow and solute transport.

After demonstrating the effectiveness of integrating geophysical logging with HT analysis for improved *K* estimation in a numerical sandbox study, the subsequent studies were conducted at the North Campus Research Site (NCRS) underlain by a highly heterogeneous glaciofluvial deposit.

(2) Study II conducted nuclear magnetic resonance (NMR) logging at the NCRS. Compared to conventional geophysical logging surveys, NMR logging can directly provide *K* estimates, as well as total porosity and effective porosity measurements. The petrophysical relationship between NMR signals and *K* was site-specifically optimized, and the NMR-derived downhole *K* profiles were compared with a variety of hydraulic measurements to evaluate their accuracy and resolution along boreholes.

(3) Study III constructed 3D *K* fields based on downhole NMR *K* profiles and evaluated the representativeness of these *K* models. Various spatial interpolation approaches were employed to generate spatial *K* patterns. A multi-level heterogeneity characterization approach was proposed to better represent the layered porous medium at the NCRS. The model performance to predict groundwater flow was examined through simulating the observed drawdown responses from multiple pumping tests.

(4) Study IV integrated NMR logging with HT analysis for improved characterization of subsurface heterogeneity. To highlight the importance of incorporating high-resolution initial *K* distributions to reduce the smoothness of *K* tomograms, a limited HT calibration dataset with fewer pumping tests and decreased observation density was utilized for model calibration. The effectiveness of this integration was evaluated through a comparative case study using varying numbers of head data for calibration and different spatial interpolation techniques for constructing initial NMR *K* models.

(5) Study V conducted a heat tracer test at the NCRS, in which a dense monitoring network was installed to record temperature responses. The ability of various characterization approaches (e.g.,

HT analysis) to accurately map *K* heterogeneity was investigated by reproducing the complex spatial distribution of the temperature breakthrough curves (BTCs). Additionally, NMR-derived effective porosity was used to map a heterogeneous porosity field. Lastly, the sensitivities of heat tracer plume migration to flow, transport, and thermal parameters were investigated.

The main contributions of these studies are: (1) conventional geophysical logging survey can provide hydrostratigraphic information to improve the resolution and accuracy of HT estimates; (2) NMR logging yields reliable downhole *K* estimates for interbedded layers of gravel, sand, silt and clay; (3) after spatial interpolation, 3D *K* models can be constructed based on NMR logging, which offers reasonable drawdown predictions to pumping tests; (4) integrating NMR logging with HT analysis can provide more representative *K* estimates consistent with the depositional environment, and the integrated models can still yield reliable *K* estimates at high resolution when only a limited head data is available for calibration; and (5) the complex temperature response from a heat tracer test can be best reproduced using HT analysis and NMR logging to represent the heterogeneous *K* and effective porosity fields. Based on their robust performance in predicting groundwater flow and solute/heat transport at different scales, this work advocates the joint use of HT analysis and borehole geophysical logging to characterize subsurface heterogeneity.

Acknowledgements

I would like to extend my sincere gratitude to all those who have supported me during my stay at the University of Waterloo for my PhD work. First and foremost, I would like to express my deepest thanks to my supervisor, Prof. Walter A. Illman, for his professional guidance, financial support, and consistent patience throughout the past six years. We have countless discussions from the very preliminary stage of each project to the completion of every article. Completing this PhD journey has been a challenging experience for me, but I could never have done it without the endless support from my supervisor.

Second, I would like to express my special thanks to Prof. Colby M. Steelman for his constructive advice, helpful edits on my manuscripts, and consistent support of my field work. It is my pleasure to know you and get your support.

I extend my heartfelt thanks to the internal member of my advisory committee, Prof. Andre Unger, for his time, thoughtful suggestions, and constructive criticism, which have helped strengthen this work. I am also sincerely grateful to the internal/external examiner, Prof. James Craig, and the external examiner of my thesis committee, Prof. Michael Cardiff, for their time reviewing this work and providing valuable comments to improve the quality of this thesis.

My sincere appreciation goes to my colleagues and program participants for their time, expertise, and thoughtful critiques. Special thanks to my friend, Mr. Zeren Ning, for the stimulating discussions on the research and the great assistance with the experiments. I hope his pursuit of the PhD degree will go well.

I would like to acknowledge the financial support provided by the China Scholarship Council and Natural Sciences and Engineering Research Council of Canada (NSERC), without which this research would not have been possible.

On a personal note, thank you, my parents. I have stayed away from home for more than six years, but you have always supported me with unconditional love.

Lastly, I appreciate all those who have contributed, directly or indirectly, to the completion of this thesis.

Chenxi Wang

June, 2025

Table of Contents

Examining Committee Membership ii
Author's Declaration iii
Statement of Contributions iv
Study I: Hydraulic Tomography Estimates Improved by Zonal Information From the Clustering of Geophysical Survey Data iv
Study II: Comparison of NMR-Derived Hydraulic Conductivity with Various Hydraulic Testing Methods v
Study III: Characterizing Spatial Heterogeneity of Hydraulic Conductivity Using Borehole NMR in a Complex Groundwater Flow System v
Study IV: Integrating NMR Logging with Three-Dimensional Steady-State Hydraulic Tomography for Improved Mapping of Hydraulic Conductivity Distribution vi
Study V: Three-Dimensional Analysis of Heat Tracer Transport with Hydraulic Conductivity Distributions from Various Approaches in a Complex Aquifer System vi
Abstract viii
Acknowledgements xi
List of Figures xviii
List of Tables xxii
1 Introduction 1
1.1 Hydrogeophysical Characterization 2
1.2 Reviews of Hydraulic Tomography Studies 6
1.3 Predictions of Solute/Heat Tracer Transport Behavior 9
2 Research Topics and Objectives 17
3 Hydraulic Tomography Estimates Improved by Zonal Information from the Clustering of Geophysical Survey Data 26
3.1 Overview 26
3.2 Introduction 27
3.3 Methodology 33
3.3.1 Clustering with Spatial Constraints 33
3.3.2 Geostatistical Inverse Modeling 35
3.4 The Experimental Design 38
3.4.1 Reference *K* Field Generation 38

3.4.2 Geophysical Survey Data 41
3.5 Clustering Validity 43
3.5.1 Zone Geometry Extraction 43
3.5.2 Evaluation of Clustering Performance 45
3.5.3 Discussion 46
3.6 Hydraulic Tomography Analyses 47
3.6.1 Model Setup 47
3.6.2 Zonation Models 49
3.6.3 Model Calibration Results 51
3.6.4 Assessment of Model Validation Performance 56
3.6.5 Solute Transport Predictions 59
4 Comparison of NMR-derived Hydraulic Conductivity with Various Hydraulic Testing Methods 66
4.1 Overview 66
4.2 Introduction 67
4.3 Background 70
4.3.1 Borehole NMR Basics 70
4.3.2 Models for *K* Estimation 72
4.4 Site and Data Descriptions 74
4.4.1 The North Campus Research Site 74
4.4.2 NMR Logging 75
4.5 Hydraulic Testing Methods 77
4.5.1 Permeameter Tests 77
4.5.2 Multi-Level Slug Tests 78
4.5.3 HPT Logging Tests 79
4.6 Model Calibration 80
4.7 Results and Discussion 81
4.7.1 NMR-Derived *K* Profiles 81
4.7.2 Comparison with Hydraulic Testing Methods 83
4.7.3 Evaluation of Petrophysical Models 85
5 Characterizing Spatial Heterogeneity of Hydraulic Conductivity using Borehole NMR in a Complex Groundwater Flow System 89

5.1 Overview 89
5.2 Introduction 90
5.3 Methods 93
5.3.1 Borehole NMR Logging 93
5.3.2 Clustering with Spatial Constraints 95
5.3.3 Multi-Level Heterogeneity Characterization 96
5.4 Site and Data Descriptions 97
5.4.1 Site Geology and Instrumentation 97
5.4.2 NMR Logging 99
5.4.3 Hydraulic Tests 99
5.5 NMR Data Analysis 101
5.5.1 NMR-Derived *K* Profiles 101
5.5.2 Variogram Analysis 102
5.5.3 Spatially Associated Clustering 104
5.6 Results 105
5.6.1 Comparative Cases of Numerical Models 105
5.6.2 Spatial Distributions of *K* Fields 108
5.6.3 Evaluation of Spatial *K* Fields 112
5.6.4 Simulation of Drawdown Responses 114
5.7 Discussion 117
5.7.1 Comparison of Drawdown Prediction Performance 117
5.7.2 Usefulness of NMR Logging for Characterizing Spatial Heterogeneity 120
6 Integrating NMR Logging with Three-Dimensional Steady-State Hydraulic Tomography for Improved Mapping of Hydraulic Conductivity Distribution 121
6.1 Overview 121
6.2 Introduction 122
6.3 Site and Data Descriptions 127
6.3.1 Site Description 127
6.3.2 Steady-State HT Survey 129
6.3.3 NMR Logging Survey 130
6.4 Hydraulic Tomography Analysis 133
6.4.1 Numerical Models of Groundwater Flow 133

6.4.2 Calibrated *K* Fields via HT Analysis 135
6.5 Evaluation of *K* Fields 141
6.5.1 Drawdown Simulations of Groundwater Flow Models 141
6.5.2 Comparison with Borehole *K* Measurements 144
6.6 Discussion 146
6.6.1 Drawdown Prediction Capabilities of Groundwater Flow Models 146
6.6.2 Implications for Characterization of Spatial Heterogeneity 148
7 Three-Dimensional Analysis of Heat Tracer Transport with Hydraulic Conductivity Distributions from Various Approaches in a Complex Aquifer System 152
7.1 Overview 152
7.2 Introduction 153
7.3 Experimental Design 158
7.3.1 The North Campus Research Site 158
7.3.2 The Heat Tracer Test 159
7.3.3 Borehole Testing Data 165
7.3.4 Hydraulic Tomography Survey 168
7.4 Heat Transport Models 168
7.4.1 Numerical Modeling 168
7.4.2 Representations of *K* and Porosity 171
7.5 Results of Heat Transport Simulations 176
7.5.1 Simulated Temperature Breakthrough Curves 176
7.5.2 Heat Tracer Plume Migration 183
7.5.3 Impacts of Transport and Thermal Properties 185
7.6 Discussion 188
7.6.1 Comparison of Predictions Using Different *K* Representations 188
7.6.2 The Advection-Dispersion Transport Model and Its Parameters 191
8 Conclusions and Recommendations 195
8.1 Summary and Conclusions 195
Study I: Hydraulic Tomography Estimates Improved by Zonal Information from the Clustering of Geophysical Survey Data 195
Study II: Comparison of NMR-Derived Hydraulic Conductivity with Various Hydraulic Testing Methods 198

Study III: Characterizing Spatial Heterogeneity of Hydraulic Conductivity using Borehole NMR in a Complex Groundwater Flow System 199

Study IV: Integrating NMR Logging with Three-Dimensional Steady-State Hydraulic Tomography for Improved Mapping of Hydraulic Conductivity Distribution 201

Study V: Predictions of Heat Transport Behaviors in a Complex Groundwater Flow System 204

8.2 Recommendations for Future Studies 206

References 209

Appendix A: Supplementary Information for Study I: Hydraulic Tomography Estimates Improved by Zonal Information from the Clustering of Geophysical Survey Data 232

Appendix B: Supplementary Information for Study III: Characterizing Spatial Heterogeneity of Hydraulic Conductivity using Borehole NMR in a Complex Groundwater Flow System 258

Appendix C: Supplementary Information for Study IV: Integrating NMR logging with Hydraulic Tomography for Improved Mapping of Hydraulic Conductivity 268

Appendix D: Supplementary Information for Study V: Three-Dimensional Analysis of Heat Tracer Transport with Hydraulic Conductivity Distributions from Various Approaches in a Complex Aquifer System 276

List of Figures

Figure 3.1: A general framework of integrating geophysical survey data to improve HT estimates 37

Figure 3.2: True field generation of K and EC for the numerical sandbox experiment 40

Figure 3.3: EC and GR logs along six boreholes 43

Figure 3.4: Four scenarios of zone geometry extracted from the clustering of ln K profiles along six boreholes 45

Figure 3.5: Spatial distribution of K fields in zonation models for different cases. 51

Figure 3.6: Spatial distribution of K tomograms for different cases 53

Figure 3.7: Scatterplots of simulated versus true ln K distribution for different cases 56

Figure 3.8: Validation scatterplots of simulated versus observed drawdowns for different cases 57

Figure 3.9: Summary of color-coded L_1 and L_2 norms for the three zones and overall validation results . 58

Figure 3.10: Relative concentration distributions (C/C_0) at 4 h, 12 h, 48 h and 96 h for all cases 60

Figure 3.11: Breakthrough curves of nine sampling ports during 96 hours of tracer injection and dissipation 62

Figure 3.12: Scatterplots of predicted versus observed C/C_0 at 12 h and 48 h at 47 sampling ports for all cases 63

Figure 3.13: Summary of the color-coded L_1 and L_2 norms for M_0, μ and σ^2 for all cases..... 64

Figure 4.1: Field site and NMR logging information..... 76

Figure 4.2: Profiles of NMR signals near CMT1 to CMT4 77

Figure 4.3: Contour maps of RMSE distribution from the grid search analysis for the calibration of the KGM model 81

Figure 4.4: Comparison of downhole *K* estimates between NMR-derived *K* using the SDR, Seevers, KGM, SOE, and the KGM model, and hydraulic measurements from permeameter, HPT logging, and slug tests 82

Figure 4.5: Scatterplots of measured *K* from hydraulic testing methods versus estimated *K* using four NMR models 86

Figure 4.6: Summary of the color-coded MAE and RMSE for the comparison between NMR-derived *K* and hydraulic measurements from permeameter and slug tests 88

Figure 5.1: A general framework for constructing a multi-level heterogeneity model of *K* from borehole NMR logs for a layered porous medium 97

Figure 5.2: Monitoring well and NMR logging locations at the NCRS 98

Figure 5.3: The NMR-derived *K* profiles along eight boreholes with co-located permeameter-based *K* measurements and borehole geologic logs 102

Figure 5.4: Clustered *K* profiles from eight NMR logs a) with and b) without spatial constraints 105

Figure 5.5: Contour plots of *K* distribution in A-A' and B-B' sections based on comparative cases of spatial modeling 109

Figure 5.6: Contour plots of *K* distribution in C-C' and D-D' sections 112

Figure 5.7: Comparison of *K* in PW3 between the model-derived *K* based on NMR logging and the permeameter *K* measurements 113

Figure 5.8: Comparison of simulated and observed drawdown 115

Figure 5.9: Summary of the color-coded NMAE and NRMSE for the comparison between the simulated and observed drawdown responses 117

Figure 6.1: Locations of monitoring wells for pumping tests and NMR logging 129

Figure 6.2: The NMR-derived *K* profiles along eight boreholes..... 132

Figure 6.3: Contour plots of NMR-derived *K* distributions in four 2D cross-sections based on the spatial correlation of NMR logging data 134

Figure 6.4: Contour plots of *K* distributions in A-A' and B-B' cross-sections based on comparative cases of HT analysis 137

Figure 6.5: Contour plots of *K* distributions in C-C' and D-D' cross-sections based on comparative cases of HT analysis. 138

Figure 6.6: Statistics of the spatial *K* fields for all NMR, HT, and integrated models..... 140

Figure 6.7: Validation scatterplots between simulated and observed drawdowns of different spatial *K* models 142

Figure 6.8: Summarized mean absolute error (L_1) and mean squared error (L_2) for the drawdown fits between model simulations and field observations for each pumping test 144

Figure 6.9: Comparison of *K* profiles at nine well locations between the model-derived *K* and the permeameter *K* measurements 146

Figure 6.10: Locations and dimensions of Aquifers 1 and 2 identified by different spatial *K* models 150

Figure 7.1: Locations of monitoring wells and testing boreholes 159

Figure 7.2: Observed temperature curves over time for all observation points during the heat tracer test 164

Figure 7.3: Downhole profiles of *K* and porosity at nine observation wells of the NCRS. 166

Figure 7.4: Contour plots of *K* distributions in A-A' and B-B' sections based on comparative *K* representations for heat transport simulations 171

Figure 7.5: Contour plots of *K* distributions in C-C' and D-D' cross-sections based on comparative *K* representations for heat transport simulations 174

Figure 7.6: The perspective view of the porosity fields within the data domain 175

Figure 7.7: The temperature breakthrough curves from field observations and model simulations for all observation ports 177

Figure 7.8: The simulated spatiotemporal evolution of heat tracer plumes in the plan view crossing PW1-5 and PW4-3, and three cross-sectional views of B-B', C-C', and D-D' 185

Figure 7.9: The simulated temperature BTCs at seven representative observation ports considering the impacts of transport and thermal parameters 187

List of Tables

Table 3.1: Summary of Validity Check Results Using CVIs 45

Table 3.2: Summary of cases for HT analyses. 48

Table 3.3: Transport parameters used in the solute transport simulation. 59

Table 4.1: Summary of the number of hydraulic measurements at each location and statistical parameters of *K* estimates from each hydraulic testing method 78

Table 4.2: Summary of optimized constants after calibration with permeameter test data 81

Table 4.3: Compilation of the calibrated SDR constants for glacial deposits given $n = 2$ 87

Table 5.1: Information on the NMR probes used in the eight NMR logs. 99

Table 5.2: Summary of pumping tests for groundwater flow predictions 100

Table 5.3: Summary of input datasets and spatial interpolation methods for comparative cases of spatial *K* models 107

Table 5.4: Statistics of *K* estimates within each layer of the spatial *K* model derived from the multi-level heterogeneity characterization of eight NMR logs (Case: *Multi-level_8*) 119

Table 6.1: Summary of the pumping tests for model calibration of the *HT_4* and *HT_8* scenarios and their validation 130

Table 6.2: Information on the initial *K* distributions and number of head data for comparative cases of HT models 135

Table 7.1: Three stages of the heat tracer test 161

Table 7.2: Summary of seven comparative cases of spatial *K* models 169

Table 7.3: Transport and thermal parameters used in heat transport simulations 170

Table 7.4: Zonal K and porosity values for the 19-layer geological zonation model 173

Table 7.5: The tracer total response strength (M_0) (°C·day) at each observation port for the observed and simulated heat transport 182

1 Introduction

Investigations of groundwater flow and solute/heat tracer transport rely heavily on the accurate knowledge of hydraulic properties (e.g., hydraulic conductivity (K), porosity (θ), and specific storage (S_S)). Therefore, hydrogeologists have used various techniques (e.g., hydraulic tests, geophysical surveys, and tracer tests) to characterize the subsurface and determine hydraulic parameters. Hydraulic testing methods usually provide direct estimates of hydraulic properties. Traditional hydraulic testing methods, including slug tests, single-hole packer tests, and pumping tests, have been extensively applied to determine K, which are widely regarded as standards in the industry. These hydraulic testing methods measure directional components of K at various scales and thus are suitable for multiple tasks. A slug test rapidly displaces water in the well, and the corresponding recovery of the water level is measured over time. Hence, the effective K near the well is measured, which is biased towards the horizontal K component, assuming radial flow toward the well screen. In a pumping test, the groundwater in the well is consistently pumped, resulting in an increasing drawdown over time. Conventionally, the time-drawdown data can be interpreted using analytical methods (e.g., the Thiem steady-state model, the transient Theis curve matching, and the Cooper-Jacob methods) to yield K estimates between the pumping and observation wells. However, these methods typically assume that the subsurface is homogeneous, regardless of its heterogeneous nature. Wu et al. (2005) found that conventional analysis of pumping tests (i.e., analytical solutions) is insufficient in predicting accurate drawdown behaviors, and the estimated K may not be representative of the aquifer.

Over the past several decades, more sophisticated hydraulic testing methods (e.g., wellbore flowmeter tests, direct-push permeameter tests, and multi-level slug tests) have been developed to characterize subsurface heterogeneity, particularly for vertical K trends. For example, a wellbore flowmeter test (Molz & Young, 1993; Rehfeldt et al., 1992) measures the effective horizontal K

component at the meter scale, assuming that the aquifer is uniformly layered and K is proportional to the inflow rate at each layer. For a direct-push permeameter test (DPP, see Butler et al., 2007), one inserts the probe into a borehole and injects a stream of water horizontally. The pressure difference at two transducers installed vertically is observed for K estimation by Darcy's law, assuming pressure equality at the same distance spherically. Over the past decade, a new direct-push hydraulic testing tool, known as the Hydraulic Profiling Tool (HPT), has been developed to map downhole K variations with higher resolution (McCall and Christy, 2020). The HPT probe injects a set flow of water into the formation while recording the injection pressure, flow rate, and the probe advance rate during the logging process. Multiple models (e.g., McCall, 2010; Borden et al., 2021; Zhao & Illman, 2022b) have been developed to estimate K from these signals. The sampling interval of direct-push HPT logging tests is 1.5 cm.

1.1 Hydrogeophysical Characterization

In addition to the more direct hydraulic testing methods, geophysical surveys, especially borehole geophysical logging techniques, are also widely used to characterize subsurface heterogeneity. During the logging process, various geophysical properties are recorded, which can be converted to hydraulic properties using petrophysical models. However, those petrophysical relationships may provide a non-unique quantification of the hydraulic property (Yeh et al., 2002). For instance, Archie's law (Archie, 1942) focuses on the pore-fluid electrical conductivity (EC), regardless of the pore-surface conduction. Therefore, high EC values may appear in clay-rich sediments, confounding interpretations of water content distribution. Another issue is that potential petrophysical models only describe a limited range of hydraulic properties: dielectric permittivity from the ground penetrating radar (Topp et al., 1980) provides an estimate of soil water content; P-wave velocity from the seismic method (Raymer et al., 1980) can be related to porosity; while the gamma-ray radiation method (Díaz-Curiel et al., 2021) yields gamma ray counts for clay content estimation.

As borehole geophysical logging is usually cost-effective and causes little to no disturbance to the subsurface, different geophysical tools can be collectively used to improve the understanding of the aquifer characteristics. Although the non-uniqueness of petrophysical models can be effectively reduced by incorporating multiple datasets, reliable quantification of *K* based on geophysical logging techniques remains a challenge. In recent years, the development of machine learning (ML) methods has facilitated *K* interpretations based on conventional geophysical datasets. Paradis et al. (2015) classified hydrofacies by clustering *K* and multiple datasets from cone penetrometer tests and a soil moisture and resistivity probe. Relevance vector machines were then used to model hydrofacies and to predict *K* based on geophysical data. Kawo et al. (2024) trained multiple ensemble ML models to relate airborne electromagnetic resistivity with borehole lithology data and predicted 3D hydrofacies and *K* profiles along streams. Nevertheless, the complex petrophysical relationships interpreted with ML methods are usually site-specific. Meanwhile, ML methods cannot yield high-resolution *K* estimates, especially within highly heterogeneous environments.

A novel geophysical technique is required to map *K* variations, which can directly relate geophysical properties to *K*. Nuclear magnetic resonance (NMR) tools were initially used in the petroleum industry to detect the formation hydraulic properties. With the development of slim-hole NMR logging tools (Perlo et al., 2013; Sucre et al., 2011; Walsh et al., 2013; Zhu et al., 2021), NMR's utility has been enabled for fast application in shallow unconsolidated sediments. The geophysical applications of NMR use the nuclear spin of the hydrogen protons in water molecules. The total amplitude of NMR signals is associated with hydrogen in water molecules, allowing for the measurement of water content. More importantly, the nuclear spin echo decay yields a temporal distribution, indicating that relaxation processes interact with disparate pore structures within the porous medium. As a result, it can also relate to the pore characteristics and be further used to interpret permeability or *K*. Various petrophysical relationships have been developed to estimate *K* from NMR responses (Coates & Dumanoir, 1973;

Dlugosch et al., 2013; Kenyon et al., 1988; Seevers, 1966; Timur, 1968; Walsh et al., 2013). Evaluation of these petrophysical relationships (e.g., Crow et al., 2022; Dlubac et al., 2013; Kendrick et al., 2021, 2023; Knight et al., 2016; Maurer & Knight, 2016) reveals that the NMR-derived *K* estimates are very consistent with *K* measurements from hydraulic tests. The optimal model fit in *K* is typically within one order of magnitude, making the NMR technique a reliable tool for high-resolution site characterization. On the other hand, a cutoff time (typically 33 ms for unconsolidated materials) can be assigned in the temporal distribution of spin echo decay to distinguish the mobile water content from the immobile (i.e., bounded) water content. Hence, the mobile-domain porosity for the saturated sediments can be estimated, which is effective to the advective transport process and essential to the accurate simulation of solute/heat tracer transport. The NMR's ability to measure mobile porosity is unique among borehole geophysical logging tools.

To better characterize heterogeneity between boreholes, patterns observed along individual boreholes are correlated using various spatial modelling techniques. Ordinary kriging estimation might be the most popular method for interpolation, which assumes that the random field of targeted properties follows a continuous Gaussian process. To capture different non-Gaussian fields, discontinuous facies models (de Marsily et al., 2005) are used to better describe geological features, e.g., the indicator estimation and simulation (Deutsch & Journel, 1998) and the Markov chain model (Carle et al., 1998). However, geological structures may not be well reflected in these models, diminishing their predictive capacity (Berg & Illman, 2015; Salamon et al., 2007; Zhao et al., 2023).

The zonation method (Dietrich & Tronicke, 2009; Doetsch et al., 2010; Dumont et al., 2018; Gueting et al., 2015; Liu et al., 2010; Paasche et al., 2010; Zhang et al., 2022) honors the geological information and describes aquifer characteristics by partitioning hydrostratigraphic units, in which the accurate delineation of zone geometry is key to subsurface heterogeneity characterization and their predictive performance (Luo et al., 2017; Zhao & Illman, 2018). However, intralayer heterogeneity is

usually neglected in zonation models, potentially limiting their representativeness to realistic hydrogeological environments.

The automatic process of zonation is usually performed by clustering. Various clustering algorithms have been applied to partition hydrostratigraphic zones from geophysical survey data. For example, *k*-means clustering (MacQueen, 1967) is a commonly used non-parametric partitioning algorithm that employs a centroid vector as a prototype for each cluster. This algorithm seeks the minimum sum of feature-to-centroid distances by iteratively assigning features to the closest centroids and generating new centroids based on the current assignments for each cluster. In addition to *k*-means clustering, hierarchical clustering (Dumont et al., 2018), fuzzy-*c* means clustering (Paradis et al., 2015), and mixture-of-Gaussian clustering (Doetsch et al., 2010) have been implemented for clustering geophysical parameters.

Despite the availability of a wide variety of clustering algorithms, the key issue is to understand the inductive principles involved in algorithms (Estivill-Castro, 2002) such that appropriate algorithms can be utilized for input datasets. For example, *k*-means clustering tends to construct clustered features forming ellipsoidal shapes. When *k*-means clustering is employed to partition geophysical observations, outliers are inevitably present in clustering results, and the number of outliers increases with the increasing heterogeneity of the subsurface. A main reason is that the spatial adjacency of samples is neglected in clustering, resulting in samples from disparate regions being partitioned into the same cluster. For example, Zhao and Illman (2022a) performed both fuzzy-*c* means clustering (a variant of the *k*-means method) and visual interpretation of HPT pressure data to delineate hydrostratigraphic layers. Results show that the zonation based on clustering did not lead to improved performance compared to the zonation via visual interpretation in complex geological environments.

Spatial correlation is an essential subsurface characteristic, which assumes that the closer the two locations are, the higher the similarity of the data is at these two locations. Pham (2001) proposed spatial

models for fuzzy *c*-means clustering, in which a spatial penalty term was added to yield spatially smooth membership functions. This study was followed by Paasche et al. (2010), in which spatial constraints were applied to objectively outline the dominant subsurface units. Other independent studies also demonstrated that the use of spatial constraints significantly removed the outliers and better reproduced geological structures (Jia et al., 2021; Kawo et al., 2024).

1.2 Reviews of Hydraulic Tomography Studies

To evaluate the representativeness of 3D *K* distributions, an effective approach is to predict drawdown responses during aquifer pumping tests and compare them to field observations. Previous work has shown that accurate predictions of drawdown responses can be difficult to achieve based on standard in-situ borehole hydraulic measurements (e.g., slug tests, packer tests), particularly in complex environments (Berg & Illman, 2015; Illman et al., 2010; Sun et al., 2024). Accurate depictions of 3D *K* distributions for reliable numerical groundwater flow predictions can be achieved using Hydraulic Tomography (HT) analysis (Berg & Illman, 2011b; Cardiff et al., 2013; Hochstetler et al., 2016; Tiedeman & Barrash, 2020).

HT was proposed and developed over the past three decades. It relies on the joint inverse modelling of hydraulic head data from multiple pumping tests. Water injection/extraction is conducted in targeted aquifers with the installation of monitoring networks, allowing for the observation of drawdown signals over large volumes. Therefore, HT is a cost-effective method for collecting non-redundant datasets on subsurface heterogeneity (Hou et al., 2023; Zha et al., 2015). To interpret the heterogeneity pattern of aquifer-aquitard systems, geostatistical and stochastic inverse methods were developed (Gutjahr & Wilson, 1989; Harvey & Gorelick, 1995; Hoeksema & Kitanidis, 1984, 1989; Kitanidis, 1995; Kitanidis & Vomvoris, 1983; Rubin & Dagan, 1987; T. C. J. Yeh et al., 1996; Yeh et al., 1995), in which geostatistical parameters of aquifer systems were introduced into the inversion, and the most likely spatial distribution of hydraulic properties and their uncertainty are estimated for accurate prediction of

drawdown responses. To date, HT has been developed as a mature technique to accurately map spatial heterogeneity of hydraulic properties through synthetic studies (Cardiff & Barrash, 2011; Gottlieb & Dietrich, 1995; Luo et al., 2020; Tso et al., 2015; Yeh & Liu, 2000; Zhu & Yeh, 2005, 2006), laboratory experiments (Berg & Illman, 2011a; Illman et al., 2007, 2008, 2010, 2012; Liu et al., 2002; Liu et al., 2007a; Luo et al., 2017; Zhao & Illman, 2021), field studies (Berg & Illman, 2011b, 2013, 2015; Bohling et al., 2007; Brauchler et al., 2013; Cardiff et al., 2013; Huang et al., 2011; Illman et al., 2009; Klepikova et al., 2020; Luo et al., 2023; Ning et al., 2023; Tiedeman & Barrash, 2020; Zha et al., 2016; Zhao et al., 2023; Zhao & Illman, 2017, 2018, 2022a, 2022b) and basin-scale aquifer characterization (Daranond et al., 2020; Liu et al., 2020a; Liu et al., 2020b; Luo et al., 2022; Xia et al., 2023; Yeh et al., 2009).

Among various geostatistical inversion codes, Successive Linear Estimator (SLE) (Yeh & Liu, 2000; see also Xiang et al., 2009; Zhu & Yeh, 2005) has been extensively applied in HT analysis, and its effectiveness in interpreting K and S_S fields has been sufficiently validated. The SLE code originates from the traditional cokriging method in geostatistics. Cokriging is a linear estimator in which the correlation between primary (e.g., hydraulic parameters) and secondary variables (e.g., observations of hydraulic head) is assumed to be linear. Since the relationship between hydraulic parameters and observations is essentially nonlinear, traditional cokriging cannot fully utilize observed data. This nonlinear issue can be addressed in SLE through iteratively updating the estimated parameter fields and their conditional covariances during the inversion process, such that the difference between the observed and simulated heads can be utilized iteratively to estimate hydraulic parameters.

Yeh and Liu (2000) developed the Sequential Successive Linear Estimator (SSLE) for steady-state HT analysis, in which steady-state head data for different pumping tests are sequentially incorporated into the geostatistical inversion. Zhu and Yeh (2005) extended this work to transient HT analysis, in which K and S_S are jointly optimized using transient time-head datasets. Although the sequential approach is beneficial to constrain the covariance matrix at a manageable size, the resulting

tomograms may vary with different sequences of incorporating multiple pumping tests into the inversion (Illman et al., 2008). Therefore, Xiang et al. (2009) proposed the Simultaneous Successive Linear Estimator (SimSLE), in which head data from all pumping tests were simultaneously interpreted to calibrate the spatial distribution of K and S_S.

Challenges remain despite the robust performance of the geostatistical inversion approach. A primary issue in subsurface heterogeneity characterization encountered in the industry is that hydraulic response data from boreholes and outcrops are often insufficient. Cardiff et al. (2013) found that significant information about subsurface heterogeneity is progressively lost after removing pumping and/or observation intervals in a field HT survey. It indicates that the deficiency of pumping data can lead to smooth tomograms (i.e., spatial distribution of hydraulic parameters).

Yeh et al. (2008a) envisioned that the fusion of different types of tomographic surveys with data from disparate sources of excitations is a future direction for improved basin-scale subsurface characterization. Accordingly, researchers considered fusion (Bohling & Butler, 2010; Yeh et al., 2008b) of hydraulic, geological, and geophysical data to improve HT performance. For example, Zha et al. (2017) treated the spatial variability of K as two-scale heterogeneity using a nested covariance function whereby site-specific qualitative stratigraphy was incorporated to improve the conditional means and covariances of primary variables (i.e., hydraulic parameters) for HT analysis. Results from the synthetic study and the field application revealed that hydraulic estimates were improved in this manner. Different from the effort by Zha et al. (2017), Zhao et al. (2016) constructed geological zonation models by extracting stratigraphic information from borehole geology logs, which characterized the subsurface as various zones prior to the inversion. Then, geological zonation models were utilized as the initial model for geostatistical inverse modelling for HT analysis. The results of Zhao et al. (2016) and subsequent studies (Luo et al., 2017, 2020; Zhao et al., 2022a; Zhao & Illman, 2017, 2018) showed that the resolution and accuracy of HT estimates were significantly improved after incorporating accurate geological

information, especially in areas with a sparse drawdown monitoring network, whereas inaccurate geological information was found to produce erroneous characterization of aquifer hydraulic properties. Similarly, initial *K* models for HT analysis can also be derived by spatially correlating the borehole *K* estimates. For example, Zhao et al. (2023) generated spatial *K* distributions based on HPT-estimated *K* along boreholes using ordinary kriging interpolation. The HPT-based spatial *K* model was further calibrated using HT analysis, resulting in improved drawdown prediction performance.

Although many efforts were made to enhance the characterization of site heterogeneity by integrating different types of prior information (e.g., geology logs, flowmeter data, and hydraulic testing methods) with HT analysis, few attempts were made to jointly use geophysical and HT surveys to determine the spatial distribution of *K*. A possible reason is that few petrophysical relationships have been built with *K* (except for some novel geophysical logging tools, e.g., NMR logging). Therefore, only hydrostratigraphic information can be inferred from the geophysical survey. However, following the workflow of Zhao and Illman (2017), the geophysically-based zonation models can be populated with zonal *K* estimates via HT analysis, which is then treated as the initial model for highly parameterized HT analysis to refine the spatial *K* distribution. Thus, additional hydrostratigraphic information from the geophysical survey can potentially be integrated in this manner.

1.3 Predictions of Solute/Heat Tracer Transport Behavior

It has consistently been a critical issue to accurately predict solute transport behavior in porous media over the last half-century. Traditionally, the classical advection-dispersion equation (ADE) has been applied to conceptualize transport models, with a homogeneous *K* assigned to the entire aquifer. Based on the classical theory, mechanical dispersion led to the temporal spreading and evolution of the tracer plume. Thus, dispersivity was employed to quantify the hydraulic mixing process, which occurs either longitudinally or transversely to the migration paths. The classical ADE for conservative tracers is:

$$-\frac{\partial}{\partial x_i}(\theta v_i C) + \frac{\partial}{\partial x_i}\left(\theta D_{ij}\frac{\partial C}{\partial x_i}\right) + q_s C_s = \theta\frac{\partial C}{\partial t} \tag{1.1}$$

where C is the solute concentration, θ is the porosity, and t is time. The first term on the left-hand side of this equation is the advective flux term, with v_i representing the average linear groundwater velocity. The second term is the dispersive flux term, with D_{ij} representing the dispersion coefficient tensor. The third term is the source/sink term, with q_s and C_s representing the fluid sink/source and the concentration of the fluid sink/source.

One landmark field experiment for investigating subsurface contaminant transport was conducted at the Canadian Forces Base (CFB) Borden site, where a natural-gradient tracer test was performed (see Mackay et al., 1986) in a mildly heterogeneous sandy aquifer. Based on the analysis of permeameter measurements (Sudicky, 1986), the geometric mean of K and the variance of ln K ($\sigma^2_{\ln K}$) for this aquifer were estimated to be 7.2×10^{-5} m/s and 0.29, respectively. At the Borden site, Sudicky et al. (1983) found that macrodispersivity increases with more heterogeneities encountered along the migration path. Therefore, a unique dispersivity value cannot be assigned to the medium to describe the evolution of the dispersed zone of the tracer. Moreover, the importance of subsurface heterogeneity was also acknowledged, which led to irregular concentration profiles and an enhancement of longitudinal dispersion. Sudicky (1986) investigated the K variability and its influence on macrodispersivity. Theories by Gelhar and Axness (1983) and Dagan (1982, 1984) were adopted to estimate macrodispersivity based on statistical analysis of K. The results showed that the observed longitudinal and horizontal transverse spreading of the tracer increased with time.

The sandy aquifer at the CFB Borden site exhibits only mild heterogeneity, yielding Gaussian forms of tracer patterns after Spatiotemporal evolution. Similar studies have been performed at Cape Cod, Massachusetts (Garabedian et al., 1991; Leblanc et al., 1991), with a variance of ln K equal to 0.24. Sudicky et al. (2010) performed numerical simulations of solute transport at the North Bay landfill site

($\sigma^2_{\ln K} = 1.79$) based on the effective *K* and macrodispersivity estimation. They found that using statistically derived parameters can also result in reliable flow and solute transport models that reasonably predicted the extent and migration rate of the observed contaminant plume.

For more complex aquifer systems with highly heterogeneous *K* fields, accurately representing the spatial heterogeneity of *K* is essential for transport simulations. Since the 2000s, a series of tracer tests (see review by Zheng et al., 2011) have been conducted at the Macrodispersion Experiment (MADE) site in Mississippi, USA. Based on flowmeter measurements (Rehfeldt et al., 1992), *K* values for the shallow aquifer at the MADE site range from 4.1×10^{-7} m/s to 1.5×10^{-2} m/s, with a variance of ln *K* equal to 4.5. Different from the transport studies on mildly to intermediately heterogeneous porous media, the highly heterogeneous nature of this porous medium led to non-Fickian behavior of the tracer plume migration. Non-Fickian dispersion can be understood as the observed temporal evolution of the tracer plume being inconsistent with predictions using the classical ADE, which is built upon Fick's first law of diffusion in homogeneous porous media.

To explain the non-Fickian behavior of the tracer plume at the MADE site, a rate-limited mass transfer process between the mobile domain and the immobile domain of the medium is introduced to solute transport (Harvey & Gorelick, 2000), forming a dual-domain mass transfer (DDMT) transport model, which assumes that transport in the mobile domain is mainly dominated by advection, while transport in the immobile domain is caused by diffusion:

$$-\frac{\partial}{\partial x_i}(\theta_m v_i C_m) + \frac{\partial}{\partial x_i}\left(\theta_m D_{ij}\frac{\partial C_m}{\partial x_i}\right) + q_s C_s = \theta_m \frac{\partial C_m}{\partial t} + \theta_{im}\frac{\partial C_{im}}{\partial t} \tag{1.2}$$

$$\theta_{im}\frac{\partial C_{im}}{\partial t} = \beta(C_m - C_{im}) \tag{1.3}$$

where θ_m and θ_{im} represent the porosity of the mobile and the immobile domain. It is assumed that transport in the mobile domain is mainly dominated by advection, while transport in the immobile domain

is caused by diffusion. Correspondingly, C_m and C_{im} are the concentrations in the mobile and immobile domains. The simulated C_m is compared to the observed concentration. Additionally, β is defined as the mass transfer rate coefficient. The higher β is, the stronger the mass transfer is between the mobile and immobile domains. Therefore, when $\beta \rightarrow \infty$, the connection is instantaneous, and the medium acts as a single-domain advective-dispersive model with the porosity approaching total porosity; when $\beta \rightarrow 0$, the system behaves as a single-domain model with total porosity approaching mobile-zone porosity, θ_m. By definition, the rate-limited mass transfer occurs due to molecular diffusion between the mobile and immobile domains. This process is not considered in the classical ADE, in which the mobile-domain porosity should theoretically be used to represent porosity in Equation (1.1).

Initially, Feehley et al. (2000) conceptualized different heterogeneous *K* fields for simulating plume migration with a tritium tracer. They found that using the DDMT model can better reproduce the observed plume cloud than the classical ADE model, particularly in terms of the longitudinal spreading of the tracer plume. By definition, the rate-limited mass transfer occurs due to molecular diffusion between the mobile and immobile domains. Liu et al. (2007b) demonstrated that the bulk mass transfer rate coefficient is governed by both diffusion and slow advection between the mobile and immobile domains. In subsequent studies, the use of the DDMT model was found to adequately reproduce the primary characteristics of the breakthrough curves (BTCs) when geological features were captured. For example, Liu et al. (2010) incorporated direct-push *K* profiling data to build a zonal *K* distribution in a single well injection-withdrawal test ("MADE-4"). The DDMT model simulated the BTCs that can represent the observed high concentration peak, timing, and late-time tailing more accurately. Zheng et al. (2011) summarized the transport studies at the MADE site and stated that the actual hydrodynamic mixing caused by subsurface heterogeneity is accounted for to a large extent by mass transfer between the mobile and immobile domains.

The DDMT model was employed as an alternative conceptualization approach for predicting transport behavior in complex groundwater flow systems, due to the insufficient characterization of *K* heterogeneity. However, researchers advocated that transport simulations should be based solely on advection and molecular diffusion, provided that subsurface heterogeneity can be accurately characterized. In the MADE studies, Salamon et al. (2007) incorporated the hole effect structure into stochastic geostatistical simulations and found that the longitudinal mass distribution profiles can be reproduced using the classical ADE with a subset of *K* realizations. Ronayne et al. (2010) represented submeter-scale, highly connected channels within a background matrix based on a multivariate Gaussian *K* field from numerous realizations of sequential Gaussian simulations. They found that the classical ADE can reasonably reproduce key characteristics of the observed BTC, including a delayed peak and a low-concentration tail, based on the heterogeneous *K* model. Bianchi et al. (2011b) confirmed that a high degree of flow and transport connectivity exists due to preferential flow paths, a concept initially suggested by Feehley et al. (2000) and Harvey and Gorelick (2000). Dogan et al. (2014) produced 3D *K* fields based on high-resolution direct-push *K* data and found that the classical ADE can adequately reproduce the solute transport behavior. Bianchi and Zheng (2016) developed 3D *K* fields based on the simulation of lithofacies and yielded reasonable concentration fits with observations, concluding that the transport at the MADE site is controlled by a network of high *K* sediments (i.e., preferential flow paths) embedded in a less permeable matrix. In summary, it is possible to accurately predict transport behavior in complex geological environments when subsurface heterogeneity in hydraulic parameters and hydraulic connectivity patterns are well captured.

HT has been demonstrated to be a robust approach for characterizing subsurface heterogeneity. Most studies (e.g., Berg & Illman, 2011b; Cardiff et al., 2013; Hochstetler et al., 2016; Tiedeman & Barrash, 2020) worked on predicting drawdown responses to independent pumping tests, while limited studies have focused on predicting solute/heat tracer transport behavior with HT-derived *K*. Ni et al.

(2009) performed a 2D HT analysis using the geostatistical inverse approach in a synthetic, mildly-heterogeneous aquifer ($\sigma^2_{\ln K}$ = 1.0) and found that the HT model can yield the most reliable predictions of BTC peaks and phases at different monitoring distances. Illman et al. (2012) designed a sandbox experiment of forced-gradient tracer tests within a mildly heterogeneous sandy aquifer ($\sigma^2_{\ln K}$ = 0.87). Model simulations were compared with the observed BTCs in the 2D monitoring network. The temporal moment analysis demonstrated that utilizing HT analysis for mapping heterogeneous *K* can facilitate high-resolution predictions of solute transport. Zhao et al. (2022) extended the work of Illman et al. (2012) and incorporated geologically-interpreted zonal information into HT analysis for improved characterization of *K* heterogeneity. They concluded that the solute transport predictions using HT-estimated *K* can be further refined when reliable borehole lithofacies information is incorporated to form the initial *K* distribution for calibration via HT. Luo et al. (2020) conducted a synthetic solute transport analysis under the regional groundwater flow condition in a basin-scale aquifer-aquitard system. They found that the HT *K* models can provide adequate predictions of solute transport behavior. All these synthetic or laboratory studies employed the classical ADE for simulations.

Only a few field tracer experiments have been conducted to evaluate the HT-estimated *K*. Jimenez et al. (2015) applied a sequential procedure of HT inversion to image the 3D *K* distribution for a mildly heterogeneous porous medium ($\sigma^2_{\ln K}$ = 0.46) and found that the main characteristics of the BTCs can be recovered when applying a suite of longitudinal dispersivity and effective porosity values; however, the tracer BTCs can also be approximated by a simplified effective *K* model with higher dispersivity. Sanchez-León et al. (2016) conducted two forced-gradient tracer tests at a more heterogeneous alluvial aquifer ($\sigma^2_{\ln K}$ = 3.0 based on variogram fitting). One tracer dataset was used to calibrate the parameters for the DDMT model (e.g., mobile-domain porosity). The result showed that using the HT-estimated *K* can successfully predict the BTC for another independent tracer test. Nevertheless, the mobile-domain porosity after calibration (i.e., 0.0195 for sandy materials) was unreasonably low.

Therefore, it is necessary to design a new field tracer test to investigate the usefulness of HT-estimated *K* for transport predictions in the complex geological environment. In this study, a heat tracer test was designed with a dense monitoring system for quantitative comparison of BTCs at high resolution. The heat tracer was used as a proxy (Anderson, 2005) for conservative tracer migration when the advective transport process is dominated in tracer migration (e.g., Doro et al., 2015; Giambastiani et al., 2013; Klepikova et al., 2016; Sarris et al., 2018; Schwede et al., 2014; Somogyvari & Bayer, 2017; Tarnawski et al., 2015; Vandenbohede et al., 2009; Wagner et al., 2014). Ma et al. (2012) found that the temperature depth profiles and temperature peak arrivals simulated by the heat transport model were in reasonable agreement with observations when the *K* distribution was obtained from the calibration of the bromide transport model. However, a temperature-related density effect may exist in heat tracer tests, causing upward flow when heated water moves through the porous medium.

The landmark field experiments (e.g., at the CFB Borden and the MADE site) and their interpretations/simulations offer invaluable experience for us to design a new field tracer test. In previous experimental designs, the quantitative comparison of observed and simulated BTCs was usually performed at limited sampling locations (e.g., the pumping well or different distances in the longitudinal direction). Moreover, depth-specific concentration profiles were rarely recorded. Reproducing BTCs at multiple locations entails the mapping of *K* heterogeneity at a higher resolution and accuracy. For example, Bianchi et al. (2011a) conducted a tracer test (referred to as "MADE-5") under forced-gradient flow conditions and monitored concentrations at two multilevel sampler wells and extraction wells. They found that the BTCs measured at seven depths in the two multilevel wells were radically different from one another in terms of shape, arrival times, and magnitude of the concentration peaks. The numerical simulations, using either the classical ADE model with sequential Gaussian simulations of *K* realizations or the DDMT model with ordinary kriged *K*, failed to reproduce the observed multilevel BTCs. In this thesis study, the tracer test data was recorded at multiple depths and in different migration directions.

Furthermore, porosity was previously determined on the basis of expert knowledge of the soil or calibration of tracer data. Hence, the most favorable porosity value was assigned to the medium, ensuring the optimal model fit with observed BTCs. In addition, effective porosity varies significantly with total porosity for finer-grained materials. In the transport studies at the MADE site, the value of total porosity was usually assigned between 0.30 and 0.35 (e.g., Bianchi et al., 2011a; Bianchi & Zheng, 2016; Feehley et al., 2000; Liu et al., 2010; Ronayne et al., 2010). However, different effective porosity (i.e., mobile-domain porosity) values were determined. Ronayne et al. (2010) assigned 75% of the total porosity to the mobile domain, yielding a mobile-domain porosity of 0.225. For the same tracer test ("MADE-4"), the mobile-domain porosity was assigned to be 0.044 by Liu et al. (2010). Feehley et al. (2000) suggested that a mobile-domain porosity should be 1/8 of the total porosity. Bianchi and Zheng (2016) confirmed the validity of this ratio by assuming a mobile-to-total porosity ratio equal to the volumetric fraction of the defined lithofacies of the "highly conductive gravel", which corresponds to preferential flow paths at the site. In reality, the effective-to-total porosity ratio for sands and gravels could be over 90%, while the mobile-domain porosity for silt is smaller than 0.1. Therefore, the spatial heterogeneity of effective porosity should be accounted for in the numerical simulation using the classical ADE, especially for highly heterogeneous porous media where the soils vary from gravel, sand, silt, to clay. In this thesis study, borehole NMR logging was employed to estimate the total porosity and effective porosity profiles along boreholes, which were used to construct more reliable spatial porosity models.

2 Research Topics and Objectives

By integrating borehole geophysical logging with HT analysis, the main goal of this thesis study aims to investigate the feasibility of improving groundwater flow and solute/heat transport prediction performance through refining the representation of hydraulic parameters (e.g., K and porosity). HT analysis may yield smooth K distributions, especially when the observation density is low. To reduce the smoothness of K tomograms, previous efforts at the NCRS used K distributions derived using other sources of data as initial guesses of K representations for the geostatistical inverse modelling of HT analysis, which has been demonstrated to be an effective approach to improve the accuracy and resolution of K tomograms (Zhao and Illman, 2017, 2018, 2022a, 2023; Luo et al., 2023). Therefore, this thesis study follows such a framework and extends it to integrate borehole geophysical logging data with HT analysis for improved subsurface heterogeneity characterization.

Furthermore, it was a formidable challenge to accurately predict tracer transport behavior in complex aquifer systems using conventional approaches to characterize subsurface heterogeneity. Since the subsurface heterogeneity in K and S_s as well as the hydraulic connectivity of the aquifer systems can be better captured via HT analysis, this thesis study conducted a heat tracer test in a highly heterogeneous glaciofluvial deposit to investigate whether the transport model using HT to map K heterogeneity is capable of adequately reproducing the observed tracer transport behavior.

To answer these questions, this thesis study is divided into five independent research topics, with each topic constituting of individual chapters addressing progressive research objectives. In this chapter, the specific objectives of each study are described. Results from five studies are presented and discussed individually in Chapters 3-7. Chapter 8 lists summaries and conclusions from each study and provides some workable recommendations for future research.

Study I: Hydraulic Tomography Estimates Improved by Zonal Information From the Clustering of Geophysical Survey Data

Due to the non-uniqueness of the petrophysical relationship between geophysical logging data and K, reliable K distributions cannot be readily derived from geophysical logging. Therefore, this study attempted to extract the hydrostratigraphic information from multiple geophysical logs. Afterwards, layer boundaries determined from geophysical logs can be interpolated and extrapolated to construct zone geometry. On the other hand, the zonal K estimates were treated as internally homogeneous K values, which can be estimated using a deterministic inversion of head data from pumping tests or using the local K measurements (e.g., from permeameter results). Therefore, geological zonation models of K were constructed to represent the layered porous media, which can be incorporated into HT analysis as the initial guess for stochastic and geostatistical inverse modelling.

One specific issue is how to effectively determine layer boundaries and match the geological zones across different boreholes. Previous studies (Zhao & Illman, 2017, 2018) used visual interpretation to determine layer boundaries along boreholes from borehole geology logs. This requires great effort, and the accuracy of zone geometry is highly dependent on the experience of the practitioners. Automatic determination of layer boundaries usually employs clustering. However, conventional clustering approaches cannot yield improved zone geometry compared to manual interpretation (see Zhao and Illman (2022a)). In this study, a spatial conditioning term was proposed to improve the clustering performance. In addition, this study also investigated the use of multiple types of geophysical logging data to reduce the non-uniqueness of petrophysical relationships. Therefore, comparative cases of HT analysis were designed to explore the most effective framework to integrate borehole geophysical logging with HT analysis.

In Study I, a numerical experiment was conducted based on a real sandbox consisting of a highly heterogeneous layered aquifer-aquitard system. Six vertical scanlines of electrical conductivity and gamma ray logging data were generated to mimic multiple borehole geophysical logging surveys. First, the clustering based on geophysical data was performed, and the clustering validity was evaluated. Then, comparative cases of HT analyses were evaluated, in which the geostatistical inversion was initialized with various K distributions using zonal information from geophysical logs. After model calibration, the representativeness of these K tomograms was assessed through predicting the drawdown responses from multiple independent pumping tests and reproducing the true concentration responses from two conservative tracer tests. Through numerical studies, it was demonstrated that the developed framework efficiently facilitates the integration of disparate borehole geophysical survey data with HT analysis. It provides important insights into fusing non-redundant geophysical and hydraulic datasets to better characterize subsurface heterogeneity, and it is of particular importance to geophysicists and hydrogeologists for guidance of operations at various sites.

The innovation points of Study I include: (1) a spatial conditioning term for clustering of multiple geophysical logging data was proposed to construct more reliable zonation models; (2) a framework was established in which borehole geophysical logs were utilized to improve HT performance; and (3) this was the first study to demonstrate the effectiveness of integrating geophysical logging data with HT analysis for improved predictions of groundwater flow and solute transport.

Study II: Comparison of NMR-Derived Hydraulic Conductivity with Various Hydraulic Testing Methods

Starting from Study II, this thesis focused on field-scale experiments and their analyses. All field experiments of this thesis were conducted at the North Campus Research Site (NCRS) located on the

University of Waterloo (UW) campus, which is underlain by a highly heterogeneous glaciofluvial aquifer system (Alexander et al., 2010; Berg & Illman, 2011b).

Study I showed that accurate zonal information and *K* estimation are essential to building a most optimal initial guess of the *K* distribution for HT analysis. Hence, a more advanced geophysical logging tool, called nuclear magnetic resonance (NMR) logging, was utilized in Study II. NMR logging can characterize subsurface heterogeneity along boreholes at a high resolution. Meanwhile, it can yield direct *K* estimates from NMR signals, as well as effective porosity measurements. Therefore, NMR logging is anticipated to provide a reliable initial guess of *K* distributions for HT analysis.

NMR logging was originally used in the petroleum industry for characterization of bedrock. Over the last decade, the application of NMR logging has been extended for characterization of unconsolidated materials, with several NMR petrophysical relationships proposed to estimate downhole *K* profiles. By comparing with co-located *K* measurements from hydraulic testing methods, previous studies showed that the model constants for these petrophysical relationships were stable at some sites; however, the model constants has been found to vary over one order of magnitude for another study (see compilation by Kendrick et al. (2023)). If stable model constants representative to specific deposits cannot be determined, then site-specific optimization is imperative to ensure the accuracy of NMR *K* estimates, limiting the utility of NMR logging for subsurface heterogeneity characterization.

Therefore, Study II first calibrated the model constants of four common NMR petrophysical relationships using permeameter measurements at the NCRS. Then, the utility of NMR logs to estimate *K* within this highly heterogeneous glaciofluvial deposit was assessed by comparing them with three types of co-located hydraulic measurements, including permeameter, multi-level slug, and direct-push hydraulic profiling tool logging tests. The calibrated model constants were listed and compared to literature-based constants from various geological deposits.

The novel aspects of Study II include: (1) the NMR-derived *K* was site-specifically calibrated for highly heterogeneous glacial deposits, with the optimized *K* evaluated using multiple hydraulic testing methods. (2) A suite of NMR model constants was proposed, which can be added to the current NMR model database for use in similar geological environments.

Study III: Characterizing Spatial Heterogeneity of Hydraulic Conductivity Using Borehole NMR in a Complex Groundwater Flow System

Previous studies (e.g., Zhao et al., 2023; Zhao & Illman, 2017, 2018) indicated that the initial *K* distributions from other data sources cannot adequately reproduce drawdown responses in pumping tests when model calibration using HT analysis is not performed. Therefore, after evaluating the *K* estimation from NMR logging along boreholes in Study II, the main research objective of Study III is to explore the potential of NMR logging to derive spatial *K* distributions, which would enable its utilization in numerical groundwater flow and transport models.

A critical step from downhole *K* estimates to spatial distributions of *K* fields is the spatial interpolation of borehole *K* data. Multiple spatial interpolation approaches were used in this study to generate spatial heterogeneity of *K* between boreholes. These approaches can be classified to three categories. The first type is the geostatistical interpolation approach. Ordinary kriging and indicator kriging estimation were used for spatial interpolation, in which the assumptions of continuous distributions or discrete facies of *K* were embedded. The second type is the geological zonation approach, in which the layered porous media at the NCRS were treated as internally homogeneous zones with sharp boundaries. Zone geometry along boreholes was delineated using either visual interpretation (see the previous results by Zhao and Illman (2017)) or the *k*-means clustering with a spatial conditioning term. The third type is a hybrid approach in which the layered porous medium was treated to be a multi-level heterogeneity model. The interlayer heterogeneity was characterized by the spatially associated

clustering, while the intralayer heterogeneity was generated by the ordinary kriging of the *K* data clustered into each zone. The representativeness of these NMR-derived spatial *K* models was evaluated by reproducing a permeameter-based *K* profile at an unsampled location and by comparing the numerically simulated drawdown responses with field observations of ten pumping tests.

Novel aspects of Study III include (1) the first use of NMR logging to model high-resolution 3D *K* distribution in a highly heterogeneous environment; (2) the depiction of inter-well *K* variations by modelling multi-level heterogeneity patterns; (3) the evaluation of model performance to predict hydraulic responses of multiple pumping tests through a comparative case study. Our findings suggest that borehole NMR can derive reliable 3D *K* fields, representative of the complex groundwater flow system at the NCRS.

Study IV: Integrating NMR Logging with Three-Dimensional Steady-State Hydraulic Tomography for Improved Mapping of Hydraulic Conductivity Distribution

After deriving 3D *K* fields from NMR logging, Study IV examined the HT performance integrated with NMR logging for improved mapping of *K* heterogeneity in a complex aquifer-aquitard system at the NCRS. When no prior information about subsurface heterogeneity is available, the geostatistical inverse modelling for HT analysis typically initializes with a homogeneous *K* field. Here, the NMR-derived spatial *K* models in Study III were treated as the initial *K* guess for the geostatistical inverse modelling and were calibrated via HT analysis. As a result, three classes of spatial *K* models were constructed, including the NMR-derived *K* models, the HT-estimated *K* models, and the integrated models by the joint use of HT and NMR logging.

Another special concern of this study is the HT performance with an insufficient dataset for calibration. Previous studies at the NCRS utilized a dense monitoring network to record drawdown

responses at depth-specific observation ports. For example, a steady-state HT calibration dataset used by Luo et al. (2023) and Zhao and Illman (2017) contained 195 head data points from seven pumping tests. Their HT results without incorporating prior information from other data sources can provide accurate predictions of drawdown responses from independent pumping tests. However, such a "dense" monitoring network may not always be available for HT applications in industry. Therefore, a limited HT dataset was designed in this study, in which only the central pumping well and four observation wells were included in the monitoring network, and head data from four pumping tests were used for model calibration. This limited HT dataset was intended to yield a smooth K tomogram. As a comparison, a full HT dataset was also included, in which eight pumping tests with all available head records were used for model calibration. The representativeness of these K tomograms from HT models and integrated models was validated by comparing the simulated with the observed drawdown from five independent pumping tests.

The innovative aspects of this study include: (1) the first field application of integrating borehole NMR data with HT analysis for improved mapping of K heterogeneity; (2) examining various spatially interpolated K models as initial K guesses for improved HT results; and (3) exploring the usefulness of integrating NMR logging with the HT analysis for reliable K estimates using a more cost-effective and practical design of the HT survey.

Study V: Three-Dimensional Analysis of Heat Tracer Transport with Hydraulic Conductivity Distributions from Various Approaches in a Complex Aquifer System

Over decades of research, HT has been demonstrated to be a robust approach to characterize the subsurface heterogeneity. The most common method to evaluate the effectiveness of HT results has been to predict drawdown responses from independent pumping tests (e.g., Berg & Illman, 2011b; Cardiff et

al., 2013; Hochestetler et al., 2016; Tiedeman & Barrash, 2020). However, to date, HT's ability to predict solute/heat tracer transport behavior for complex aquifer systems has not been rigorously assessed.

In terms of drawdown predictions, the pressure signals measured at observation points are propagated through the space from the source location (i.e., pumping or injection). Therefore, non-unique representations of 3D *K* fields can exist, yielding comparable overall drawdown fits with field observations. In comparison, the conservative solute/heat tracer migrates to the observation points along the dominant advective paths, necessitating more accurate imaging of the preferential flow pathways. Therefore, the accurate prediction of solute/heat tracer transport behavior relies heavily on the unique representations of *K* tomograms.

In Study V, a new tracer test was conducted, in which chilled water was used as the heat tracer and injected into the subsurface at the NCRS. The temperature responses were recorded in a dense monitoring network. The primary research objective was to reproduce the heat tracer transport behavior based on *K* distributions from HT analysis. To highlight the HT performance, different subsurface characterization approaches were used to map *K* heterogeneity at the NCRS. First, a homogeneous and anisotropic *K* was used to represent the entire 3D *K* field. Additionally, the geostatistical interpolation approach using ordinary kriging estimation and the geological zonation approach were employed to generate the spatial heterogeneity patterns, with *K* measurements or estimates from permeameter tests, hydraulic profiling tool (HPT) tests, and pumping tests. The last approach is HT analysis, which uses the geostatistical inverse modelling to generate highly parameterized 3D *K* distributions.

The determination of spatial porosity heterogeneity was usually neglected in previous studies (e.g., see a series of solute tracer transport studies at the MADE site (Zheng et al., 2011)). In this thesis study, NMR logging provides measurements of mobile-domain water content, which was converted to the effective porosity for the advective transport process in the saturated zone. Therefore, another research objective was to assess the importance of mapping the spatial heterogeneity of effective porosity to heat

transport predictions. Here, the spatial heterogeneity of effective porosity was determined by populating a 19-layer geological zonation model with NMR-derived effective porosity measurements.

The simulated temperature breakthrough curves were compared with field observations. The temporal moment analysis was employed to quantify the model fits. Lastly, the sensitivities of heat transport behavior to transport and thermal parameters, including porosity, dispersivity, and thermal conductivity, were investigated.

Innovative aspects of this study include: (1) a heat tracer test is performed by injecting chilled groundwater, and a dense monitoring network is used to track plume migration and spreading; (2) NMR logging allows for the estimation of a heterogeneous effective porosity field, which was previously hard to quantify using other in situ testing methods; and (3) the use of HT analysis provides reliable K estimates for reasonable predictions of the spatiotemporal evolution of the heat tracer plume based on the classical advection-dispersion equation.

3 Hydraulic Tomography Estimates Improved by Zonal Information from the Clustering of Geophysical Survey Data

Citation:

Wang, C., & Illman, W. A. (2023). Hydraulic Tomography Estimates Improved by Zonal Information From the Clustering of Geophysical Survey Data. *Water Resources Research*, *59*, e2023WR035191. https://doi.org/10.1029/2023WR035191

3.1 Overview

Hydraulic Tomography (HT) has been demonstrated as a robust approach to characterize subsurface heterogeneity through the inverse modeling of multiple pumping data. However, smooth or even erroneous tomograms can result when insufficient observations are involved in the inversion. In this study, the feasibility of integrating geophysical surveys into HT analysis is investigated. First, *k*-means clustering is utilized to extract zonal information from borehole geophysical logs, and a new type of spatial constraints containing geological knowledge is proposed to obtain improved hydrostratigraphic boundaries along boreholes. Next, zonation models are constructed by applying clustering-based zone geometry and populating zonal estimates of hydraulic conductivity (K) from analyzing pumping data. Afterwards, zonation models are treated as the initial guess of spatial variability in the geostatistical inversion of HT analysis. Additionally, local K measurements can be utilized to further improve HT estimates. Comparative cases of HT analyses are designed for a numerical sandbox experiment to highlight the HT performance integrated with geophysical surveys, in which the geostatistical inversion is initialized with: (1) a homogeneous K field; (2) zonation models built by the clustering of disparate

geophysical surveys with/without spatial constraints; and (3) zonation improved by incorporating local K measurements. Based on ln K field comparisons and validation through predictions of drawdowns and tracer plume migration from independent tests not used in the calibration effort, we find that integration of geophysical surveys into HT analysis by clustering with spatial constraints is demonstrated as an effective approach, and local K measurements can further improve HT estimates.

3.2 Introduction

Investigations of groundwater flow and solute transport rely heavily on the accurate knowledge of hydraulic properties (e.g., hydraulic conductivity (K) and specific storage (S_s)). Among various techniques to characterize the subsurface for hydraulic parameters, pumping tests provide drawdown responses of large-scale aquifer systems, and K and S_s can be estimated through the interpretation of hydraulic head data monitored at observation ports. Traditionally, analytical solutions (Hantush, 1960; Neuman, 1972; Theis, 1935) have been used to interpret pumping tests with an assumption that the subsurface is homogeneous. However, the estimates of hydraulic parameters from such traditional analyses are insufficient in predicting accurate drawdown behaviors and characterizing the heterogeneous nature of the subsurface (Wu et al., 2005).

Hydraulic tomography (HT) was proposed and developed over the past two decades. It relies on the joint inverse modeling of hydraulic head data from multiple pumping tests. Water injection/extraction can be conducted in targeted aquifers with installation of monitoring networks such that drawdown signals are observed over large volumes. Therefore, HT is a cost-effective way to collect non-redundant datasets about subsurface heterogeneity (Hou et al., 2023; Zha et al., 2015). To interpret the heterogeneity pattern of aquifer-aquitard systems, geostatistical and stochastic inverse methods were developed (Gutjahr & Wilson, 1989; Harvey & Gorelick, 1995; Hoeksema & Kitanidis, 1984, 1989; Kitanidis, 1995; Kitanidis & Vomvoris, 1983; Rubin & Dagan, 1987; T.-C. J. C. J. Yeh et al., 1996; Yeh et al., 1995), in which geostatistical parameters of aquifer systems were introduced into the inversion, and the most likely

spatial distribution of hydraulic properties and their uncertainty are estimated for accurate prediction of drawdown responses. To date, HT has been developed as a mature technique to accurately map spatial heterogeneity of hydraulic properties through synthetic studies (Cardiff & Barrash, 2011; Gottlieb & Dietrich, 1995; Luo et al., 2020; Tso et al., 2015; Yeh & Liu, 2000; Zhu & Yeh, 2005, 2006), laboratory experiments (Berg & Illman, 2011a; Illman et al., 2007, 2008, 2010, 2012; Liu et al., 2002; Liu et al., 2007a; Luo et al., 2017; Zhao & Illman, 2021), field studies (Berg & Illman, 2011b, 2013, 2015; Bohling et al., 2007; Brauchler et al., 2013; Cardiff et al., 2013; Huang et al., 2011; Illman et al., 2009; Klepikova et al., 2020; Tiedeman & Barrash, 2020; Zha et al., 2016; Zhao & Illman, 2017, 2018, 2022b) and basin-scale aquifer characterization (Daranond et al., 2020; Liu et al., 2020a; Liu et al., 2020b; Luo et al., 2022; Xia et al., 2023; Yeh et al., 2009).

Among various geostatistical inversion codes, Successive Linear Estimator (SLE) (Yeh & Liu, 2000; see also Xiang et al., 2009; Zhu & Yeh, 2005) has been extensively applied in HT analysis, and its effectiveness and robustness in interpreting K and S_s fields have been sufficiently validated. The SLE code originates from the traditional cokriging method in geostatistics. Cokriging is a linear estimator in which the correlation between primary (e.g., hydraulic parameters) and secondary variables (e.g., field observations) is assumed to be linear. Due to the fact that the relationship between hydraulic parameters and observations are essentially nonlinear, traditional cokriging cannot make full use of observed data. This nonlinear issue can be addressed in SLE through iteratively updating the estimated parameter fields and their conditional covariances during the inversion process such that the difference between observed and simulated head can be utilized iteratively to estimate hydraulic parameters.

Despite the robust performance of the geostatistical inversion approach, challenges remain. A main issue on subsurface heterogeneity characterization encountered in the industry is that hydraulic response data from boreholes and outcrops are always insufficient. Cardiff et al. (2013) found that significant information about subsurface heterogeneity is progressively lost after removing pumping

and/or observation intervals in a field HT survey. It indicates that the deficiency of pumping data can lead to smooth tomograms (i.e., spatial distribution of hydraulic parameters). Yeh et al. (2008a) proposed a vision that fusion of different types of tomographic surveys with data from disparate sources of excitations is a future direction for improved basin-scale subsurface characterization. Accordingly, researchers consider fusion (Bohling & Butler, 2010; Yeh et al., 2008b) of hydraulic, geological and geophysical data to improve HT performance. For example, Zha et al. (2017) treated the spatial variability of K as two-scale heterogeneity using a nested covariance function whereby site-specific qualitative stratigraphy was incorporated to improve the conditional means and covariances of primary variables (i.e., hydraulic parameters) for HT analysis. Results from both synthetic study and field application reveal that hydraulic estimates are improved in this manner. Different from the effort by Zha et al. (2017), Zhao et al. (2016) constructed geological zonation models by extracting stratigraphic information from borehole geology logs, which characterizes the subsurface as various zones prior to the inversion. Then, geological zonation models are utilized as the initial model for geostatistical inverse modeling for HT analysis. The results of Zhao et al. (2016) and subsequent studies (Luo et al., 2017, 2020; Zhao et al., 2022; Zhao & Illman, 2017, 2018) show that the resolution and accuracy of HT estimates are significantly improved after incorporating accurate geological information especially in the area with a sparse drawdown monitoring network, whereas inaccurate geological information is found to produce erroneous characterization of aquifer hydraulic properties.

Geophysical surveys are also widely used to characterize subsurface heterogeneity. Conventional geophysical investigation relies on the petrophysical analysis relating hydraulic parameters (e.g., K, porosity, water content) with geophysical parameters (e.g., electrical conductivity (EC), dielectric permittivity) through empirical relationships. A widely-used example is Archie's law (Archie, 1942), which can relate water content or effective porosity in saturated soil with EC. A drawback to Archie's law is that it only regards the pore-fluid part of EC, while the pore-surface part of EC, which often dominates

in clay-rich deposits, is not considered. In addition to EC, other empirical relationships also have been established between various geophysical and hydraulic parameters. For example, Topp et al. (1980) related dielectric permittivity to porosity for the Ground Penetrating Radar (GPR) method, while Raymer et al. (1980) associated P-wave velocity with porosity for the seismic method. As there is no perfect correlation existing between hydraulic and geophysical parameters, such petrophysical analysis is always site-specific and highly non-unique (Yeh et al., 2002).

Furthermore, few petrophysical relationships are built with *K*, which directly represents the ability of medium to transmit water. Pollock and Cirpka (2008, 2010, 2012) used electrical potential perturbations to invert *K* via time-lapse electrical resistivity tomography (ERT). In a conservative tracer test, EC levels are significantly increased due to increasing concentration. Therefore, a fully-coupled hydrogeophysical inversion was introduced, in which *K* field is estimated from inversion of the mean arrival time of electrical potential perturbations and hydraulic heads. Some new geophysical techniques have been developed during the past decade. For example, the Nuclear Magnetic Resonance (NMR) technique is promising in estimating *K* using the NMR-derived porosity and relaxation time (Walsh et al., 2013).

As geophysical surveys are usually cost-effective and bring less-to-no disturbance to the subsurface, disparate geophysical survey data can be collectively analyzed to reduce non-uniqueness of petrophysical relationships. The zonation method (Dietrich & Tronicke, 2009; Doetsch et al., 2010; Dumont et al., 2018; Gueting et al., 2015; Paasche et al., 2010; Paradis et al., 2015; Zhang et al., 2022; Zhao & Illman, 2022a) is extensively adopted by geophysicists to describe aquifer characteristics combining disparate geophysical survey data, in which data from multiple surveys are partitioned into hydrostratigraphic zones by clustering. Meanwhile, zonal hydraulic parameters can either be computed by applying petrophysical relationships or populated by averaging borehole hydraulic measurements from hydraulic characterization techniques (e.g., grain-size correlation, permeameter tests, slug tests, etc.).

Various clustering algorithms have been applied to partition hydrostratigraphic zones from geophysical survey data. For example, *k*-means clustering (MacQueen, 1967) is a commonly applied non-parametric partitioning algorithm, which employs a centroid vector as a prototype for each cluster. This algorithm seeks for the minimum sum of feature-to-centroid distances by iteratively assigning features to the closest centroids and generating new centroids based on the current assignments for each cluster. In addition to *k*-means clustering, hierarchical clustering (Dumont et al., 2018), fuzzy-*c* means clustering (Paradis et al., 2015; Zhao & Illman, 2022a), and mixture-of-gaussian clustering (Doetsch et al., 2010) have been implemented for clustering of geophysical parameters.

In spite of the availability of a wide variety of clustering algorithms, the key issue is to understand the inductive principles involved in algorithms (Estivill-Castro, 2002) such that appropriate algorithms can be utilized for input datasets. For example, *k*-means clustering tends to construct clustered features forming ellipsoidal shapes. Moreover, outliers are inevitably present in clustering results, and the number increases with the increasing heterogeneity of subsurface. The main reason is that spatial adjacency of samples is neglected in clustering such that samples from disparate regions are partitioned into the same cluster.

Spatial correlation is an essential subsurface characteristic, which assumes that the closer the two locations are, the higher the similarity of data at these two locations. Pham (2001) proposed spatial models for fuzzy *c*-means clustering, in which a spatial penalty term was added to yield spatially smooth membership functions. This study was followed by Paasche et al. (2010), in which spatial constraints were applied to objectively outline the dominant subsurface units.

Though many efforts were made to improve the ability of utilizing geophysical survey data to interpret subsurface heterogeneity, most of their performances on the prediction of groundwater flow and solute transport have not been quantitatively validated. From our knowledge, the geostatistical inversion of data from a HT survey can yield robust hydraulic estimates among various in-situ aquifer

characterization techniques (e.g., Berg & Illman, 2011a; Illman et al., 2010). However, when the number of pumping tests and monitoring data are sparse, smooth estimated parameters can result. Hence, our main goal is to integrate non-redundant geophysical survey data with HT analysis. Referring to the workflow of Zhao et al. (2016), zonation models from geological knowledge can be utilized as the initial *K* distribution for geostatistical inverse modeling to successfully improve HT estimates. This enlightens us to utilize geophysical survey data to improve HT performance in a similar manner. In this study, the first purpose is to establish a general framework to integrate geophysical survey data into HT analysis. Second, a new type of spatial constraints is proposed to alleviate the number of outliers, in which geological knowledge can be incorporated. At last, the importance of local *K* measurements to improve HT estimates in such a framework is investigated. To examine these issues, a numerical study based on a real sandbox consisting of a highly-heterogeneous aquifer-aquitard system is constructed such that groundwater flow and solute transport patterns of different cases are known perfectly. Comparative cases are designed to highlight the improved performance of HT analysis when geophysical data are integrated, in which the geostatistical inversion is initialized with: (1) a homogeneous *K* field; (2) zonation models built by the clustering of disparate geophysical surveys with/without spatial constraints; and (3) zonation models improved by incorporating local *K* measurements. Through synthetic studies, we demonstrate that the developed framework efficiently facilitates the integration of disparate borehole geophysical survey data with HT analysis. It provides important insights on fusing non-redundant geophysical and hydraulic datasets to better characterize subsurface heterogeneity, and it is of particular importance to geophysicists and hydrogeologists for guidance of operations at various sites.

3.3 Methodology

3.3.1 Clustering with Spatial Constraints

Consider a dataset $\boldsymbol{D}_1 = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_m\}$ containing m independent geophysical survey data. For the i^{th} observation ($1 \le i \le m$), $\boldsymbol{X}_i = \left(X_{i,1}; X_{i,2}; \dots; X_{i,n}\right)$ is an n-dimensional feature vector denoting n types of structural information from geophysical surveys (e.g., borehole measurements including EC, gamma ray (GR) counts, P-wave velocity, etc.). The structural information is normalized to form the sample space $\boldsymbol{D}_1^* = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m\}$ for clustering. For the j^{th} structural information ($1 \le j \le n$) of the i^{th} measurement, the normalized structural feature $x_{i,j}$ is:

$$x_{i,j} = \frac{X_{i,j} - \overline{X_j}}{\sigma_{X_j}} \tag{3.1}$$

where $\overline{X_j}$ and σ_{X_j} are the mean and standard deviation of the j^{th} structural information, respectively. We assume that $\boldsymbol{D}_1^*$ is partitioned into k clusters, while setting $\boldsymbol{C}_r$ as the r^{th} cluster ($1 \le r \le k$) with $\boldsymbol{\mu}_r$ representing the mean vector of this cluster. A squared Euclidean distance is assigned as the distance metric, and the objective function for k-means clustering with no spatial penalty term is:

$$E = \sum_{r=1}^{k} \sum_{\boldsymbol{x}_i \in \boldsymbol{C}_r} \|\boldsymbol{x}_i - \boldsymbol{\mu}_r\|_2^2 \tag{3.2}$$

where E is the error sum of squares representing the degree of dissimilarity within a cluster among all samples, and $\| \|_2$ is the Euclidean norm. Each sample is assigned to its closest mean vector (i.e., the centroid), and consequently the minimum E is determined. Then, mean vectors $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_k\}$ for all clusters are recalculated for the next iteration based on the current assignments of samples. Equation (3.2) is iteratively solved until convergence criteria are met (i.e., E is no longer reduced) to produce optimal clustering results. Essentially, Equation (3.2) is the classical representation of k-means clustering.

To reduce outliers and incorporate the spatial adjacency nature of samples into clustering, a new type of spatial constraints is utilized. The concept of this spatial term is adopted from the Inverse Distance Weighted (IDW) interpolation method (Lu & Wong, 2008) used in geography, in which the inverse distance is used to assign larger weights to closer observation points for prediction. For the i^{th} observation, $\boldsymbol{Y_i} = \left(Y_{i,1}; Y_{i,2}; Y_{i,3}\right)$ records three-dimensional coordinates at observation locations. A dataset of the corresponding spatial coordinates of structural information is defined as $\boldsymbol{D}_2 = \{\boldsymbol{Y}_1, \boldsymbol{X}_2, \dots, \boldsymbol{Y}_m\}$, and this dataset is incorporated into the input survey data $\boldsymbol{D}_1$ as additional dimensions. When the distance from a sampling location to a cluster centroid is closer, the calculated E is smaller. Therefore, this point is more likely to be assigned to a cluster in its neighborhood.

Raw spatial coordinates are normalized before incorporated into clustering to render structural features and spatial coordinates comparable for computations. For the l^{th} spatial coordinate of i^{th} measurement, the normalized coordinates $y_{i,l}$ is:

$$y_{i,l} = \frac{Y_{i,l}}{\lambda_l} \tag{3.3}$$

where λ_l is the average dimension of stratigraphic units in the l^{th} spatial direction, which can be inferred from site-specific geology, e.g., preliminary analysis of geophysical imaging, core sample analysis, in-situ geological investigations, etc. For the new input data set $\boldsymbol{D}_0 = \{(\boldsymbol{X}_1, \boldsymbol{Y}_1), (\boldsymbol{X}_2, \boldsymbol{Y}_2), \dots, (\boldsymbol{X}_m, \boldsymbol{Y}_m)\}$ containing spatial information, the normalized sample space is $\boldsymbol{D}_0^* = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2), \dots, (\boldsymbol{x}_m, \boldsymbol{y}_m)\}$. For the r^{th} cluster $\boldsymbol{C}_r$, $(\boldsymbol{\mu}_r, \boldsymbol{\nu}_r)$ is the mean vector reflecting both structural and spatial features. Meanwhile, a spatial penalty term is added to E for k-means clustering with spatial constraints:

$$E^* = \sum_{r=1}^{k} \sum_{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \boldsymbol{C}_r} \|\boldsymbol{x}_i - \boldsymbol{\mu}_r\|_2^2 + \beta \sum_{r=1}^{k} \sum_{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \boldsymbol{C}_r} \|\boldsymbol{y}_i - \boldsymbol{\nu}_r\|_2^2 \tag{3.4}$$

where E^* is the error sum of squares with spatial features. Factor β controls the trade-off between distances in structural and spatial terms. Thus far, a unit distance in the spatial term represents that the point-to-centroid distance is an average dimension of stratigraphic units, while a unit distance in the structural term is equivalent to a standard deviation among geophysical survey observations. The impact of spatial constraints is increased when fewer data types of structural information are integrated, while its impact is reduced when multiple geophysical survey logs with non-redundant information are incorporated into clustering. Besides, if there is no preference for weight assignments of the two terms, a default value of 1 can be applied for the trade-off factor β. However, the assignment of the β value is bound to geological knowledge, which can be slightly increased when the medium is well-stratified, while it can be decreased when the medium is poorly stratified. Nevertheless, note that for β=0 this expression (3.4) reduces to Equation (3.2), and the spatial penalty term does not affect the calculation of E^*.

3.3.2 Geostatistical Inverse Modeling

As a geostatistical inverse method, the Simultaneous Successive Linear Estimator (SimSLE, Xiang et al., 2009) is used to yield highly-parameterized K and S_s estimates given observations from head response data during HT surveys. To circumvent the non-linearity between hydraulic parameters (e.g., ln K) and head observations, SimSLE adopts a linear relationship iteratively to condition HT estimates in which weighting coefficients for observations are introduced to calibrate HT estimates. Weighting coefficients are determined by involving spatial covariance functions into the computation. Conditional covariances and weighting coefficients are updated successively such that observations are fully utilized during the inversion, which is a major difference from other geostatistical inverse modeling approaches.

The SimSLE algorithm initializes through a classical cokriging method. Prior information is provided through a covariance function and assigning the initial distribution of hydraulic parameters. On one hand, the initial covariance function of ln K is described by assuming an exponential model with

given estimates of variance and correlation lengths. On the other hand, careful selection of the initial hydraulic parameter field works for both the establishment of sensitivity matrix using an adjoint state sensitivity method (see Zhu & Yeh, 2005 for details) and the initial guess of HT estimates for model calibration. Our target is to integrate geophysical sources of prior information into the HT model by providing an improved initial guess of hydraulic parameter field. A more detailed description of the SimSLE algorithm can be found in Xiang et al. (2009).

3.3.3 Framework to Improve HT Estimates

A framework is established to employ the clustered geophysical survey data into HT analysis by constructing zonation models as the initial distribution of hydraulic parameters. A flowchart of this framework is shown in Figure 3.1. The main steps are as follows: (1) first, multiple geophysical survey logs are normalized and then incorporated into the sample space for *k*-means clustering. Spatial terms are determined for the simulation domain and incorporated into clustering to provide more constraints. During the clustering procedure, hydrostratigraphic zones (*k*) are partitioned, which contain zonal information from geophysical survey logs; (2) next, a commercial software Leapfrog Geo (Seequent Ltd., 2021) is used to interpolate and extrapolate layer boundaries among boreholes to generate zone geometry. A volumetric algorithm called the Fast Radial Basis Function (Carr et al., 2001) is applied in the Leapfrog Geo software to produce first-pass surfaces and continuous smooth models utilizing a global neighborhood searching algorithm from categorical data along boreholes; and (3) geophysical measurements are converted to *K* values if petrophysical relationships exist (e.g., NMR). Alternatively, if local *K* measurements/estimates from grain size analyses, borehole hydraulic tests (e.g., permeameter, slug, and/or single-holes tests or from the Hydraulic Profiling Tool (McCall & Christy, 2020; Zhao et al., 2023) are available, the geometric mean of collected *K* measurements in each zone can be calculated and utilized as an initial realization of the effective mean value in this zone; (4) following the generation of zone geometry, zonal *K* values are optimized using head response data from pumping tests by coupling a

model-independent parameter estimation code PEST (Doherty, 2015) with the groundwater flow simulator in VSAFT3 (Variably Saturated Flow and Transport 3-D model, Yeh et al., 1993). If no local K measurements are available, an unconditional effective mean of K for the entire simulation domain is applied. Thus far, zonation models are constructed by delineating zone geometry from the clustering of geophysical survey data and filling zonal mean values of hydraulic parameters by analyzing pumping data; and (5) zonation models are assigned as the initial field of hydraulic parameters in the calibration of HT analysis through geostatistical inverse modeling. Then, SimSLE will update the conditional effective parameters and reduce its uncertainty at each iteration to yield calibrated HT estimates. The inversion process stops when either the biggest head misfit between simulated and observed head at all observations locations or the change of ln K variance during two successive iterations is smaller than prescribed tolerances, and the optimal HT estimates are obtained. The estimated K and S_s fields (if necessary) can then be utilized in various numerical models for groundwater flow and contaminant transport predictions.

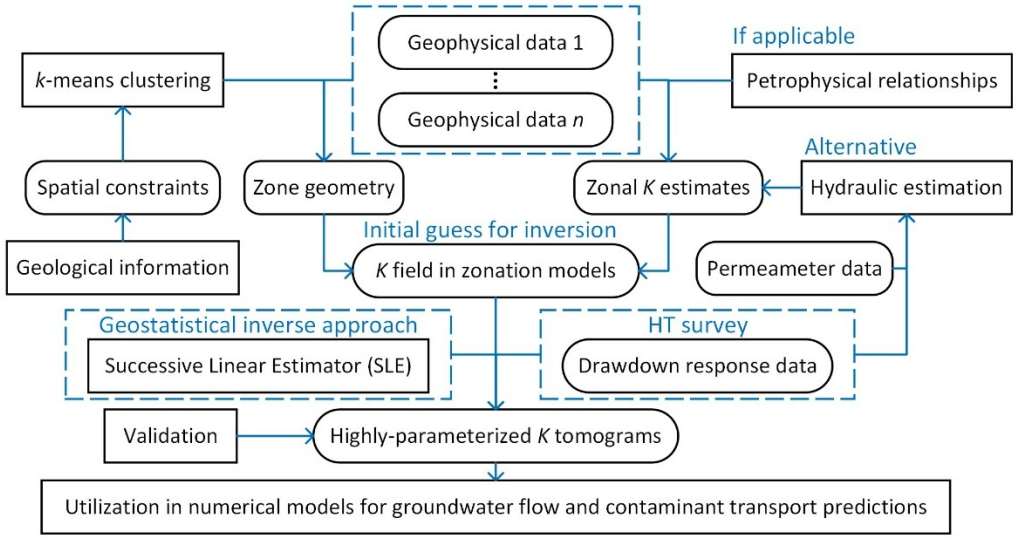


Figure 3.1: A general framework of integrating geophysical survey data to improve HT estimates.

3.4 The Experimental Design

3.4.1 Reference K Field Generation

To better assess the HT performance with initial zonal information for high-resolution characterization of subsurface heterogeneity, a numerical case based on an actual two-dimensional laboratory sandbox is designed, in which the experimental setup is fully controlled, including having perfect knowledge of the true *K* field, error-free head response data for pumping tests and boundary conditions. Specifically, the layer structure from a laboratory sandbox (Illman et al., 2010) is initially utilized to construct the primary layer boundaries (Figure 3.2a) in this experiment. Various sands of different grain sizes were deposited cyclically in the sandbox, and therefore an aquifer with interfingering layers was constructed for the simulation of a realistic sedimentary deposit. Overall, 18 layers were built in a 172.72 cm (length) by 82.28 cm (depth) two-dimensional numerical aquifer. Due to the fact that only sediment types from coarse to fine sands are utilized in the laboratory sandbox, *K* in the aquifer ranges from 3.20×10^{-4} m/s to 2.60×10^{-3} m/s (Illman et al., 2010), which does not satisfy our target to construct a highly-heterogeneous aquifer-aquitard system. To fulfill this requirement, hydraulic properties from a highly heterogeneous aquifer-aquitard system at the NCRS (North Campus Research Site, located at the University of Waterloo) was introduced to the numerical sandbox aquifer. A detailed description of this site can be found in Alexander et al. (2011) and Berg and Illman (2011b). The simulation domain was discretized into 120 × 58 elements with average dimensions of 1.44 cm × 1.40 cm in the *x*- and *z*-directions, respectively.

The contour plot of the reference *K* field for the numerical highly-heterogeneous aquifer-aquitard system is shown in Figure 3.2b. The heterogeneity pattern of the numerical aquifer is designed through two steps: 1) the large-scale heterogeneity pattern following the layer structure from the actual laboratory sandbox, and the geometric mean ($K_G = 1.17\times10^{-6}$ m/s) and variance of ln *K* ($s^2_{\ln K} = 6.93$) of the

modeling results of Zhao and Illman (2017, 2018, 2022a) are utilized to create the reference K field for the simulation domain. The interlayer variation of K values among each of the 18 layers was generated in MATLAB whereby a log-normal distribution is built for the interlayer heterogeneity of K. The effective values of K for the 18 layers are listed in Table A1; and 2) small-scale heterogeneity is considered within each of the 18 layers. The variance of ln K for intralayer heterogeneity is assumed to be 1/15 of the major variance of ln K (i.e., 6.93/15 = 0.462) in the aquifer. A random field generator called FGEN developed by Robin et al. (1993) using the direct Fourier Transform method is utilized to independently generate the heterogeneity within each layer. An average correlation length and depth of 28.79 cm and 2.71 cm, respectively, for each layer are provided in the FGEN generator. These two values are assumed to be 1/3 of the average correlation length and depth of the simulation domain, which are estimated using the SGEMS variogram modeling software (Remy et al., 2007). Hence, a weakly-stationary random field is generated for each layer with a specified mean, variance, and correlation distances. Overall, the K_G for the entire aquifer is 1.75×10^{-6} m/s, while the $s^2_{\ln K}$ is 6.54.

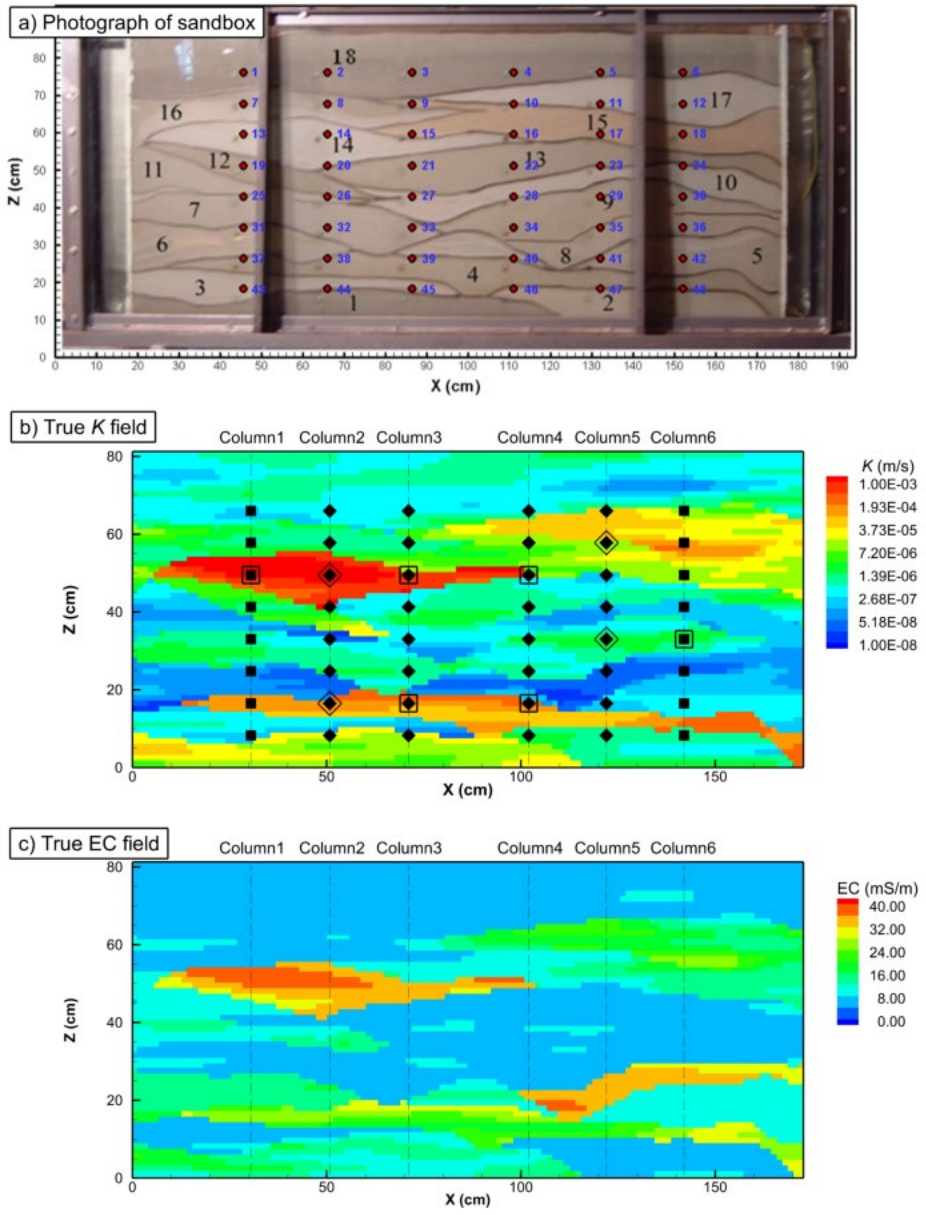


Figure 3.2: True field generation of *K* and EC for the numerical sandbox experiment. a) A photograph of the laboratory sandbox in which numbers of 18 layers are denoted in black, and 48 observation ports are indicated using red circles. b) Spatial distribution of the synthetically generated true *K* field based on the deposition pattern shown in figure 3.2a. Dashed lines show six columns of boreholes where observation ports are installed. 32 solid diamonds represent observation ports for model calibration, while 16 additional observation ports used for model validation are denoted as solid squares. Large hollow diamonds and squares stand for the pumping ports for the model calibration and validation processes, respectively. c) Spatial distribution of the corresponding EC field converted from *K* distribution. Co-located dashed lines indicate the locations of borehole geophysical surveys yielding EC and GR logs.

3.4.2 Geophysical Survey Data

A critical issue in hydrogeophysical characterization is that petrophysical relationships are always highly non-unique. To map the complex hydrogeophysical environment, a non-monotonic relationship between EC and *K* is built for the generation of a true EC field. As there is no direct relationship between EC and *K* (Slater, 2007), grain size information is introduced to associate EC with *K*. The Hazen method (Hazen, 1892) is applied to link the reference grain size with *K* for this numerical experiment. In terms of EC generation, two components, surface conductivity and electrolyte conductivity, are considered in bulk EC (σ), which exhibit disparate correlation patterns to the reference grain size. Both the clay effect and specific surface area (Slater & Lesmes, 2002; Weller et al., 2013) are considered in the generation of surface conductivity, while a form of Archie's law (Archie, 1942; Doveton, 1986) is used for the estimation of electrolyte conductivity from effective porosity. Meanwhile, Urumović (2016) provides a model in which effective porosity is a function of reference grain size. Detailed descriptions of EC data generation can be found in Text S1. Thus, a relationship can be built between the electrolyte conductivity and the reference grain size, and the bulk conductivity is a linear superposition of all components. The pore-surface and the pore-fluid parts of the reference EC field are shown in Figure A1a and A1b, respectively. Figure 3.2c shows the reference bulk EC field. EC logging data along six drillholes in the simulation domain can be extracted from the reference EC field.

Another borehole geophysical survey called the gamma ray (GR) radiation is involved in the logging data for clustering to assess the progressive performances of constructing zonation models. GR logging is a method of measuring natural gamma radiation in sediments along boreholes. It is usually recorded in American Petroleum Institute (API) units and used in a qualitative manner as it is impacted by various factors (e.g., well diameter, layer thickness, and radioactive elements). Potassium, a source of natural gamma radiation, usually occurs in clay minerals. Thus, two classes are divided for unconsolidated sediments: clay minerals with a relatively high GR, and the rest with an insignificant

proportion. Therefore, GR radiation is regarded as a good supplementary data type to EC logs. A classical model defined by Poupon and Gaymard (1970) is used to generate GR counts along boreholes:

$$I_{clay} = \frac{GR - \gamma_{min}}{\gamma_{max} - \gamma_{min}} \tag{3.5}$$

where I_{clay} denotes clay content, while γ_{max} and γ_{min} are, in principle, the maximum and minimum values provided by processing GR logs. GR logs measured at the NCRS by Gamma39 from Geonics Ltd. is processed to obtain γ_{max} and γ_{min}. A random field with a normal distribution is introduced into gamma ray profiles along boreholes. To conform with model discretization, a logging interval of 1.40 cm is used. Hence, 58 measurements of EC and GR along each borehole are obtained. EC and GR logs along six columns of the numerical aquifer-aquitard system are provided in Figure 3.3.

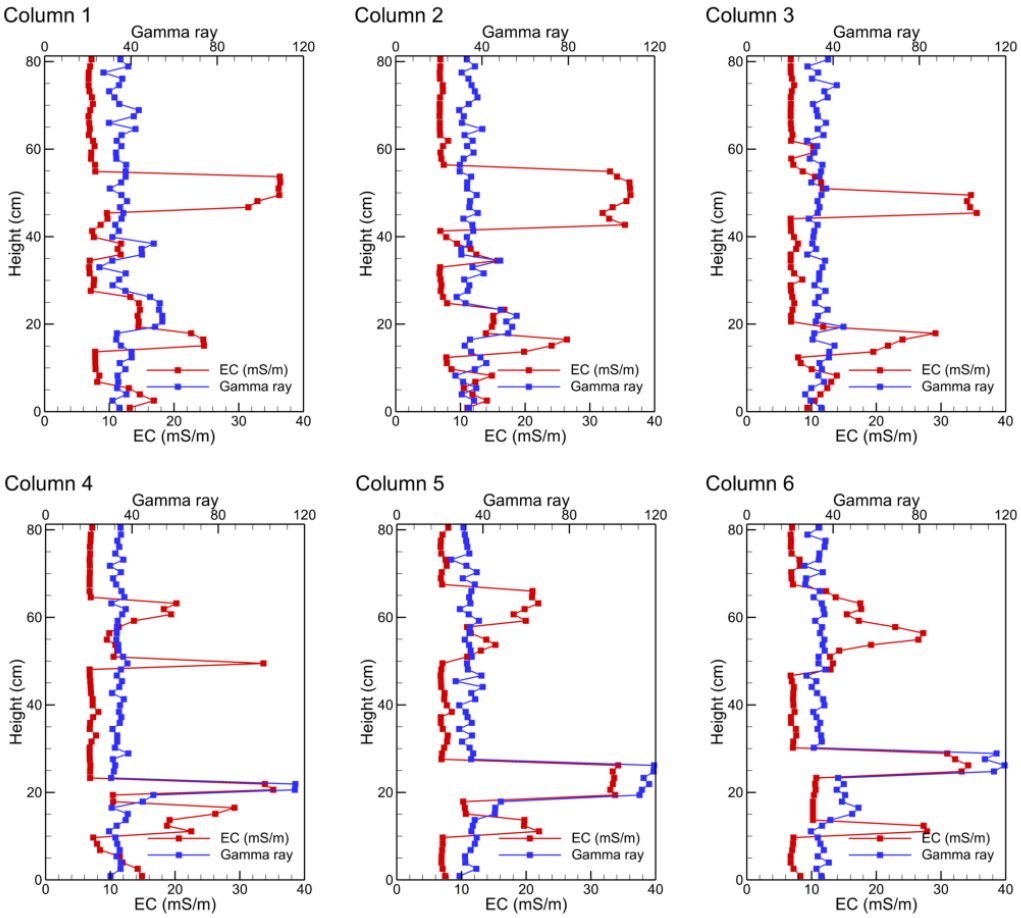


Figure 3.3: EC and GR logs along six boreholes. EC and GR profiles are labelled in red and blue, respectively.

3.5 Clustering Validity

3.5.1 Zone Geometry Extraction

The effectiveness of clustering with proposed spatial constraints is assessed before applied to HT analysis. For this purpose, K logging data are used as samples in clustering and converted to ln K profiles to extract zone geometry. In this manner, the uncertainty from petrophysical relationship can be excluded. In terms of the initialization of clustering, the k-means ++ algorithm (Arthur & Vassilvitskii, 2007) can be utilized to provide careful seedings of initial mean vectors. Furthermore, the “elbow” curve method (Kodinariya & Makwana, 2013) is used to select the optimal k from the output. As the objective function E has a decreasing trend with an increasing k, the "elbow" represents an inflection point at which the diminishing trend of the objective function curve slows down such that k after this point may exceed the actual number of data groups.

Geological knowledge can be employed to provide spatial constraints in clustering. In this study, spatial constraints are inferred from analyzing downhole logging data. A trial-and-error process of downhole k-means clustering with vertical constraints is conducted in the following manner: 1) corresponding to different k, a series of average unit thicknesses are computed as results of the entire logging distance over k. The resulting average unit thickness is utilized in clustering to yield various partitioning results of logging data for each k; 2) the objective function E regardless of spatial terms are plotted over k, in which the “elbow” point is determined, and the average unit thickness corresponding to the current k is an appropriate estimate of a vertical constraint at this location; and 3) a vertical constraint of the entire medium can be estimated by calculating the mean of vertical constraints at all drillhole locations. As there is no information on the average unit length of the medium provided in this case, it is

roughly estimated by multiplying the average thickness with a factor (e.g., 6–10 times) based on the common length-to-thickness ratio of strata.

For this trial-and-error process, six ln K profiles along Columns 1–6 are incorporated into clustering, respectively. The objective function E is computed for each clustering run. For simplicity, the relative sum of squares (E/E_1) is then calculated as the ratio of E at the current k over E at $k = 1$. Figure A2 presents the curves reflecting E/E_1 over k for Columns 1–6. "Elbow" points are selected and labelled with red circles on the plots. Afterwards, the corresponding average unit thickness can be calculated. In practice, the "elbow" points cannot always be unambiguously identified as their determination is dependent on visual observation. Therefore, an alternative set of "elbow" points (labelled with blue deltas) is also selected to produce another set of estimated average unit thicknesses, and we can see to what degree the selection of average unit dimensions can affect the clustering performance.

Four scenarios of zone geometry extraction (Figure 3.4) are considered in the validity check. Figure A3 shows the "elbow" curves and how k is determined in these scenarios. The k-means clustering without spatial constraints is considered in Scenario i. As there are abundant outliers in the clustering result regardless of spatial constraints, only four clusters can be determined for Scenario i. Conversely, a variety of spatial terms are tested for the rest of the scenarios to check the performance of spatial constraints.

Next, in Scenario ii, the preferential "elbow" point is selected to yield a vertical constraint of 6.91 cm, while a factor of 10 is assigned to yield a horizontal constraint of 69.1 cm. Consequently, eight clusters are partitioned in this scenario. For Scenario iii, an alternative "elbow" point is selected to yield a vertical constraint of 14.34 cm. Meanwhile, the same factor is assigned to yield a horizontal constraint of 143.4 cm. Therefore, 10 groups are obtained through clustering. Lastly, for Scenario iv, the preferential "elbow" point is also applied, but a factor of 6 is now used to obtain a horizontal constraint of 41.5 cm. As a result, ln K profiles are divided into 10 clusters in Scenario iv as well. Furthermore, a priori true ln K

field is assigned as the reference model of clustering results, and the zone geometry from the reference model is shown in Figure 3.4e.

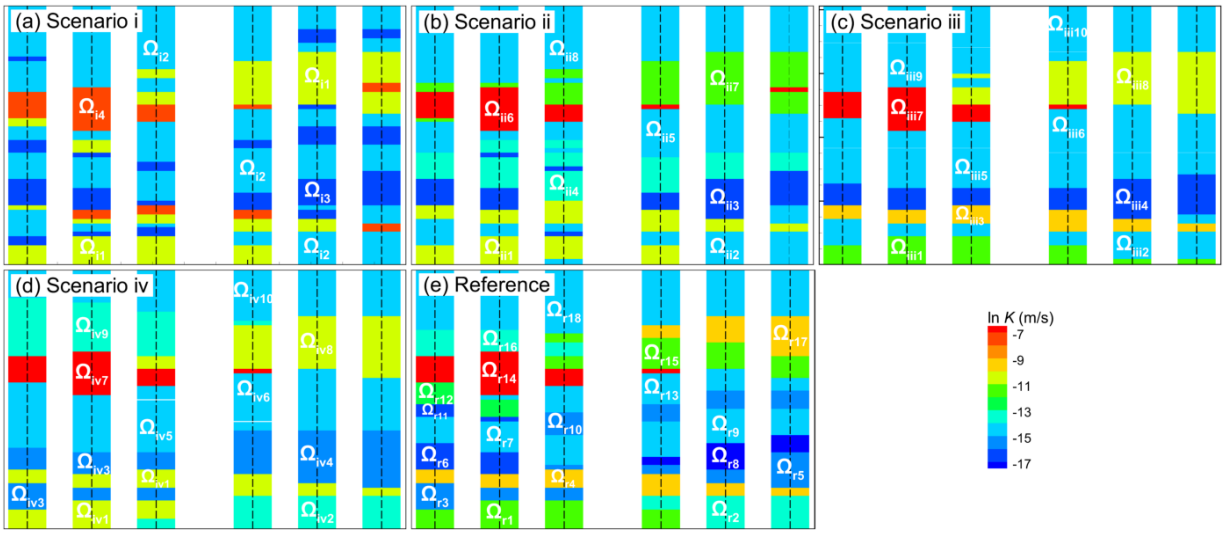


Figure 3.4: Four scenarios of zone geometry extracted from the clustering of ln *K* profiles along six boreholes, including: a) Scenario i: zone geometry by *k*-means clustering without spatial constraints; b) Scenario ii: zone geometry by *k*-means clustering with a preferential vertical constraint and a horizontal constraint generated by multiplying a factor of 10; c) Scenario iii: zone geometry by *k*-means clustering with an alternative vertical constraint and a horizontal constraint generated by multiplying a factor of 10; d) Scenario iv: zone geometry by *k*-means clustering with a preferential vertical constraint and a horizontal constraint generated by multiplying a factor of 6; and e) is the zone geometry from the reference model. The effective mean of ln *K* measurements is assigned to each zone for visual observation.

3.5.2 Evaluation of Clustering Performance

For the evaluation of clustering performance, clustering results with/without spatial constraints are quantitatively compared to the reference model. Three types of external Cluster Validity Indices (CVIs) (Fowlkes & Mallows, 1983; Jaccard, 1912; Rand, 1971) are employed to compare the similarity of zone geometry between clustering results and the reference model. Detailed descriptions of the three CVIs are included in Text A2. The CVIs range from 0 to 1. The larger the CVIs are, the better clustering results can reproduce the reference model. A summary of the comparison is in Table 3.1.

Table 3.1: Summary of Validity Check Results Using CVIs.

Scenarios	Number of Clusters (k)	Spatial Constraints (cm)		Cluster Validity Indices (CVIs)		
		Horizontal	Vertical	JC	FMI	RI
i	4	N/A	N/A	0.19	0.41	0.71
ii	8	69.10	6.91	0.46	0.64	0.93
iii	10	143.40	14.34	0.52	0.71	0.93
iv	10	41.50	6.91	0.44	0.61	0.93

3.5.3 Discussion

Figure 3.4 indicates that zone geometry can be satisfactorily extracted from ln *K* profiles. In particular, the farther the samples deviate from the mean of sample space, the better the clusters containing these samples can be partitioned. In Scenario i, biased representation of effective ln *K* values at each zone occurs, especially for Cluster Ω_{i1} and Ω_{i4}. As the effective mean for each zone represents cluster centroids, the biases reflect unreasonable partitioning of samples in Scenario i, which might stem from the assignment of samples from faraway regions into one cluster regardless of spatial adjacency. When spatial constraints are incorporated (Scenario ii), the zone geometry and zonal mean of ln *K* values can better reproduce the characteristics in the reference model. Besides, we notice that the optimal k can be determined more readily through visual observation in the "elbow" curve analysis for the scenarios with spatial constraints (see Figure A3). It reveals that the dissimilarity of different clusters in soil characteristics can be better identified with the use of spatial constraints.

Quantitatively, the change in three CVIs shows similar trends for various scenarios. Scenario iii leads the validity check performance among all scenarios, while Scenarios ii and iv follow closely after Scenario iii. In contrast, Scenario i lags significantly behind the other scenarios. These results indicate that clustering results with spatial constraints exhibit higher similarity to the reference model. The alternative spatial constraints applied in Scenario iii result in a small influence on the objective function E^*. The leading performance of Scenario iii indicates that the weights of spatial constraints can be reduced when reliable sources of samples (i.e., ln *K* profiles) are clustered for the interpretation of hydrostratigraphy. Nevertheless, the significance of HT analysis integrated with the clustering of

geophysical survey logs should be examined through the simulation of groundwater flow and solute transport.

3.6 Hydraulic Tomography Analyses

3.6.1 Model Setup

There are six columns of boreholes within the aquifer-aquitard system, and eight ports (see Figure 3.2) are distributed along each borehole for the observation of drawdown data during pumping tests. The region between Columns 2 and 5 is the main pumping zone, in which four pumping tests are conducted along Columns 2 and 5 at relatively permeable locations, including Ports 2-3, 2-7, 5-2 and 5-5 (e.g., Port 2-3 means the third port from the top along Column 2). Besides, 31 observation ports in total from Columns 2 to 5 (i.e., drawdown observations at pumping ports are excluded) are utilized to observe drawdowns for the calibration of zonation and geostatistical inverse models. In addition, six independent pumping tests are conducted at Ports 1-3, 3-3, 3-7, 4-3, 4-7, and 6-5 to generate data for the model validation effort, in which observation records along six columns are all included. Pumping tests are conducted under steady-state flow conditions such that constant drawdown observations are obtained. In total, 124 steady-state drawdown data are used for model calibration, while 282 steady-state drawdown data are utilized for model validation. Pumping rates (Q) for all pumping tests are listed in Table A2. In terms of boundary conditions, constant head boundaries with a value of 100 cm are assigned at top and lateral boundaries of the sandbox, while a no-flow boundary is assigned at the sandbox bottom. After modeling groundwater flow, a forced-gradient tracer test is conducted in the numerical aquifer to evaluate HT estimates.

Previous studies (e.g., Luo et al., 2017; Zhao et al., 2016; Zhao & Illman, 2017, 2018) in HT analysis have demonstrated that the highly-parameterized geostatistical inverse modeling approach can characterize the subsurface heterogeneity well and successfully predict drawdown behaviors. Therefore,

this case study mainly focuses on the comparison of HT performances by utilizing varying zonal information from geophysical survey logs as initial guesses using the geostatistical inverse approach. Accordingly, the same drawdown data are utilized for all inversion cases. The main difference is the varying initial distribution of hydraulic parameters (i.e., *K* in this study) utilized in model calibration. Again, initial models of HT analysis are established by constructing zone geometry from the clustering of geophysical information and filling with zonal effective *K* by coupling the groundwater flow simulator in VSAFT3 with PEST. In total, six cases including 10 scenarios are designed to highlight the significance of HT analysis integrated with geophysical survey logs.

In Case I, a homogeneous field of *K* is assumed as an initial model. In Case II, only borehole geophysical survey logs are involved in *k*-means clustering to extract zonal information. Cases III and IV are designed on the basis of Case II. In Case III, spatial constraints are estimated and then included in the clustering to produce a different interpretation of zone geometry, while in Case IV, point *K* measurements are integrated into zonal *K* estimation to provide more constraints. Both spatial constraints and point *K* measurements are included in Case V. Besides, two different scenarios (i.e., "-a" and "-b" scenarios) are designed for each case from Cases II to V using different geophysical survey logs. Six columns of EC logs are used for the clustering in the "-a" scenarios, while six columns of collocated EC and GR logs are included for the clustering in the "-b" scenarios. Furthermore, the true layer structure is provided as zone geometry in Case VI. A summary of all cases is listed in Table 3.2.

Table 3.2: Summary of cases for HT analyses.

Case number	Borehole geophysical logging types	Additional information for zonation
I	N/A	N/A
II-a	EC	N/A
II-b	EC and GR	N/A
III-a	EC	Spatial constraints for clustering
III-b	EC and GR	Spatial constraints for clustering
IV-a	EC	Local *K* measurements available
IV-b	EC and GR	Local *K* measurements available

V-a	EC	Spatial constraints and local *K* measurements
V-b	EC and GR	Spatial constraints and local *K* measurements
VI	N/A	True layer structure

3.6.2 Zonation Models

The clustering of EC data along six boreholes (see Figure A4) is utilized to yield an estimate of vertical constraint, while a factor of 10 is assigned to obtain an estimate of horizontal constraint. As a result, the average unit length and the average unit thickness for spatial constraints are 98.21 cm and 9.821 cm, respectively. The "elbow" curve method is applied to determine *k* for different cases (see Figure A5). After clustering results are produced for all cases, zone geometry is delineated by interpolation and extrapolation of borehole zonal information using Leapfrog Geo, and zonal *K* values are estimated from drawdown data of the four pumping tests. Figure 3.5 is the spatial distribution of *K* in zonation models for different cases.

In Case II-a, four zones are partitioned by the clustering of EC logs only. Figure 3.5a indicates that in general, it delineates the layer structure of the aquifer-aquitard system, but the shape and volume of layers obtained for this zonation model differs from the true stratigraphy. Besides, the estimation of *K* values cannot satisfactorily reproduce the interlayer heterogeneity of *K* pattern. One significant misinterpretation in Figure 3.5a is that Layers 5, 6, and 8 (layer numbers are indicated on Figure 3.2a) are erroneously interpreted as layers with intermediate-to-high *K*, which should in fact be aquitards with low-*K* materials. While in Case II-b, the estimation of *K* values in the three layers (Figure 3.5b) is considerably improved due to better delineation of zone geometry by the joint clustering of EC and GR logs. It indicates the ability of reproducing the variation of hydraulic properties by geophysical survey increases with the combination of multiple non-redundant geophysical surveys, while using EC logs can only partially capture the hydrostratigraphy. However, the resolution of the zonation model in Figure 3.5b is still relatively low.

When spatial constraints are included in clustering, 11 and 13 zones are partitioned through the clustering of EC logs only and through the clustering of EC and GR together, respectively. Figures 3.5c and 3.5d are the zonation models for Cases III-a and III-b, respectively. Compared to the zonation model for Case II-a, the resolution of layer structure in Case III-a is significantly improved, and the interlayer heterogeneity of *K* can be better estimated. In addition, the delineation of layer structure is further improved in the zonation model for Case III-b (Figure 3.5d). However, the estimation of zonal *K* is still biased especially in that adjacent zones in geometry cannot be populated with appropriate parameters (e.g., in Layers 14 and 15, also in Layers 1 and 4).

Through visual observation, zonation models for Case IV (see Figure 3.5e for Case IV-a and Figure 3.5f for Case IV-b) are almost the same with Case II in both zone geometry and zonal *K* assignments. Conversely, significant differences in the estimated parameters of zonation models for Case V can be found in comparison to Case III. Eleven zones are partitioned in Case V-a (see Figure 3.5g), which reflects well the main characteristics of the true hydrostratigraphy in terms of both layer structure and zonal parameter estimation. The hydrostratigraphic zone characterization for Case V-b (Figure 3.5h) is further enhanced.

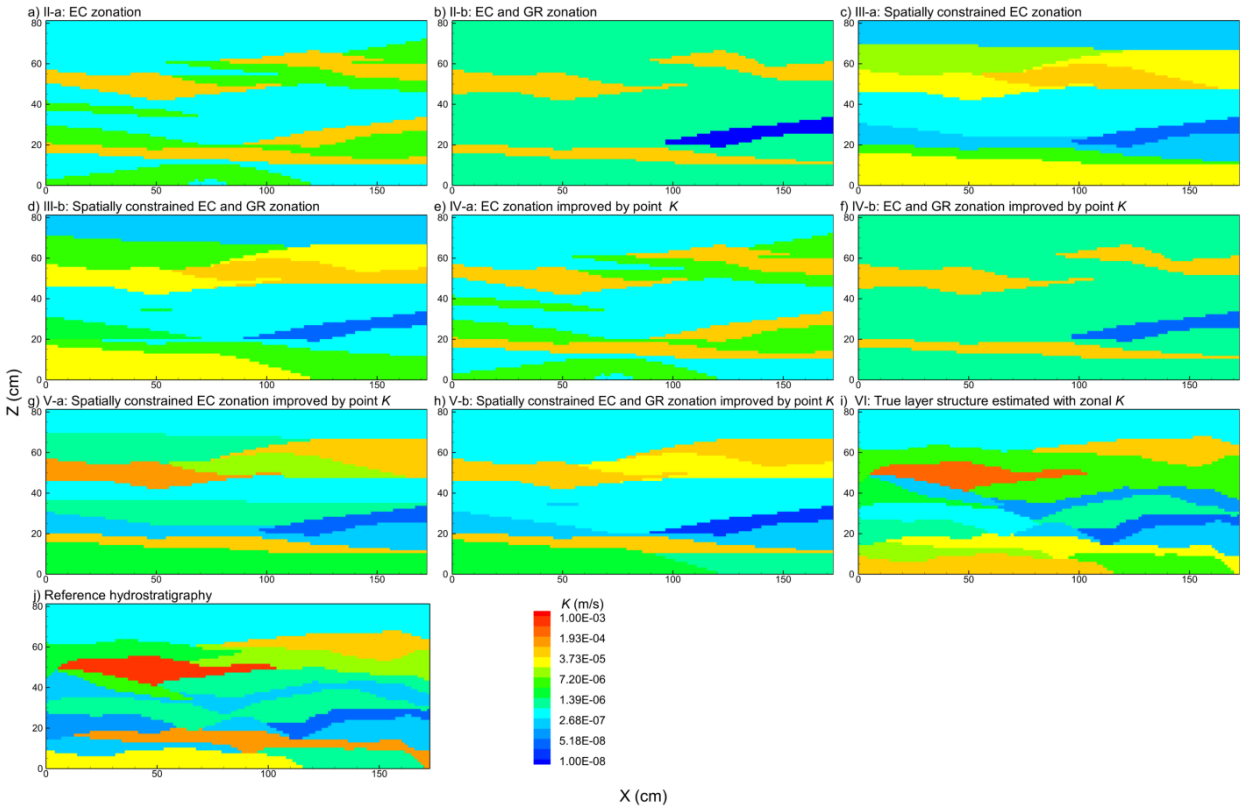


Figure 3.5: Spatial distribution of *K* fields in zonation models for different cases. Zonation models are constructed by: a) Case II-a: clustering from EC logs; b) Case II-b: clustering from EC and GR logs; c) Case III-a: spatially constrained clustering from EC logs; d) Case III-b: spatially constrained clustering from EC and GR logs; e) Case IV-a: clustering from EC logs and improved with local *K* measurements; f) Case IV-b: clustering from EC and GR logs and improved with local *K* measurements; g) Case V-a: spatially constrained clustering from EC logs and improved with local *K* measurements; h) Case V-b: spatially constrained clustering from EC and GR logs and improved with local *K* measurements; i) Case VI: filling zonal *K* values to the zone geometry from the true layer structure; and j) shows the reference hydrostratigraphy.

3.6.3 Model Calibration Results

The zonation models are calibrated using SimSLE built in VSAFT3. Model calibration results are assessed by comparing simulated with observed drawdowns at observation ports. Two model performance metrics, including the mean absolute error (L_1) and the mean squared error (L_2), are utilized to evaluate their performances:

$$L_1 = \frac{1}{n}\sum_{i=1}^{n}|s_i - s_i^*| \tag{3.6}$$

$$L_2 = \frac{1}{n}\sum_{i=1}^{n}(s_i - s_i^*)^2 \tag{3.7}$$

where n is the total number of drawdown data included. s_i and s_i^* indicate the i^{th} simulated and observed drawdown data, respectively. The smaller the L_1 and L_2 norms are, the better the models are calibrated. As the L_2 norm can better magnify large discrepancies in drawdown, it is also utilized to determine where the iteration stops. The L_2 norm is plotted as a function of the iteration number, and the iteration at which the inflection of curve flattens is selected for model output. In addition, a linear fit line is provided for each case, and the coefficient of determination (R^2) is computed as follows:

$$R^2 = 1 - \frac{\sum_i^n (s_i - \widehat{s_i})^2}{\sum_i^n (s_i - \overline{s})^2} \tag{3.8}$$

where $\widehat{s_i}$ and $\overline{s}$ are the predicted drawdown in the linear model fit and the mean value of all simulated drawdowns, respectively. Figure A6 shows scatterplots of simulated versus observed drawdowns of model calibration results from various cases. The L_1 and L_2 norms reveal that zonation models for all cases are calibrated well by the geostatistical inversion of multiple pumping test data. The small differences in model performance metrics may only result from the different selection of iteration number. Figure 3.6 is spatial distribution of K tomograms for all cases. The true K field (see Figure 3.6k) is also provided for comparison.

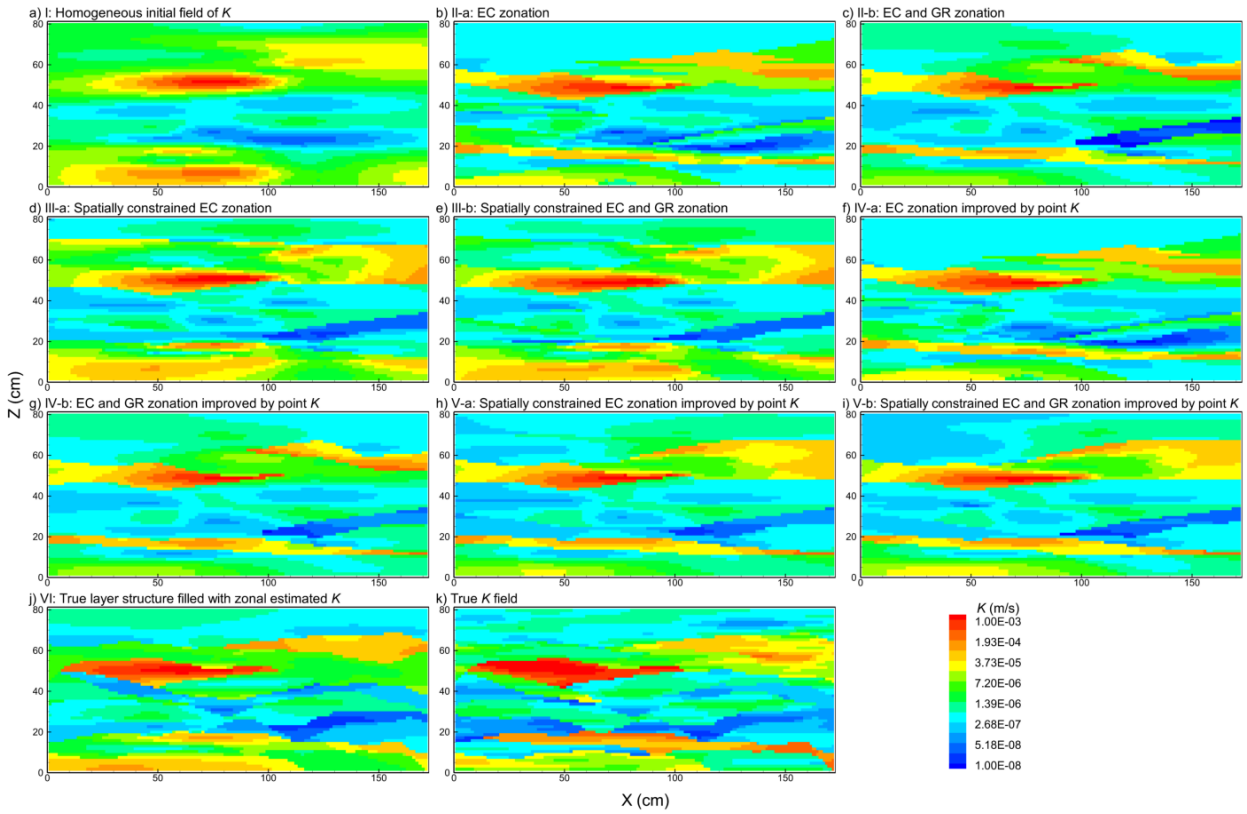


Figure 3.6: Spatial distribution of *K* tomograms for different cases. Geostatistical inversions are initialized with: a) Case I: homogeneous *K* field; zonation models from b) Case II-a: clustering of EC; c) Case II-b: clustering of EC and GR; d) Case III-a: spatially constrained clustering of EC; e) Case III-b: spatially constrained clustering of EC and GR; f) Case IV-a: clustering of EC, and improved with local *K* measurements; g) Case IV-b: clustering of EC and GR, and improved with local *K* measurements; h) Case V-a: spatially constrained clustering of EC, and improved with local *K* measurements; i) Case V-b: spatially constrained clustering of EC and GR, and improved with local *K* measurements; j) Case VI: true layer structure. k) shows the true *K* field for comparison.

In Case I, the robust performance of the HT analysis with the geostatistical inverse approach using SimSLE is shown as anticipated. High-*K* and low-*K* regions of the aquifer-aquitard system (Figure 3.6a) can be generally characterized although the resolution of HT estimates is relatively low. To some degree, the locations and shapes of high-*K* and low-*K* layers are different from the true *K* field (Figure 3.6k). As the average observation interval used in this study is close to the vertical correlation scale (Yeh & Liu, 2000), satisfactory observation density should have been achieved. However, with data from four pumping tests, we can still see that the mapping of subsurface heterogeneity is not satisfactory for a

highly-heterogeneous aquifer-aquitard system. It supports the conclusion in Yeh et al. (2008b) and Bohling and Butler (2010) that the fusion of hydraulic, geological and geophysical characterization techniques is strongly encouraged to obtain better estimates of the *K* tomogram.

Qualitatively, the resolution of *K* tomogram is significantly improved with the use of zonal information as the initial model. The *K* tomogram (Figure 3.6b) in Case II-a successfully reproduces three high-*K* layers (i.e., Layers 4, 14 and 17), but two aquitards (i.e., Layers 6 and 8) are poorly mapped. Based on the performance of Case II-a, the *K* tomogram (Figure 3.6c) of Case II-b can reproduce Layer 8 well as it is successfully identified in its zonation model. Conversely, low-*K* layers are well calibrated in Case III-a (Figure 3.6d), and two high-*K* layers, Layers 14 and 18, are also successfully captured, which are originally not clearly characterized in its zonation model. It demonstrates the highly-parameterized HT analysis using the SLE algorithm can further identify new aquifer characteristics which are not mapped in its zonation model. However, *K* is over-estimated in Layer 1, and the geometry of Layer 4, which should be a long and thin aquifer, is not well characterized for Case III-a. Besides, due to the discrepancy between the layer structure in the zonation model and the true stratigraphy, misinterpretation of the layer connection between Layers 16 and 18 is evident. Nevertheless, when GR logs are included in the clustering for Case III-b, this misinterpretation disappears, and the layer structure is better delineated in HT estimates (Figure 3.6e).

Next, point *K* measurements are involved to improve zonal *K* estimation in Cases IV and V. However, *K* tomograms for Cases IV-a and IV-b are similar to those from Cases II-a and II-b through visual observation. Nonetheless, significant improvements to *K* tomograms can be observed from Cases V-a and V-b in comparison to Case III. The *K* tomogram for Case V-a successfully reproduces all main characteristics of the heterogeneous aquifer-aquitard system, including better interpretation of Layer 1 and the delineation of Layer 4. Meanwhile, improper interpretation of layer boundaries between Layers

16 and 18 has been rectified in the K tomogram for Case V-a. Likewise, the K tomogram for Case V-b is also improved compared to Case III-b.

To quantitatively compare K distributions from model calibration, scatterplots of estimated versus true ln K distribution are provided in Figure 3.7. Case VI leads all other cases with $L_1 = 0.98$ and $L_2 = 1.67$. This result is not surprising because the true layer structure is provided in Case VI. In contrast, Case I ranks behind all other cases. When zonation models are utilized as the initial guess for the inversion, L_1 and L_2 norms in Case II-a are significantly smaller than Case I, and they keep decreasing in Case II-b. When spatial constraints are included in clustering, L_1 and L_2 norms in Case III-a are slightly increased in comparison to Case II-a, while L_1 and L_2 norms in Case III-b are slightly decreased in comparison to Case II-b. As a result, the L_2 norm ranges from 3.35 to 3.58 for these four cases. Even though Case III-b performs the best among the cases integrated with indirect geophysical information only, the advantage of Case III-b is too small to demonstrate whether the inclusion of spatial constraints is beneficial to the characterization of K heterogeneity in terms of the comparison of ln K distribution. Next, L_1 and L_2 norms reduce steadily in Cases IV-a and IV-b in comparison to Cases II-a and II-b, but on the other hand, they decline significantly in Cases V-a (e.g., $L_2 = 2.68$) and V-b (e.g., $L_2 = 2.84$) in comparison to Cases III-a and III-b.

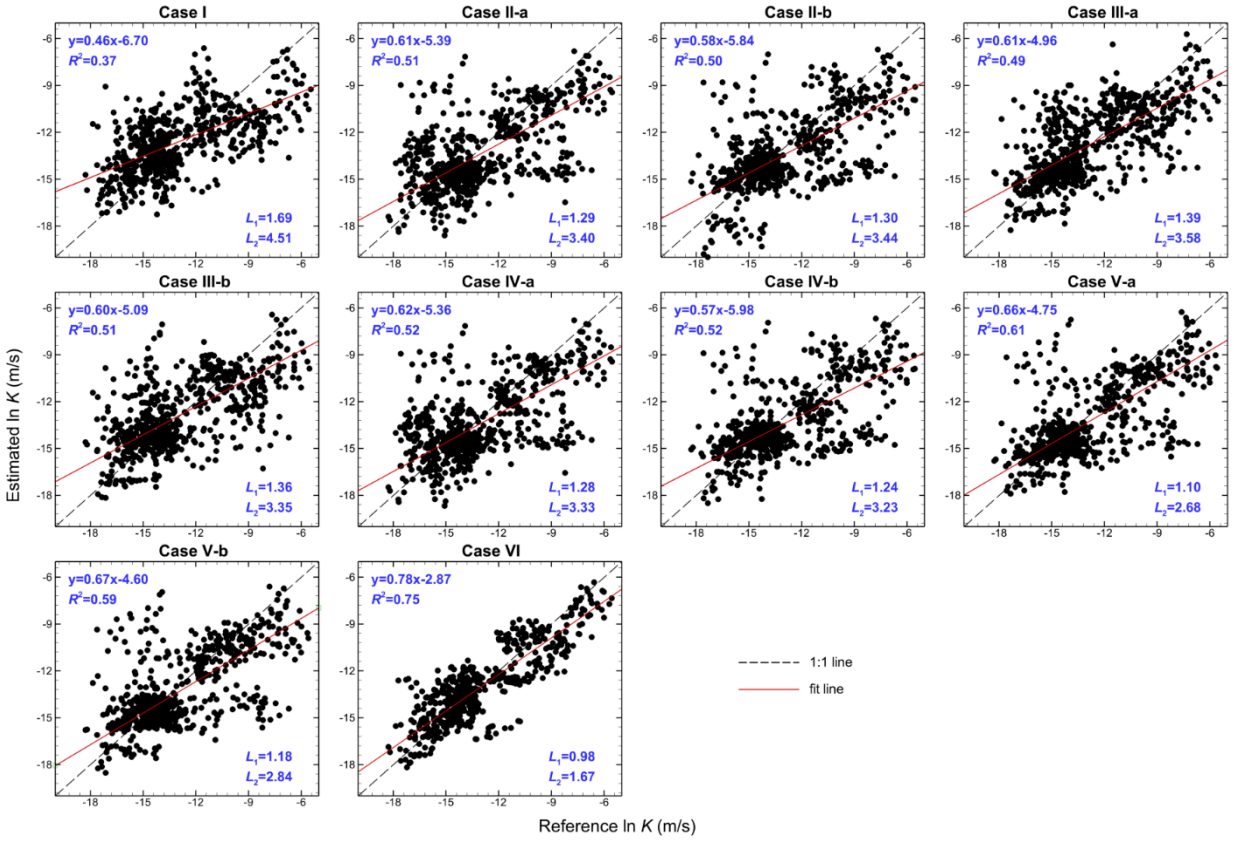


Figure 3.7: Scatterplots of simulated versus true ln K distribution for different cases. Linear model fits are provided as red solid lines in each plot, together with the linear-fit equation as well as values of R^2, and L_1 and L_2 norms.

3.6.4 Assessment of Model Validation Performance

HT results are then validated by simulating drawdown data from six independent pumping tests. Locations of six pumping tests are distributed over the entire simulation domain. To better assess model validation performance, three regions are divided in terms of pumping locations: first, the upper pumping zone, including the upper half of the region between Columns 2 and 5, in which two high-K zones (i.e., Layers 14 and 17) are involved; second, the lower pumping zone, including the lower half of the region between Columns 2 and 5, in which two high-K zones (i.e., Layers 1 and 4) and two low-K zones (Layers 6 and 8) are contained; and the last one is the outer region beyond the main pumping zone, in which no observation ports are located, and all of the aforementioned high-K and low-K zones are contained to

some degree. Scatterplots of simulated versus observed drawdowns at 47 observation ports are provided in Figure 3.8. L_1 and L_2 norms are computed as well for each pumping test, and a summary is provided in Figure 3.9. Color scales from green to red are utilized to exhibit the difference among all cases, in which the dark green represents the best performance, while the dark red represents the worst.

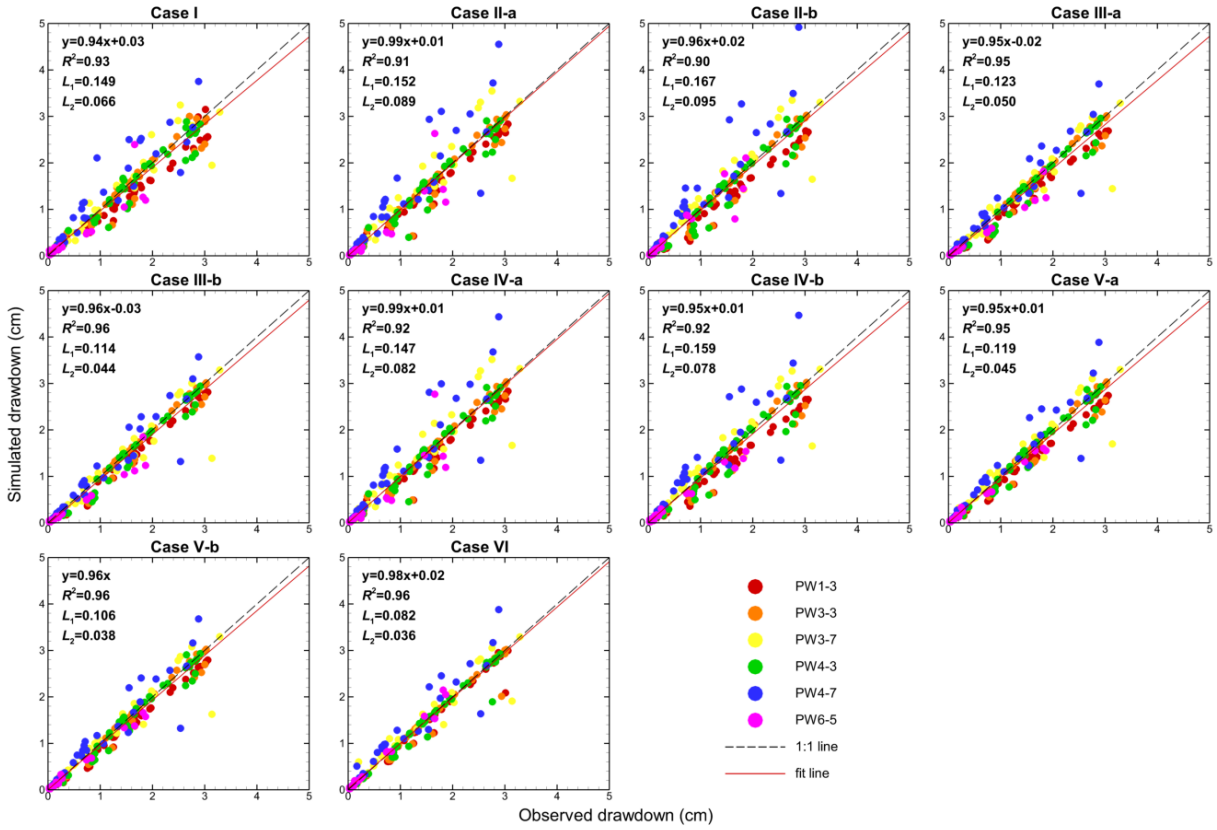


Figure 3.8: Validation scatterplots of simulated versus observed drawdowns for different cases. Observation data of various pumping locations are highlighted with different colors. R^2, L_1 and L_2 norms are indicated on each plot.

Case	The upper half				The lower half				The outer region				Overall		Rank
	PW3-3		PW4-3		PW3-7		PW4-7		PW1-3		PW6-5				
	L_1	L_2	L_1	L_2	L_1	L_2	L_1	L_2	L_1	L_2	L_1	L_2	L_1	L_2	
I	0.124	0.045	0.136	0.048	0.115	0.062	0.209	0.128	0.223	0.078	0.087	0.037	0.149	0.066	6
II-a	0.101	0.037	0.116	0.040	0.181	0.124	0.267	0.245	0.165	0.049	0.084	0.041	0.152	0.089	9
II-b	0.130	0.049	0.141	0.052	0.162	0.098	0.275	0.267	0.241	0.083	0.055	0.023	0.167	0.095	10
III-a	0.107	0.034	0.102	0.033	0.114	0.076	0.136	0.072	0.206	0.061	0.070	0.021	0.123	0.050	5
III-b	0.097	0.026	0.095	0.026	0.121	0.083	0.136	0.068	0.163	0.039	0.074	0.023	0.114	0.044	3
IV-a	0.099	0.034	0.114	0.036	0.177	0.117	0.250	0.213	0.163	0.046	0.082	0.045	0.147	0.082	8
IV-b	0.131	0.049	0.140	0.051	0.148	0.088	0.235	0.183	0.244	0.085	0.055	0.014	0.159	0.078	7
V-a	0.097	0.028	0.106	0.030	0.116	0.066	0.162	0.090	0.193	0.053	0.040	0.005	0.119	0.045	4
V-b	0.084	0.019	0.095	0.021	0.109	0.066	0.147	0.079	0.159	0.036	0.040	0.006	0.106	0.038	2
VI	0.057	0.025	0.060	0.023	0.109	0.054	0.166	0.082	0.075	0.027	0.027	0.004	0.082	0.036	1

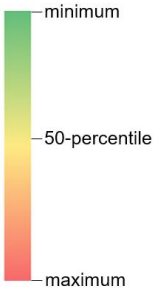


Figure 3.9: Summary of color-coded L_1 and L_2 norms for the three zones and overall validation results. A ranking is also provided to show differences among results.

Generally, calibrated models for all cases can satisfactorily predict drawdowns from independent pumping tests via the geostatistical inverse approach. Owing to the inclusion of error-free drawdown data and the design of small vertical monitoring intervals, relatively small values of L_1 and L_2 norms are computed. Again, Case VI ranks the first among all cases, while Case V-b follows closely behind Case VI. Specifically, Case V-b is comparable to Case VI in terms of model validation performance in the lower half of the main pumping zone, which indicates that Case V-b can successfully interpret multiple interfingering high-K and low-K zones in the sense of drawdown predictions. Afterwards, the other three cases containing spatial constraints in the clustering, namely Cases III-b, V-a, and III-a, rank in sequence after Case V-b. The gap of the L_2 norms among these three cases is narrow, but their L_2 norms, regardless of locations of pumping ports, are significantly smaller than for the cases using clustering without spatial constraints.

Different from the results of ln K comparison, four cases without the incorporation of spatial constraints in clustering, including Cases II-a, II-b, IV-a and IV-b, perform poorly for the model validation phase. The overall L_2 norm in Case I is even slightly better than these four cases. In particular, the interpretation of the lower half in Case I is substantially better than these four cases in the sense of

drawdown predictions. It reveals that zonation models constructed by clustering without spatial constraints are not always reliable for drawdown predictions.

3.6.5 Solute Transport Predictions

In addition to improved drawdown prediction performance, the ability of HT estimates to predict the migration of a contaminant plume is also tested to further highlight the significance of integrating geophysical information. A numerical forced-gradient dipole tracer test is designed in the heterogeneous aquifer-aquitard system, in which a steady-state flow condition is established with constant water injection and extraction. Then, a conservative tracer (e.g., bromide) with relative concentration (C/C_0) of 1.0 is injected at Port 1-5 at a rate of 0.1 mL/s, while water is extracted at Port 6-1 at a rate of 0.2 mL/s. Tracer test simulation is conducted using HydroGeoSphere (HGS) (Aquanty Inc., 2025) in this study. Tracer injection lasts for 48 hours, and an extra 48 hours of recovery follows, in which only clean water is injected at the same rate. All boundaries are assigned as impermeable boundaries during the solute transport experiment. An effective porosity of 0.3, and a tortuosity of 0.6 are assigned for the entire simulation domain. Furthermore, determination of longitudinal dispersivity refers to a study (Zhao et al., 2022) based on the prototype sandbox of our numerical experiment, in which a homogeneous longitudinal dispersivity was estimated to fit breakthrough curves (BTCs) of a tracer test. Thus, a longitudinal dispersivity of 0.154 m from an estimation based on a K field calibrated through the geostatistical inversion of HT analysis (i.e., Case 2a in Zhao et al., 2022) is selected for homogeneous longitudinal dispersivity value in our simulation domain for all cases. Correspondingly, the transverse dispersivity is assigned to be 1/5 of the longitudinal dispersivity. Table 3.3 provides a list of transport parameters for simulation.

Table 3.3: Transport parameters used in the solute transport simulation.

Transport parameters	Values
Effective porosity	0.3

Longitudinal dispersivity	0.154 m
Transverse dispersivity	0.031 m
Tortuosity	0.6

C/C_0 distributions at 4 h, 12 h, 48 h and 96 h for all cases and the true K field are provided in Figure 3.10. In terms of C/C_0 distributions for the true K field, the tracer plume initially migrates from the influent port (i.e., Port 1-5) to the effluent port (i.e., Port 6-1) along high-K Layer 14 at 12 h. Then, the tracer plume continues to evolve. It passes through the discontinuous low-K region (consisting of Layer 6 and Layer 8) and migrates along the high-K Layer 4 at 48 h. After 48 h of dissipation, most of the tracer mass is extracted.

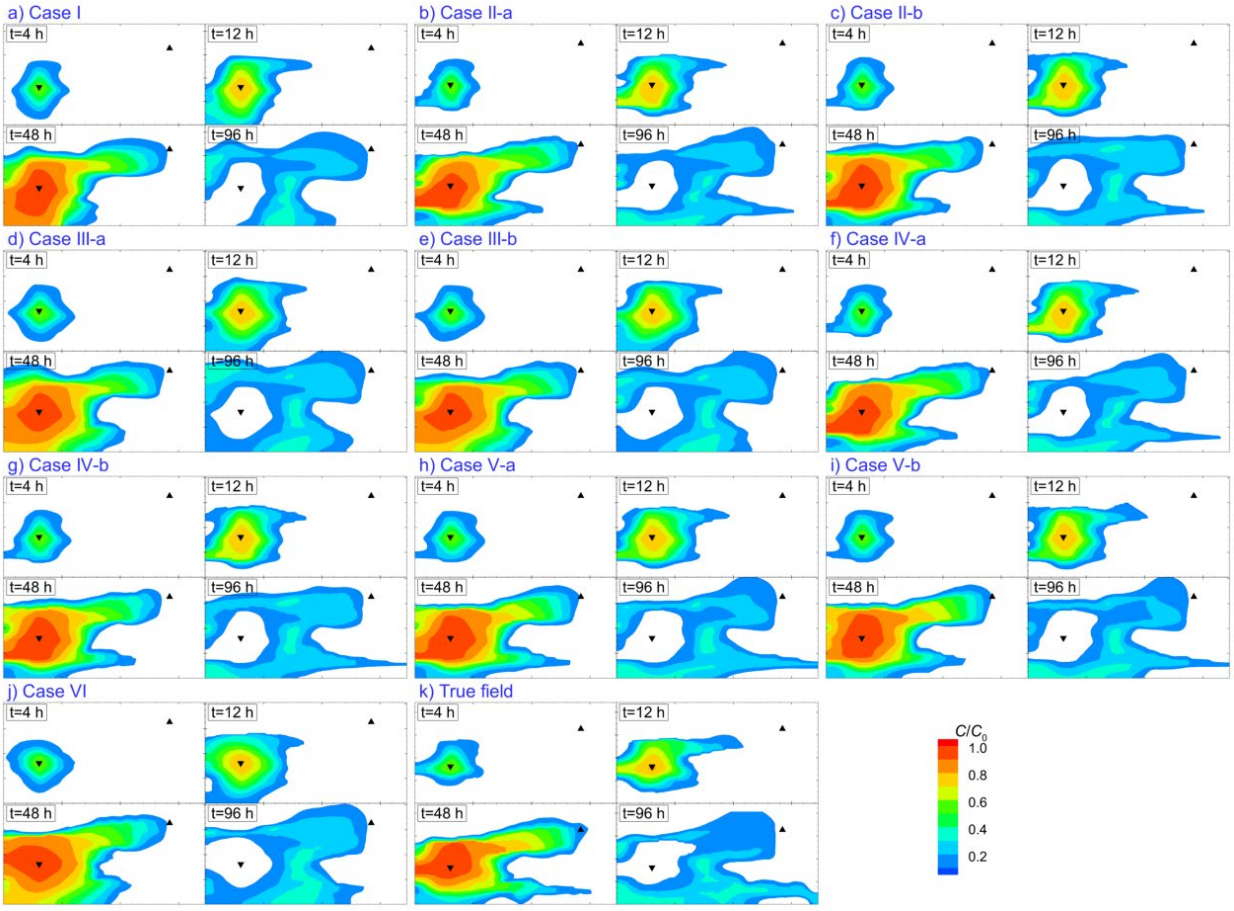


Figure 3.10: Relative concentration distributions (C/C_0) at 4 h, 12 h, 48 h and 96 h for all cases. a) through j) represent C/C_0 profiles for Cases I, II-a, II-b, III-a, III-b, IV-a, IV-b, V-a, V-b and VI. k) shows the C/C_0 distributions based on the spatial distribution of true K field. Influent and effluent ports in each

subplot are denoted with downward and upward-pointing triangles, respectively. Contours with C/C_0 under 0.1 are removed for clarity.

Concentration data at sampling ports are collected every two hours. In terms of distances from the source, nine representative sampling ports are utilized to yield BTCs during 96 hours of tracer migration, including three classes: (1) near the source: Ports 1-4, 2-5 and 1-7; (2) at intermediate distances: Ports 3-3, 3-7 and 4-3; and (3) at large distances: Ports 5-3, 5-2 and 6-1. BTCs of the nine sampling ports are plotted in Figure 3.11. For clarity, BTCs of six representative cases are provided in this figure, including Cases I, II-b, III-b, IV-b, V-b and VI. In addition, scatterplots of simulated versus observed C/C_0 at 12 hours and 48 hours from all sampling ports are provided in Figure 3.12, while Figure A7 contains scatterplots for samples recorded every two hours.

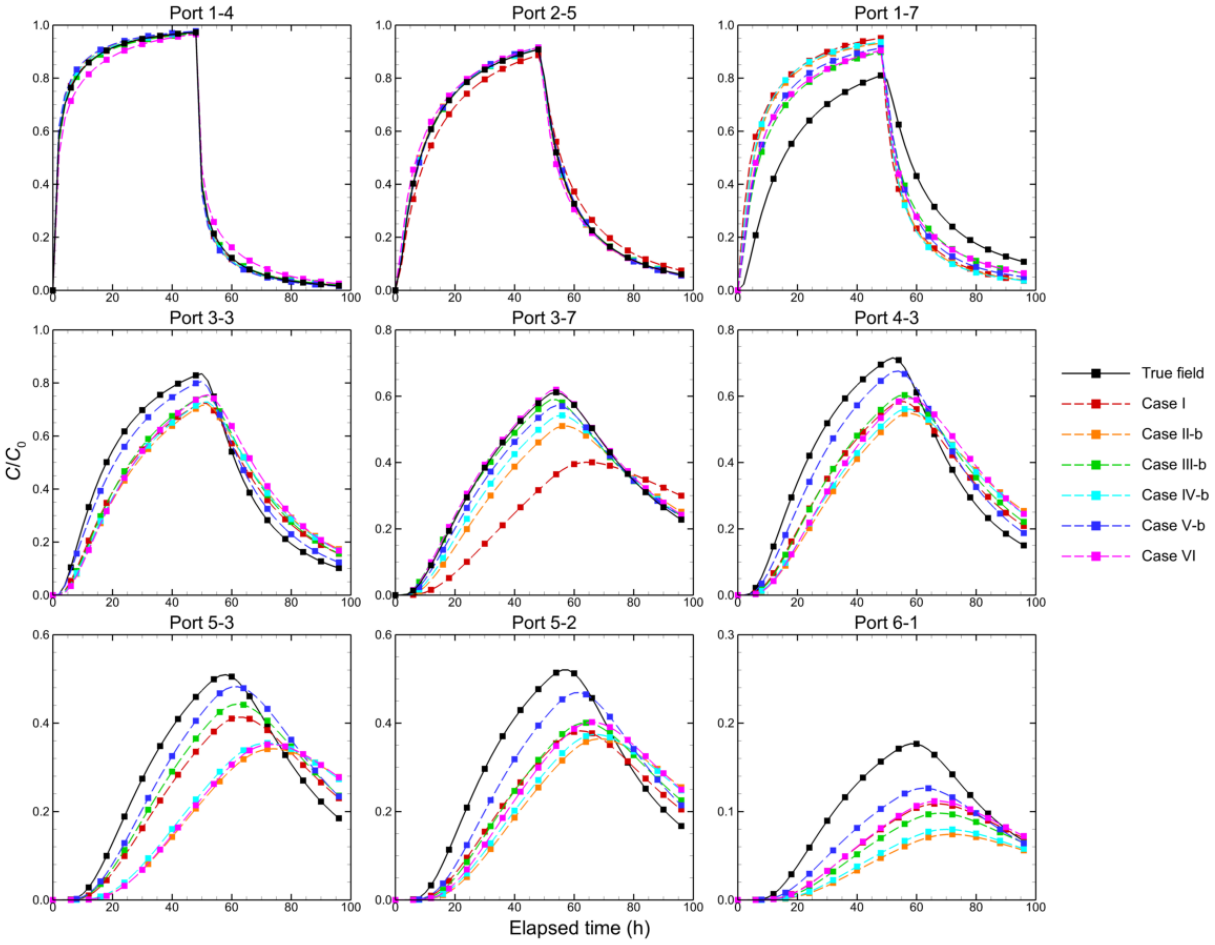


Figure 3.11: Breakthrough curves of nine sampling ports during 96 hours of tracer injection and dissipation. The upper three ports (Ports 1-4, 2-5 and 1-7) are near the source, the intermediate three ports (Ports 3-3, 3-7 and 4-3) are at intermediate distances, and the lower three ports (Ports 5-3, 5-2 and 6-1) are at large distances from the influent port.

Based on the BTCs in Figure 3.11, no case leads in the prediction performance at all sampling ports, but the BTCs of Case V-b fit the observation (i.e., BTCs based on the true K field) better in most sampling ports. More importantly, we notice that differences among the cases are marginal at sampling ports very close to the influent port, but larger differences arise as the tracer cloud migrates through the highly-heterogeneous aquifer-aquitard system. The prediction performances vary more significantly, and the leading performance of Case V-b is more prominent for the BTCs of the sampling ports far away from the influent port (see the bottom three BTCs in Figure 3.11). Therefore, it reveals that differences in tracer migration prediction emerge as it encounters more heterogeneity along the migration paths.

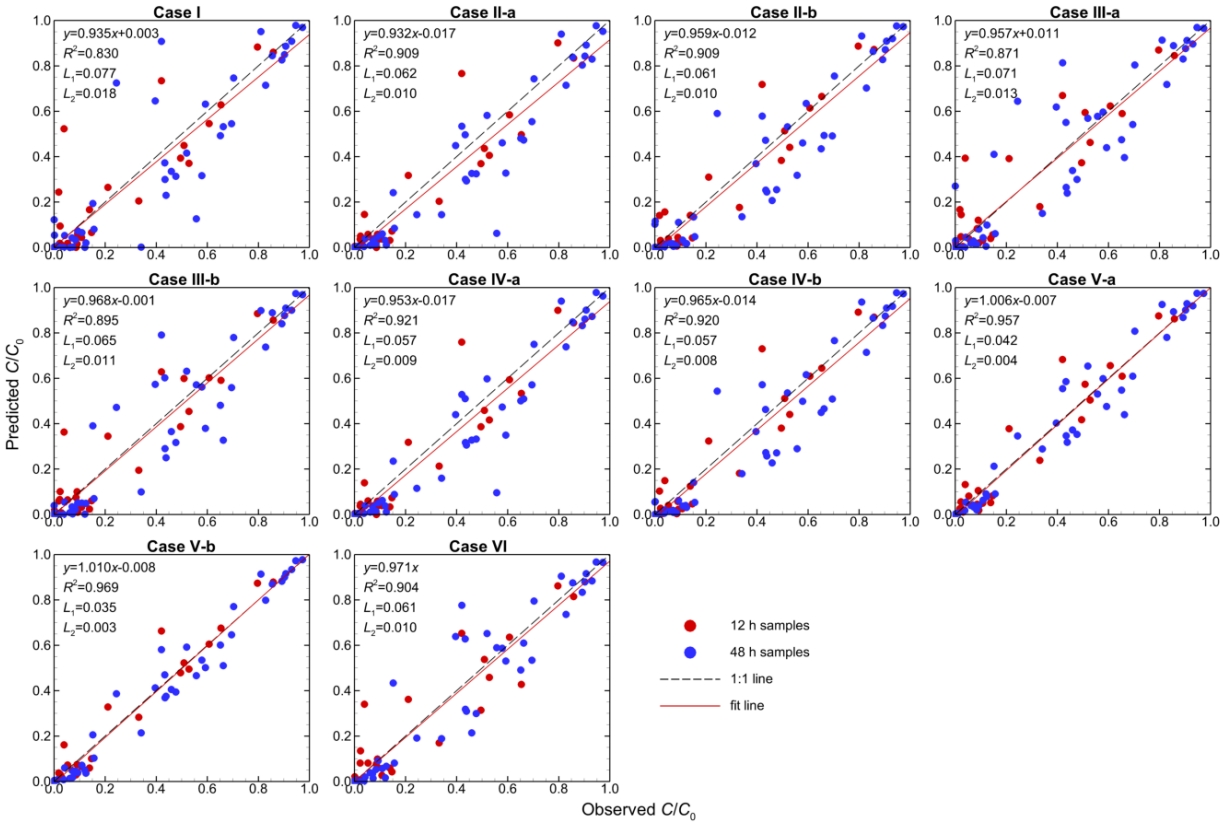


Figure 3.12: Scatterplots of predicted versus observed C/C_0 at 12 h and 48 h at 47 sampling ports for all cases. R^2, L_1 and L_2 norms are indicated on each plot.

In addition to BTCs, further quantitative comparison of simulated versus predicted C/C_0 at two timesteps is shown in Figure 3.12. Generally, clear improvements on model performance metrics (i.e., L_1, L_2 and R^2) occur when the initial guess for inversion from geophysical survey data is included and when both spatial constraints in clustering and point measurements for zonal estimation are provided for more accurate zonation models. Nevertheless, the temporal evolution of tracer clouds cannot be fully captured with only snapshots of concentration distributions. Next, temporal moment analysis is applied to further characterize the breakthrough data at all sampling ports.

The n^{th} temporal moments (M_n) of concentration (C) at location (x, y) at time (t) are given by:

$$M_n = \int_0^\infty t^n C(x, y, t) dt \tag{3.9}$$

The zeroth (M_0), first (M_1) and second (M_2) temporal moments can be obtained using the moment generation function above by setting n as 0, 1, and 2, respectively. The temporal moments are computed through numerical integration of breakthrough data using the trapezoidal rule. Note that concentration records at all 47 sampling ports (i.e., exclusive of the influent port) are included.

The total mass of solute (M_0) passing through the sandbox aquifer at a time interval of 2 h is obtained. Figure A8 shows the scatterplots of simulated versus observed M_0. Meanwhile, the first normalized moment of BTCs at all sampling ports are used to estimate the mean arrival time (μ) of the center of tracer mass, and the variance (σ^2) of BTCs are also calculated by:

$$\mu = \frac{M_1}{M_0} \tag{3.10}$$

$$\sigma^2 = \frac{M_2}{M_0} - \left(\frac{M_1}{M_0}\right)^2 \tag{3.11}$$

Figures A9 and A10 show the scatterplots of simulated versus observed μ and σ^2 for all cases, respectively. Figure 3.13 provides a summary of the computed L_1 and L_2 norms for M_0, μ, and σ^2.

Case	M_0 (h)		μ (h)		σ^2 (h^2)		Rank
	L_1	L_2	L_1	L_2	L_1	L_2	
I	5.50	63.40	5.88	59.95	51.69	6334.38	10
II-a	4.25	35.97	4.07	26.18	34.35	2360.31	6
II-b	4.55	35.33	5.31	40.35	47.18	3783.29	9
III-a	4.84	46.68	5.09	40.05	39.49	3188.39	8
III-b	4.52	37.14	4.97	34.93	40.25	2952.38	7
IV-a	3.85	29.57	4.00	24.34	33.84	2142.53	4
IV-b	3.94	26.51	4.77	32.62	41.08	2743.10	5
V-a	2.38	9.61	2.89	12.02	21.20	728.44	2
V-b	2.34	9.74	2.84	11.08	24.45	998.18	1
VI	3.42	24.60	4.45	28.29	32.86	1560.87	3

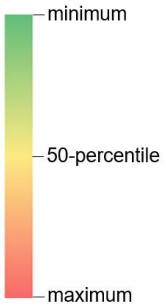


Figure 3.13: Summary of the color-coded L_1 and L_2 norms for M_0, μ and σ^2 for all cases. A ranking of all cases is provided on the right-most column.

The L_1 and L_2 norms of M_0, μ, and σ^2 in Figure 3.13 show similar trends for different cases. Cases V-b and V-a rank the first two places among all cases. Next, the involvement of local K measurements in Cases IV-a and IV-b does not make substantial improvements compared to Cases II-a and II-b. It reveals that a more reliable zonation model for the initial guess of inversion relies on both accurate delineation of zone geometry and estimation of zonal parameters. Differences in model performance metrics of Cases II-a, II-b, III-a and III-b are insignificant, but these four cases perform consistently better than Case I. Besides, the performance of Case III does not exceed Case II, which is different from the findings in drawdown predictions. It may result from the inaccurate characterization of high-K Layer 4, which is a critically long aquifer having a significant contribution to solute migration. Likewise, the inaccurate characterization of Layer 4 renders the non-ideal performance of Case VI.

To further evaluate the solute transport prediction performances of different cases, simulation of an additional tracer test is performed, in which tracer is injected at Port 6-1 at a rate of 0.1 mL/s, while water is extracted at Port 1-5 at a rate of 0.2 mL/s. The other simulation parameters remain the same as the primary tracer test. A detailed description of this additional tracer test is in Text S3. C/C_0 distributions

at various timesteps for all cases are provided in Figure A11, and all other plots regarding the solute transport analyses of this tracer test are in Figures A12–A18. Different from prediction of drawdown responses in pumping tests, in which signals propagate from pumping to observation ports, tracer clouds migrate from the influent to the effluent ports, and it evolves over a large extent of the simulation domain with elapsed time. As the tracer clouds of the two tests cover the majority of the simulation domain, it is reasonable to state that the conclusions drawn from the results of these two simulations are representative of the solute transport prediction performances of Cases I–VI. Based on the analyses of solute transport predictions for the additional tracer test, Case II-b performs even poorer than Case I (see Figures A13 and A18). It demonstrates that the incorporation of spatial constraints in clustering improves the reliability of zonation models. The other findings from the analyses of this additional tracer test corroborate the conclusions from the primary test.

4 Comparison of NMR-derived Hydraulic Conductivity with Various Hydraulic Testing Methods

Citation:

Wang, C., Steelman, C. M., Ning, Z., Walsh, D. O., & Illman, W. A. (2025a). Comparison of NMR-Derived Hydraulic Conductivity with Various Hydraulic Testing Methods. *Groundwater*, *63*(5), 713–724. https://doi.org/10.1111/gwat.70016

4.1 Overview

Borehole Nuclear Magnetic Resonance (NMR) can be used to estimate the hydraulic conductivity (*K*) of unconsolidated materials. Various petrophysical models have been developed to predict *K* based on NMR response, with considerable efforts on optimizing site-specific constants. In this study, we assess the utility of NMR logs to estimate *K* within highly heterogeneous glaciofluvial deposits by comparing them with three types of co-located hydraulic testing methods, including permeameter, multi-level slug, and direct-push hydraulic profiling tool (HPT) logging tests. Four NMR models, including Schlumberger-Doll Research (SDR), Seevers, Sum-of-Echoes (SOE), and Kozeny-Godefroy (KGM), were applied to construct *K* profiles at four locations with model constants optimized using permeameter-based *K*. Model constants suitable for glaciofluvial deposits were provided. Results showed that NMR logging can provide reliable *K* estimates for interbedded layers of sand/gravel, silt, and clay. Through cross-hole comparison of NMR-derived *K* profiles, *K* trends and magnitudes of aquifers/aquitards were readily characterized. Quantitatively, the NMR-derived *K* coincided with hydraulic-testing *K*, with optimal model fits within one order of magnitude. We noticed that (1) Seevers performed similarly but no better than SDR in predicting permeameter and slug testing measurements; (2) SOE yielded slightly better predictions than SDR; (3) the removal of porosity in SDR did not deteriorate its prediction, and the

optimized SDR constant resembled the literature-based values for glacial deposits; and (4) KGM yielded the optimal fits with slug-based *K*, demonstrating its reliable performance. Lastly, we made recommendations on selecting suitable petrophysical models.

4.2 Introduction

Borehole Nuclear Magnetic Resonance (NMR) tools have been used in the petroleum industry for over 50 years to determine formation properties such as permeability and porosity. With the development of slim-hole NMR logging tools (e.g., Sucre et al., 2011; Perlo et al., 2013; Walsh et al., 2013; Zhu et al., 2021), applications have extended to the characterization of unconsolidated sediments within the environmental and engineering sectors. Borehole NMR for unconsolidated sediments provides high-resolution site characterization of hydraulic properties (e.g., hydraulic conductivity (*K*) and water content) and features a fast in-situ application deployed either in wireline logging or through a direct-push technique (Morozov et al., 2024). Unlike conventional neutron source tools, NMR does not have an active nuclear source, making it easier to deploy in the environmental sector.

Various NMR models have been developed to estimate *K* from NMR signals. Most models are empirically derived with model constants acting as fitting parameters (e.g., Schlumberger-Doll Research (SDR), Seevers, and Sum-of-Echoes (SOE)). The selection of NMR signals (e.g., porosity and relaxation time) in NMR models also affects the quantitation of model constants. For instance, Maurer and Knight (2016) found that NMR-derived porosity does not improve the *K* prediction in the SDR model. In addition, attempts were made to replace empirical with physical parameters, e.g., the Kozeny-Godefroy model (KGM, see Dlugosch et al., 2013).

Several studies have explored the predictive performance of these models across different geological conditions and have found that model constants in unconsolidated materials often differ significantly from consolidated materials due to varying sedimentation and tortuosity. Therefore, the site-

specific optimization of NMR models through calibration with hydraulically derived *K* measurements is usually performed for more reliable *K* estimation. Dlubac et al. (2013) was the first study to optimize model constants for unconsolidated materials, with numerous subsequent studies presenting unique optimizations based on site conditions (Crow et al., 2022; Kendrick et al., 2021, 2023; Knight et al., 2016; Maurer & Knight, 2016; Parsekian et al., 2015). The comparison of NMR-derived *K* with hydraulic testing methods in these previous studies has generally shown good agreement. For example, Knight et al. (2016) optimized the SDR constants with *K* calibration datasets from three sites within alluvial deposits (fluvial, riparian, agricultural sediments) and found only minor differences in the SDR constants. This finding indicates the potential of NMR logging to provide *K* estimates with reasonable uncertainty for unconsolidated materials when site-specific calibration is not available. However, on occasion, model constants have been shown to be inconsistent between sites, indicating potential limitations in model utility, particularly in lower-*K* materials. For instance, Kendrick et al. (2023) have compiled model constants for SDR and SOE from the literature (Dlubac et al., 2013; Kendrick et al., 2023; Knight et al., 2016; Walsh et al., 2013), which were calibrated to different types of sediments. Results show that these model constants can vary over an order of magnitude across different geological materials.

Another potential limitation of existing model constants is the nature of the hydraulic testing method used to calibrate NMR models. Various hydraulic testing methods have been used to calibrate these empirical models, including direct-push permeameter (DPP) tests, wellbore flowmeter (WBF) logging tests, and multi-level slug tests. Varying scales and components of *K* are measured and related to NMR response, potentially leading to a study-specific calibration bias. For example, DPP tests (Butler et al., 2007) have been applied in several NMR comparison studies (e.g., Maurer and Knight, 2016; Kendrick et al., 2021, 2023). They are conducted by injecting a stream of water horizontally and measuring the pressure difference at two transducers with a spacing of ~0.4 m vertically for *K* estimation by Darcy's law, assuming pressure equality at the same distance spherically. Furthermore, NMR data have

been compared with WBF logging tests by Dlubac et al. (2013), which measure the effective horizontal *K* component at meter scale, assuming that the aquifer is uniformly layered and *K* is proportional to the inflow rate at each layer. Multi-level slug tests were utilized to fit with the SDR model in Walsh et al. (2013) and Crow et al. (2022). *K* measurements from slug tests are biased towards the horizontal component, assuming radial flow toward the screen. Various hydraulic testing methods can cause considerable differences in the optimization of NMR-derived *K* estimates, with implications for predictions of groundwater flow and contaminant transport. Sun et al. (2024) evaluated *K* estimates from various hydraulic testing methods, including grain-size analysis, multi-level slug tests, permeameter tests, direct-push hydraulic profiling tool (HPT) logging, and hydraulic tomography (HT), by comparing observed drawdown from pumping tests with numerically modelled drawdown. Results show that different hydraulic testing methods have led to varying drawdown prediction performances.

Unlike conventional hydraulic testing methods, a NMR probe stimulates the porewater within a circular distance around the borehole and measures the mobility of the hydrogen within the pore space to infer *K* without assuming any potential directional component of *K*. However, given the heterogeneous and anisotropic characteristics of geologic material, hydraulic testing methods often yield different representative estimates of *K*. Therefore, it is valuable to investigate how NMR-derived *K* models fit with co-located *K* measurements from different hydraulic testing methods, in which varying scales and components of *K* are sampled.

In this work, we examined the capability of NMR logging to characterize subsurface heterogeneity and estimate *K* in highly heterogeneous glaciofluvial deposits ranging from sand/gravel, silt, and clay, which exhibit more than a five-order range in *K* (m/s). Permeameter-based *K* measurements sampled at a similar interval to NMR logging were used as the *K* calibration dataset. Then, we assessed the predictive performance of NMR logging with two conventional methods, including permeameter tests and multi-level slug tests, and another in-situ logging method, the direct-push hydraulic profiling tool

(HPT) tests, by comparing their co-located *K* profiles. Four NMR petrophysical models, including SDR, Seevers, SOE, and KGM, were used to quantitatively evaluate model fits with physical *K* measurements. The calibrated model constants were listed and compared to the literature-based constants from various geological conditions. Recommendations on selecting an optimal NMR model were provided for a highly heterogeneous glaciofluvial environment.

4.3 Background

4.3.1 Borehole NMR Basics

A net nuclear spin angular momentum is possessed by the hydrogen protons in water molecules, which is used in the geophysical application of NMR methods (see review by Behroozmand et al., 2015). In terms of borehole NMR measurements, an NMR probe is lowered down an open bedrock hole or a PVC-lined monitoring well. A permanent magnet built into the probe is utilized to generate a static magnetic field, B_0. Proton spins become polarized in an equilibrium state aligned parallel to B_0, which is defined as the longitudinal direction. Hydrogen protons precess at the Larmor frequency in the B_0 field, with the precession frequency and the magnitude of the magnetization vector proportional to the B_0 field strength. A pulsed oscillating magnetic field, B_1, tuned to a specific Larmor frequency, is applied in the plane transverse to B_0 by the NMR probe to rotate the magnetization of the hydrogen protons at a desired radial distance from the NMR probe. A cylindrical shell of deposits is sampled during this process. When proton spins return to the equilibrium state (i.e., relaxation), a spin-echo decay is detected by the receiving coil. For multiple pores with varying geometry, the spin-echo decay curve, $S(t)$, can be described as a multiexponential decay function:

$$S(t) = \sum_i A_{0i} e^{-t/T_{2i}} \tag{4.1}$$

where i is the pore index, A_{0i} is the initial amplitude for each proportion of T_{2i}, and T_{2i} is a characteristic time (i.e., the relaxation time) for pores with different T_2. A_0 can be normalized by the initial amplitude of the spins in pure water to quantify water content. The $S(t)$ signal can be processed using a Laplace inversion to obtain the multiexponential decay time distribution, showing the amplitudes as a function of T_2.

The measured T_2 is a linear combination of three relaxation mechanisms (Kleinberg & Horsfield, 1990):

$$T_2^{-1} = T_{2S}^{-1} + T_{2B}^{-1} + T_{2D}^{-1} \tag{4.2}$$

where T_{2S} is the surface relaxation time, T_{2B} is the bulk fluid relaxation time, and T_{2D} is the diffusion-gradient relaxation time, since refocusing pulses have a reduced efficiency, which can be mitigated using a sufficiently short echo time (Walsh et al., 2013). The surface relaxation time, T_{2S}, ranges from several milliseconds to a second, while T_{2B} is around 1–3 s for water. The surface relaxation time typically dominates the relaxation response in unconsolidated materials, corresponding to the whole pore that is sampled during relaxation. In such a fast diffusion regime, T_{2S}^{-1} (with a unit of 1/s) can be related to the geometry of the pore in the following manner (Brownstein & Tarr, 1979):

$$T_{2S}{}^{-1} = \rho \frac{S}{V} \tag{4.3}$$

where ρ is a material-specific surface relaxivity (m/s), governed by the presence of paramagnetic ions on the pore surface, measuring its ability to enhance relaxation, and S/V is the surface area to volume ratio of the pore space (1/m).

4.3.2 Models for K Estimation

Empirical relationships of NMR models follow the Kozeny-Carman type of equations. One form of the K-C equation can be written as follows to estimate hydraulic conductivity, K (Seevers, 1966):

$$K = \frac{\varrho g \phi}{\mu \tau (\frac{S}{V})^2} \tag{4.4}$$

where ϱ is water density, g is the gravitational acceleration, ϕ is porosity, μ is dynamic viscosity, and τ is tortuosity. The relationship between T_2 and S/V in Equation (4.3) is utilized to estimate K.

Therefore, T_2 is the main indicator of K in the SDR model (Kenyon et al., 1988), which can be written as:

$$K_{SDR} = C_{SDR} \phi^m T_{2ML}^n \tag{4.5}$$

where C_{SDR}, m, and n are site-specific parameters, in which the C_{SDR} constant mainly accounts for mineralogical parameters (e.g., surface relaxivity). T_{2ML} is the mean-log T_2 used to characterize the T_2 distribution:

$$T_{2ML} = exp\,(\frac{\sum A_i ln\,(T_{2i})}{\sum A_i}) \tag{4.6}$$

Generally, n is assigned to be 2, which follows the assumption of the fast-diffusion regime. For consolidated materials, m is originally assigned a value of 4 (Kenyon et al., 1988) to account for variations in the ratio of pore throats to pore openings. In recent studies for unconsolidated aquifers, m = 0 (Crow et al., 2022; Maurer & Knight, 2016), m = 1 (Kendrick et al., 2023; Knight et al., 2016), or m = 2 (Dlubac et al., 2013) were usually adopted. Here, m = 0 was adopted such that porosity is removed from the SDR equation.

The Seevers model (Seevers, 1966) has a similar form to SDR. The main difference is that the contribution of bulk fluid relaxation is incorporated independently:

$$K_{Seevers} = C_{Seevers}\phi^m[({T_{2ML}}^{-1} - {T_{2B}}^{-1})^{-1}]^n \tag{4.7}$$

where C_{Seevers} is the Seevers constant, while $m = 1$ and $n = 2$ are typically utilized. The relaxation time of bulk water (in seconds), T_{2B}, can be estimated as follows (Dlugosch et al., 2013):

$$T_{2B} = 3.3 + 0.044(\theta - 35) \tag{4.8}$$

where θ is the temperature in degrees Celsius.

The SOE model estimates K as a linear function of the squared spin echo decay curve, $S(t)$. The SOE signal is defined as:

$$SOE = \int_0^\infty S(t)dt \tag{4.9}$$

and K can be obtained by:

$$K_{SOE} = C_{SOE}(SOE)^n \tag{4.10}$$

where C_{SOE} is the SOE constant, and n is the exponent normally assigned to be 2 (e.g., in Walsh et al., 2013). As the integral of $S(t)$ is used to substitute $T_{2\text{ML}}$ from inversion, the SOE model is theoretically more reliable in a high-noise environment.

In KGM, the material-specific NMR parameters are utilized to replace the empirically derived factors (Dlugosch et al., 2013):

$$K_{KGM} = \frac{\varrho g}{8\tau^2\mu}\phi\left(-\frac{D}{\rho} + \sqrt{\left(\frac{D}{\rho}\right)^2 + \frac{4DT_{2B}T_2}{T_{2B} - T_2}}\right)^2 \tag{4.11}$$

where D is the diffusion constant of bulk water (in m^2/s), which can be estimated as follows (Maurer & Knight, 2016):

$$D = (1.0413 + 0.039828\theta + 0.00040318\theta^2) \times 10^{-9} \quad (4.12)$$

4.4 Site and Data Descriptions

4.4.1 The North Campus Research Site

The NMR data were collected at the North Campus Research Site (NCRS) located on the University of Waterloo campus in Waterloo, Ontario, Canada. Previous studies have shown that the shallow surface consists of glaciofluvial deposits and till ranging from gravel/sand, silt, and clay, forming alternating layers of aquifer and aquitard units (e.g., Alexander et al., 2011; Berg and Illman, 2011b; Zhao and Illman, 2017, 2018). A dense and stiff Catfish Creek Till located at approximately 16 meters below ground surface (mbgs) is regarded as a hydraulic barrier at the bottom of the aquifer system of interest. A sandy-to-clayey silt sequence, namely the Maryhill Till, is deposited over the Catfish Creek Till. Thin erosional remnants of the Tavistock Till overlie the Maryhill Till, which is below the top organic soil. Alternating layers of aquifer and aquitard units were found based on borehole geology logs (refer to Alexander et al., 2011; Berg and Illman, 2011b). The primary "aquifer zone" is situated between 8 and 13 mbgs, with low K materials separating two aquifer layers. The upper aquifer layer consists of sand to sandy silt, and the lower aquifer layer is composed of sandy gravel.

The NCRS has nine multi-level wells configured in a square pattern (Figure 4.1a). Four Continuous Multichannel Tubing (CMT) wells were installed at the midpoint of each side. CMT tubes are 18 m long with seven channels to varying depths in the formation. The CMT tubing has a diameter of 0.10 m. A 0.17 m section was cut for each channel at the desired depth and wrapped with a fine mesh to build a discrete observation point with a ~2 m spacing between adjacent points.

The drilling of the CMT wells was completed using a hollow-stem auger with a diameter of 0.26 m. Continuous samples of sediments were obtained during the drilling process of each CMT well for core logging and permeameter testing in the lab. During installation of the CMT wells, each observation port was encased in 0.3 m to 0.6 m filter packs consisting of #0 silica sand filled above and below each observation port, with bentonite pellets used to isolate the filter packs.

4.4.2 NMR Logging

A slim-hole NMR logging system (Dart^{TM} DP140N, Vista Clara Inc.) was used to collect borehole NMR data adjacent to the CMT wells. The distance between NMR logs and existing CMT wells was < 1 m. The DP140N slim-hole probe is specifically designed to work with a Geoprobe direct-push rig. The DP140N has a diameter of 3.56 cm and a length of 1.33 m. The vertical resolution of the DP140N is 0.22 m. Two coils are used to generate a cylindrical response of 12.7 cm and 15.2 cm radially from the center of the probe. Together, these measurements can isolate the signal from formation material within a thin cylindrical shell outside the drilling alteration zone. The measurement spacing of the four CMT logs was 0.15 m.

An illustration of how the Dart works with the direct-push rig is shown in Figure 4.1b. First, hollow-stem rods with a closed cutting shoe are advanced to the desired depth by the percussion hammer. The shoe is then opened, and the DP140N is inserted through the rods until it reaches the bottom of the hole. The rods are retracted while the probe remains fixed. As the sensor exits the bottom of the casing and becomes exposed to the formation, a specially designed key at the top of the probe lifts it upward while remaining fully exposed to the formation. NMR measurements are collected in a stepwise manner with the retraction of the drill rods performed in pre-defined vertical sampling increments (e.g., 0.15 m). Figure 4.1c is a field photo of NMR logging at CMT1. The NMR logging deployed with direct-push rigs enables the profiling of soil conditions without the need for permanent well installations or potential

impacts of well construction materials (e.g., bentonite seals). To date, the Dart is the only NMR logging system that can be deployed with the direct-push rig.

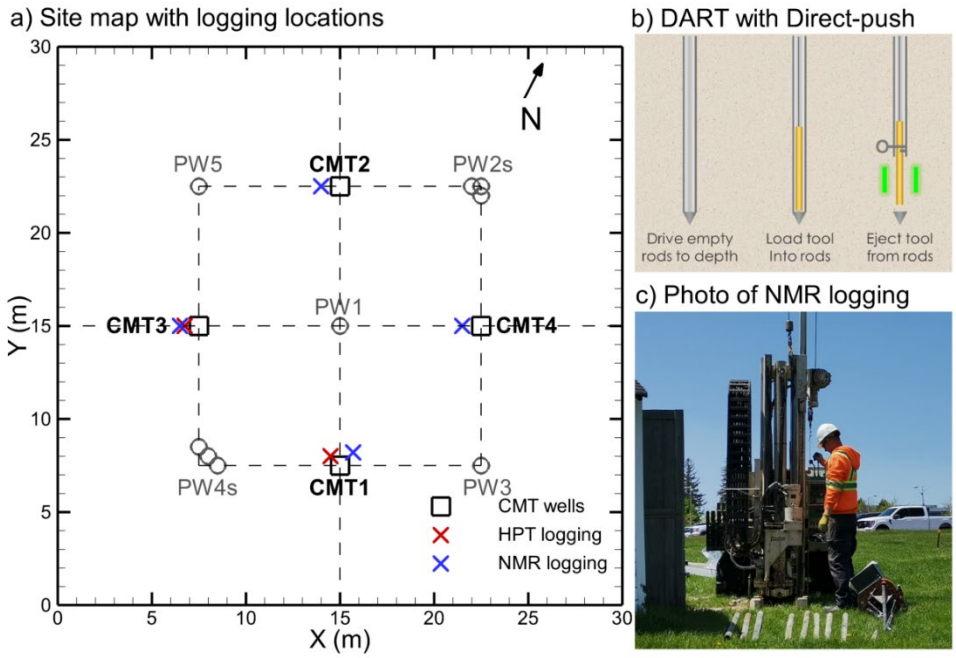


Figure 4.1: Field site and NMR logging information. a) A site map with locations of pumping wells, HPT logs, and NMR logs. The geographic coordinates of CMT1 are 43°28'36.4" N, 80°32'44.6" W. The other wells at the NCRS not involved in this study, including PW1, PW2s, PW3, PW4s, and PW5, are labelled in grey. b) Illustration of how the Dart system works with the direct-push technique. c) A field photo of NMR logging with the direct-push rig proximal to CMT1 at the NCRS.

In total, 211 discrete NMR measurements were collected at the four CMT wells. In our NMR dataset, the signal-to-noise ratio was always less than 10%, indicating high-quality measurements. The NMR signals from CMT1 to CMT4 are plotted in Figure 4.2. Three signals, including T_{2ML}, *SOE*, and total water content (θ), are provided as potential indicators for K estimation utilized in petrophysical models. In terms of water level measurements for all pumping wells at the NCRS (see Figure 4.1a), the lowest water level measured on the day of the NMR logging survey was 3.78 mbgs, so we use 3.78 mbgs as the upper limit of the depth for all NMR-derived K. The depth of exploration approached 12 mbgs for all four NMR logs. In addition, the geologic units for the four CMT wells are shown in each plot for

comparison, which were compiled from the original borehole logs by Zhao and Illman (2017) and utilized in subsequent studies (Zhao and Illman, 2018; Luo et al., 2023; Zhao et al., 2023).

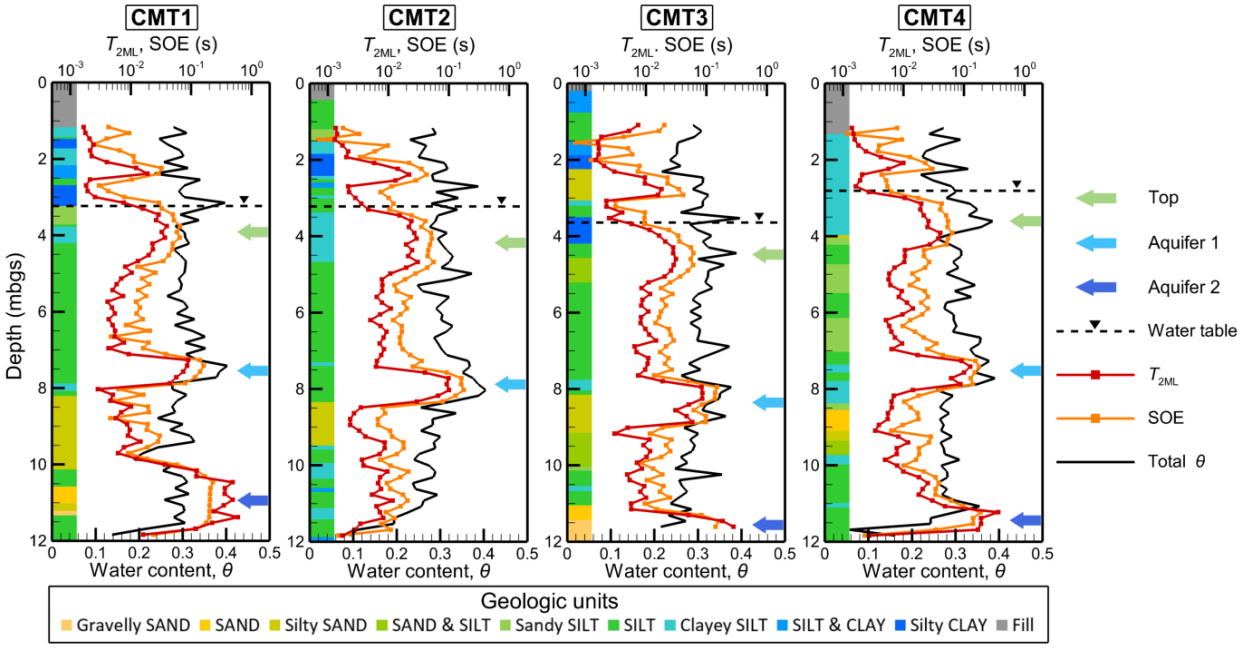


Figure 4.2: Profiles of NMR signals near CMT1 to CMT4. NMR signals include the main indicators of *K* in the petrophysical models (i.e., T_{2ML} and *SOE*) and total water content. Geologic units are also shown for comparison, while the water table is indicated with a horizontal dashed line. The locations of three conductive layers are indicated using colored arrows.

4.5 Hydraulic Testing Methods

Previously reported hydraulic testing results collected at the four CMT wells of the NCRS were used to assess the performance of NMR logging. The hydraulic testing data used to compare with NMR *K* estimates and the statistical parameters of *K* estimates for each hydraulic testing method are summarized in Table 4.1.

4.5.1 Permeameter Tests

For permeameter tests, a 10-cm diameter split spoon sampler was used to collect continuous vertical core samples at an interval of around 10 cm during the drilling of CMT wells. The core samples

were delivered to the laboratory for falling-head permeameter tests by Alexander et al. (2011). Within the depth range of this study, 248 permeameter *K* measurements were available.

The sampling interval of permeameter tests is 10 cm, and the sampling interval of NMR logging is 15 cm. For statistical analysis, we paired NMR-derived *K* estimates with permeameter-based *K* measurements, which were co-located within 10 cm, resulting in 171 paired *K* measurements for CMT1, CMT2, CMT3, and CMT4.

Table 4.1: Summary of the number of hydraulic measurements at each location and statistical parameters of *K* estimates from each hydraulic testing method, including the geometric mean of *K*, the variance of ln *K*, and the maximum and minimum *K* values.

Methods	Number of measurements					Statistical parameters			
	Total	CMT1	CMT2	CMT3	CMT4	$\overline{K}$ (m/s)	$\sigma^2_{\ln K}$	Max *K* (m/s)	Min *K* (m/s)
Permeameter	248	64	69	60	55	7.16e-7	5.66	2.81e-4	2.17e-9
Slug tests	16	4	4	4	4	4.76e-6	4.31	1.04e-4	1.66e-7
HPT logging	1078	539	-	539	-	1.32e-5	1.56	1.95e-4	4.60e-7

4.5.2 Multi-Level Slug Tests

The four CMT wells each have four channels within the range of NMR logging, providing 16 sampling ports for which Alexander et al. (2011) completed slug tests. A slug of water was injected into each port for falling-head tests, which induced a 0.5 m rise in the static water level. The Hvorslev method Hvorslev, 1951 was utilized to analyze the slug test data.

CMT channels were installed at an interval of 2 m. It is assumed that the *K* measurements from slug tests were representative of sediment conditions over a total depth of 2 m proximal to the ~0.2 m screen and that flow was isotropic. Given the larger sampling distance of the slug tests, we upscaled the NMR-derived *K* by averaging the NMR samples within the corresponding 2-m depth range. We calculated the geometric mean of the six NMR samples above and below the midpoint of each screen. In summary, 16 co-located *K* estimates were obtained, associated with the slug tests. Due to the limited

depth range of NMR logging, fewer NMR samples were available to match the slug *K* measurements in some screens near the top or bottom of the downhole profiles.

4.5.3 HPT Logging Tests

The HPT logging test is a direct-push injection logging technique. The HPT probe injects a flow of water into the formation while recording the injection pressure, flow rate, and the probe advance rate during the logging process (McCall & Christy, 2020). A physically-based equation modified from Borden et al. (2021) was used to estimate *K* (m/s):

$$K = E(6.113 \times 10^{-8} V D^2 + 5.891 \times 10^{-8} Q) P_c^{-1.017} \quad (4.13)$$

where *E* is an empirically derived hydraulic efficiency factor, *V* is the probe advance rate (cm/s), *D* is the probe diameter (cm), *Q* is the water injection rate (mL/min), and P_c is the corrected injection pressure (kPa). The best-fit value of *E* after being validated by 23 HPT profiles is 2.02 by Borden et al. (2021).

The co-located HPT measurements to NMR logs were near CMT1 and CMT3 (see Figure 4.1a). The sampling interval of direct-push HPT logging tests was 1.5 cm. Table 4.1 illustrates that the $\sigma^2_{\ln K}$ of HPT-based *K* was significantly smaller than the other two methods, which may be due to the fact that both upper and lower boundaries of *K* estimation were limited. Liu et al. (2012) pointed out that the upper boundary of *K* in HPT logging tests is around 6.9×10^{-4} m/s, over which the injection rate is too small to produce a measurable pressure response. On the other hand, the maximum detected injection pressure in the HPT sensor is around 689 kPa, which also limits the HPT technique to estimate *K* in low-permeable materials. In our dataset of HPT logging, the lower boundary of *K* is 4.60×10^{-7} m/s. Therefore, HPT logging was only adopted for qualitative comparison of *K* profiles with NMR logging.

4.6 Model Calibration

Model constants were calibrated using the K measurements from permeameter tests. A linear regression after log transform was performed to calibrate the model constants in the SDR, Seevers, and SOE model. As a non-linear relationship exists between K and the model constants in KGM, a grid search analysis (Kendrick et al., 2023; Maurer & Knight, 2016) was conducted to find the optimal constants, in which thousands of forward model runs were performed to find the optimal constants, which corresponded to the least root mean squared error (RMSE). Wide ranges of ρ (i.e., 1-300 μm/s) and τ (i.e., 1 to 4) were applied to cover the possible values. Figure 4.3 shows the contour maps of RMSE distribution in the grid search analysis. We found that RMSE gradually decreased with increasing ρ and τ. Dlugosch et al. (2013) found that the KGM-estimated K is not sensitive to the variation of τ. Here, we fixed τ to 1.5 and sought an optimal ρ for the glaciofluvial sediments at the NCRS. The optimized model constants are summarized in Table 4.2. As porosity is still adopted in the SDR equation for some studies, we also provided the calibrated C_{SDR} when m = 1. However, this scenario was not included in the comparison below.

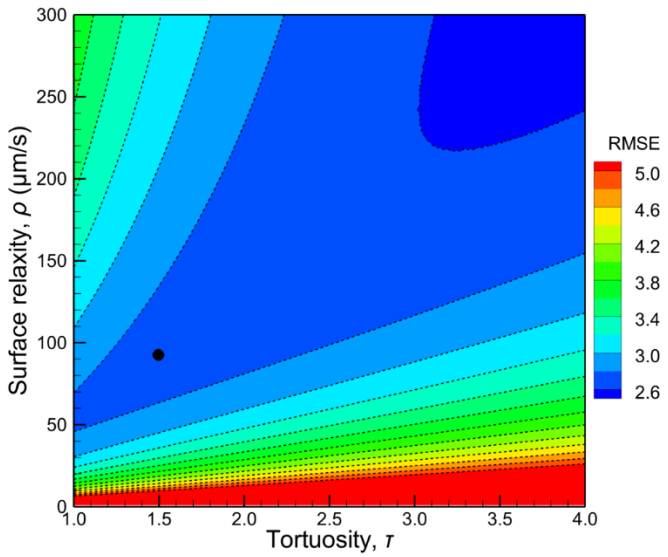


Figure 4.3: Contour maps of RMSE distribution from the grid search analysis for the calibration of the KGM model, in which the smallest RMSE leads to the optimized values for model constants. The black spot denotes the optimal ρ with the smallest RMSE when $\tau = 1.5$.

Table 4.2: Summary of optimized constants after calibration with permeameter test data

NMR model	Optimized model constants		Other parameters
	Data domain	Well scale	
SDR	C_{SDR} = 3.86e-3 m/s^3	C_{SDR} = 2.70e-3–7.21e-3 m/s^3	$m = 0$, $n = 2$
SDR*	C_{SDR} = 0.0130 m/s^3	C_{SDR} = 0.0093–0.0244 m/s^3	$m = 1$, $n = 2$
Seevers	$C_{Seevers}$ = 0.0127 m/s^3	$C_{Seevers}$ = 0.0091–0.0241 m/s^3	$m = 1$, $n = 2$
SOE	C_{SOE} = 9.02e-4 m/s^3	C_{SDR} = 5.71e-4–1.43e-3 m/s^3	$n = 2$
KGM	ρ = 91 μm/s	ρ = 71–127 μm/s	$\tau = 1.5$

* represents the SDR equation containing porosity.

4.7 Results and Discussion

4.7.1 NMR-Derived K Profiles

Figure 4.4 shows the *K* profiles from four NMR models and three types of hydraulic testing methods. Only the *K* estimates in the saturated zone (i.e., below the water table) were incorporated into this figure. According to the SDR estimates, the highest and lowest *K* values were 1.36×10^{-3} m/s and 1.16×10^{-8} m/s, respectively, exhibiting a *K* variation of more than five orders of magnitude.

Based on the visual observations of vertical *K* trends in Figure 4.4, three conductive layers can be identified from matching moderate-to-high *K* segments of NMR-derived *K* profiles. A moderately conductive layer was found right below the water table (defined as the Top layer) with a mean *K* of 10^{-6} m/s. Furthermore, two main aquifers can be characterized: the first one (defined as Aquifer 1) was located at ~8 mbgs, with an average *K* of 10^{-5} m/s; the second aquifer (defined as Aquifer 2) was located from 10–12 mbgs, with an average *K* of 10^{-4} m/s. Aquifer 1 was thin (~1 m thick) but laterally continuous, dipping slightly toward the west. Aquifer 2 exhibited a more variable layer thickness, with the thickest sediments near CMT1. It dipped sharply to the northwest and eventually pinched out in CMT2. Based on the SDR estimation, the segments in the *K* profiles (see Figure 4.4) with *K* values over 1×10^{-5} m/s and 1×10^{-6}

m/s are shaded to highlight the highly and moderately conductive layers, respectively. Hence, the Top layer, Aquifer 1, and Aquifer 2 are shaded in red, yellow, and blue in Figure 4.4.

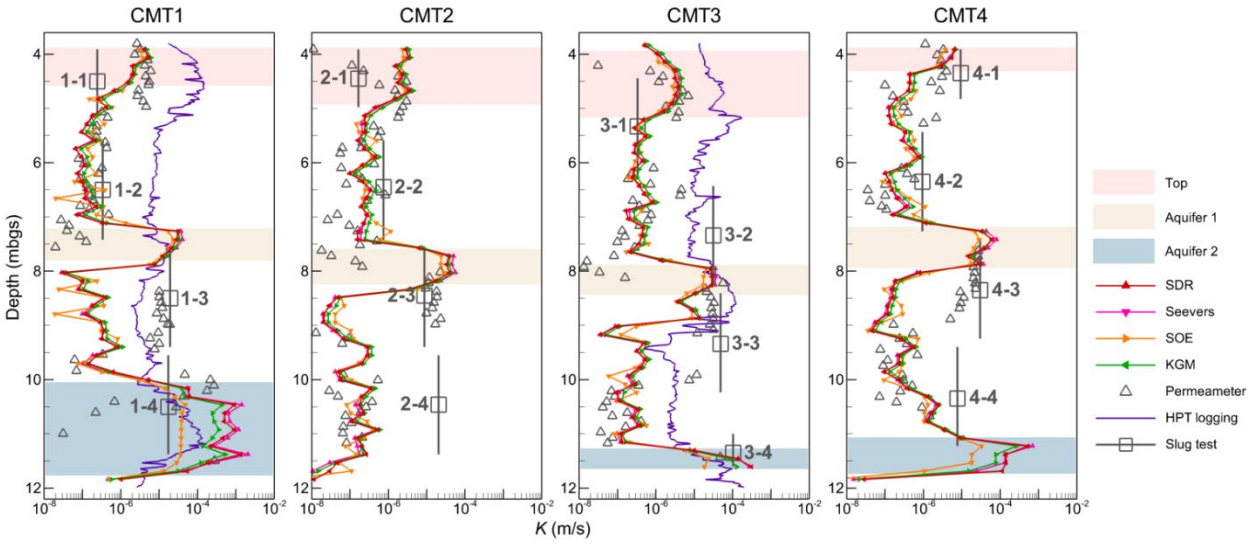


Figure 4.4: Comparison of downhole *K* estimates between NMR-derived *K* using the SDR, Seevers, KGM, SOE, and the KGM model, and hydraulic measurements from permeameter, HPT logging, and slug tests. HPT logging was only conducted in CMT1 and CMT3. The vertical bars for slug tests denote their correspondence to NMR-derived *K*. Red, yellow, and blue shadings stand for the conductive layers: Top, Aquifer 1, and Aquifer 2.

Visual analysis of the plotted NMR-derived *K* profiles indicated that four models resemble each other in general trends. Particularly, the Seevers curves almost overlapped with the SDR curves. Furthermore, relatively smooth vertical *K* variations were obtained by the SOE model for Aquifer 2. The SOE model yielded approximately one order of magnitude lower *K* estimates for Aquifer 2 and consistent *K* estimates for the other layers, compared to the calibrated SDR model. This is consistent with the downhole profiles of NMR signals shown in Figure 4.2, where the SOE values were generally larger than the T_{2ML} values with a stable gap for the downhole profiles, while the detected SOE signals were smaller than the T_{2ML} signals for Aquifer 2 in CMT1, CMT3, and CMT4. The KGM model also yielded *K* profiles with similar trends and magnitude to the SDR curves, while its *K* estimates for Aquifer 2 were by a factor of 2-3 lower than the *K* estimates by the SDR model.

4.7.2 Comparison with Hydraulic Testing Methods

Based on a visual analysis, the NMR-derived *K* profiles were compared with downhole *K* estimates from three hydraulic testing methods, respectively. First, the NMR-derived *K* profiles matched reasonably with the permeameter-based *K*. The trends and *K* magnitudes of alternating aquifers and aquitards characterized by the two methods were generally consistent across different profiles. However, discrepancies occurred in the depth correlation at ~9 mbgs, which was also found in the comparison between NMR signals and lithology in Figure 4.2. We noticed that the *K* measurements from permeameter tests were not continuous along CMT wells. For instance, records were lost from 7.5–8.5 mbgs in CMT1 and from 7.0–7.7 mbgs in CMT4. Nevertheless, the *K* magnitudes from NMR logging had good correspondence with permeameter measurements for each conductive layer (i.e., the Top layer, Aquifers 1 and 2) and the aquitard layers in between them, although depth misalignment existed. Borehole NMR data were obtained using in-situ measurements, which should lead to more accurate mapping of depth ranges for aquifer layers. On the other hand, soil compression and loss may occur during the core sampling process, which was previously reported by Alexander et al. (2011) for the NCRS data. Additionally, core samples for permeameter testing were not in situ tested in the subsurface. These factors may result in inaccurate depth correlation of aquifer layers. Such depth misalignment for permeameter measurements was reasonable, particularly for complex layered porous media, such as the NCRS. Moreover, it has been considered in previous studies (e.g., Berg and Illman 2011b; Zhao and Illman 2018) at the NCRS that low-*K* materials separating the main “aquifer zone” from 8–13 mbgs were discontinuous, providing hydraulic windows, e.g., in CMT1. Based on the NMR-derived *K* profiles (see Figure 4.2), we found that the main “aquifer zone” was at least fully separated into two aquifers in the CMT locations within the NCRS.

Slug-based *K* was typically not consistent with the NMR-derived *K* at the same depth. For instance, the slug-based *K* in CMT1-3 was more than one order of magnitude larger than the NMR-

derived *K* at the same depth. The mismatch is reasonable as flow in slug tests is towards both horizontal and vertical directions, sampling a larger volume with greater amounts of connected high-*K* materials. Here, the location of CMT1-3 was right below Aquifer 1, whose geometric mean of *K* from NMR logging was in the same order of magnitude as the slug-based *K* in CMT1-3. Likewise, most slug-based *K* values (except for CMT2-4) were close to the NMR *K* values of the conductive layers within the potential sampling distance of the slug tests (e.g., a depth range of 2 m). The large discrepancy between the slug-based and NMR-derived *K* values in CMT2-4 can likely be attributed to the fact that slug testing sampled a larger scale with higher hydraulic connectivity. Still, this scale of hydraulic connection was not captured by permeameter tests or NMR logs (e.g., horizontal laminations in sedimentary structures). However, the aquifer characteristics for such a complex layered porous medium cannot be effectively reproduced at a high resolution with slug test results.

Both NMR and HPT logging can recognize a three-layer aquifer system (i.e., the Top layer and Aquifers 1 and 2) at the NCRS, and their depth correlation was very consistent. Given that NMR and HPT logging provided consistent locations of aquifer layers, the characterization of aquifer layers with their depth correlation between various boreholes based on NMR logging was more reliable than the result based on permeameter measurements. Discrepancies existed in the estimation of *K* for different layers. For example, HPT logging overestimated *K* values of the Top layer but underestimated *K* values of Aquifer 1 in CMT1, leading to higher *K* estimates for the Top layer than Aquifer 1. Also, the HPT-characterized aquitards that intersected Aquifers 1 and 2 had an average *K* on the order of magnitude of 10^{-6} m/s, while the corresponding *K* estimates from permeameter testing and NMR logging were on the order of magnitude of 10^{-8} or 10^{-7} m/s. Confined by the upper and lower boundaries of measured *K*, HPT logging cannot adequately reproduce variable *K* ranging from gravel to clay for this glaciofluvial aquifer system. In comparison, NMR estimates were more representative of low *K* materials.

4.7.3 Evaluation of Petrophysical Models

To evaluate the *K* predictive performance of NMR logging using various petrophysical models, scatterplots of NMR-derived *K* using four petrophysical models versus co-located hydraulically measured *K* based on permeameter and slug tests were made (see Figure 4.5). Due to the narrow range of *K* measurements from HPT logging, the quantitative comparison with HPT-based *K* was not considered.

Scatter points within the scope of two dashed lines were regarded to exhibit satisfactory model fits, in which the deviation in *K* was no more than one order of magnitude (m/s). Most scatter points for comparison with permeameter tests were located near the 1:1 line, covering the range from 10^{-8} to 10^{-3} m/s. Due to the depth misalignment of permeameter measurements, some outliers with large misfits existed on either the upper-left or lower-right side of the plot, which were beyond the scope of satisfactory model fits. However, comparable numbers of large misfits occurred on both sides of the 1:1 line, so their impact on the calibrated model constants was reduced. For slug tests, the KGM model slightly overestimated the hydraulically measured *K*, while the other three models (i.e., SDR, Seevers, and SOE) conversely underestimated the co-located *K* values from slug tests, with more scatter points located on the lower-right side of the 1:1 line.

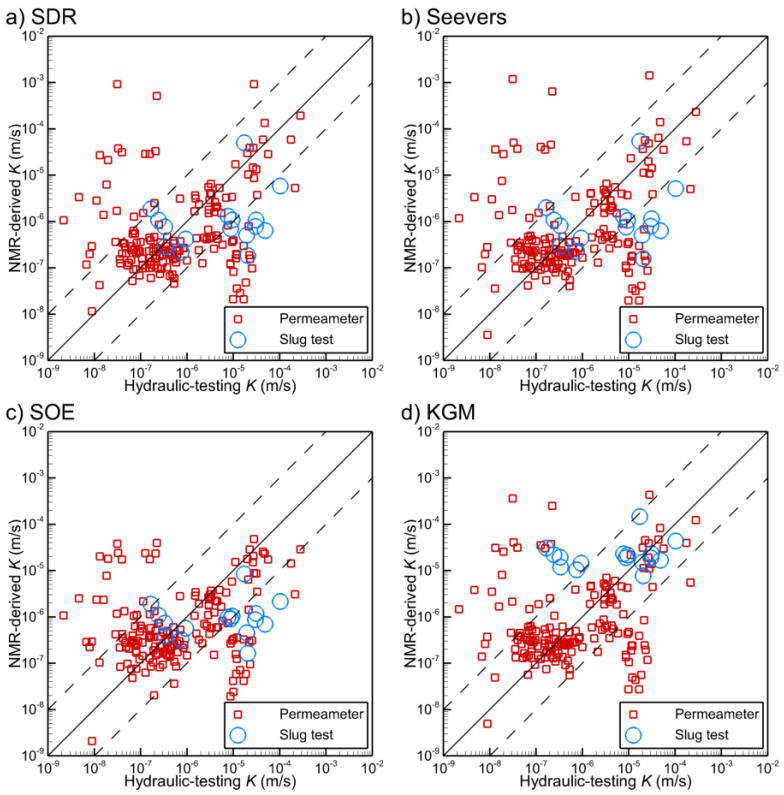


Figure 4.5: Scatterplots of measured *K* from hydraulic testing methods versus estimated *K* using four NMR models. a) to d) represent *K* estimates using the SDR, Seevers, SOE, and KGM model. The solid line is a 1:1 line indicating the perfect match, while the two dashed lines indicate the scope in which the model fit is within one order of magnitude.

Two metrics, mean absolute error (MAE) and root-mean-squared error (RMSE), were computed to quantify the NMR model fits with hydraulic measurements. Figure 4.6 summarizes the color-coded MAE and RMSE in the statistical analysis. We attached one more scenario (i.e., SDR*) where porosity was included in the SDR equation with $m = 1$ (see Equation (4.5)).

As the calculated MAE and RMSE were not reduced for both comparisons with permeameter and slug tests (see Figure 4.6), incorporating porosity in the SDR model did not improve the prediction of *K*. This result was consistent with the visual observations of total porosity profiles for the CMT wells (see Figure 4.2), in which relatively smooth vertical variations of total porosity were found, unable to indicate the locations of aquifer layers. Likewise, the stress on bulk relaxation did not improve the estimation of *K*, as the model fits of Seevers always lagged behind the SDR results. These findings corroborated previous

studies (e.g., Maurer and Knight, 2016; Kendrick et al., 2023) for the evaluations of the SDR and Seevers models at other sites.

Furthermore, the calibrated SDR constants in this study resembled the calibrated values for other glacial deposits (e.g., Crow et al., 2022; Kendrick et al., 2023). Table 4.3 lists calibrated C_{SDR} for glacial deposits given $n = 2$. The soil descriptions and the hydraulic testing methods used for K calibration datasets are also provided. Here, our calibrated values resembled the value by Crow et al. (2022) given $m = 0$ (i.e., porosity is not included) and the value by Kendrick et al. (2023) given $m = 1$. The differences in the calibrated C_{SDR} were less than 20%. In a previous study, Knight et al. (2016) calibrated C_{SDR} for alluvial deposits at three different sites and found stable C_{SDR} ranging from 0.08–0.09 m/s^3. Although the calibrated constants can still vary between different geological environments (glacial vs. alluvial), our findings confirmed that calibrated SDR constants were reasonably consistent for a given unconsolidated sediment, such as the glacial deposits at different sites.

Table 4.3: Compilation of the calibrated SDR constants for glacial deposits given $n = 2$

Location (Source)	SDR Model	Calibrated C_{SDR} (m/s^3)	Well-scale range (m/s^3)	K calibration datasets	Geologic materials
Ontario, Canada (this study)	m = 0	3.86e-3	2.70e-3–7.21e-3	Falling-head Permeameter tests	Gravel to clay
	m = 1	0.0130	0.0093–0.0244		
Ontario, Canada (Crow et al. 2022)	m = 0	3.2e-3	1.4e-3–4.6e-3	Multi-level slug tests	Gravel to silt
Wisconsin, U.S.A. (Kendrick et al. 2023)	m = 1	0.0155	0.0132–0.0213	Direct-push permeameter tests	Sand to silt

Compared to the SDR model, the SOE model result fitted slightly better with permeameter-based K, while their predictions for slug-based K were comparable. The relatively improved performance of SOE may be associated with the reduced noise of the SOE signal compared with T_{2ML}, which was previously reported by Walsh et al. (2013).

In terms of the summarized MAE and RMSE in Figure 4.6, KGM yielded the top two model fits with permeameter-based K and the optimal model fits with slug-based K. However, the locally optimized

ρ varied strongly across sites with similar sedimentations: the calibrated ρ in this study was 91 μm/s given $\tau = 1.5$, while the calibrated ρ was 172 μm/s given $\tau = 1.143$ by Kendrick et al. (2023). Therefore, whether the KGM constant remains consistent across similar geological environments still requires further investigation.

NMR models	Permeameter		Slug test	
	MAE	RMSE	MAE	RMSE
SDR	1.95	2.79	2.32	2.67
SDR*	1.96	2.81	2.33	2.69
Seevers	1.97	2.83	2.34	2.69
SOE	1.93	2.66	2.31	2.72
KGM	1.90	2.73	1.97	2.52

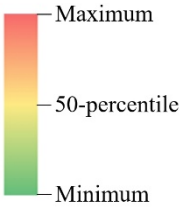


Figure 4.6: Summary of the color-coded MAE and RMSE for the comparison between NMR-derived K and hydraulic measurements from permeameter and slug tests. Color is coded column-wise, in which maximum and minimum metric values are highlighted in red and green, respectively. The change in color is linear to the change in metric values.

5 Characterizing Spatial Heterogeneity of Hydraulic Conductivity using Borehole NMR in a Complex Groundwater Flow System

Citation:

Wang, C., Steelman, C. M., & Illman, W. A. (2025b). Characterizing Spatial Heterogeneity of Hydraulic Conductivity using Borehole NMR in a Complex Groundwater Flow System. *Water Resources Research*, *61*, e2024WR039717. https://doi.org/10.1029/2024WR039717

5.1 Overview

Borehole nuclear magnetic resonance (NMR) logging can yield estimates of hydraulic conductivity (K) in unconsolidated sediments. Previous studies focused on establishing petrophysical models relating NMR responses to K and calibrating model constants for optimized K estimation. However, research has yet to explore the potential of NMR logging to derive spatial K distributions, which would enable its utilization in numerical groundwater flow and transport models. In this study, we construct various spatial K models based on NMR logging data. Characterization of spatial heterogeneity between NMR logs is explored using: 1) geostatistical interpolation approaches, including ordinary kriging and indicator kriging, 2) a zonation approach using clustering with spatial constraints for improved extraction of zone geometry, and 3) a hybrid model of multi-level spatial heterogeneity nesting a zonal representation with zonally kriged K. The representativeness of NMR-derived spatial K models is evaluated by reproducing a permeameter-based K profile at an unsampled location and by comparing the numerically simulated drawdown responses with field observations of ten pumping tests. Results reveal that the spatially associated zonation model can effectively represent the K patterns between boreholes.

Incorporating intralayer heterogeneity further refines the characterization of *K* heterogeneity, achieving optimal drawdown predictions. More importantly, its drawdown prediction performance remains stable with a limited NMR dataset. This study provides a framework for using high-resolution NMR-derived *K* profiles from multiple boreholes to characterize spatial heterogeneity at sub-meter scales in a highly heterogeneous, layered geologic deposit.

5.2 Introduction

Accurately estimating hydrogeological properties (e.g., hydraulic conductivity (*K*)) is essential to investigations of groundwater flow and solute transport in the porous medium. Borehole geophysical logging is extensively used for hydrogeophysical characterization within the engineering and environmental sectors due to its fast application and minimum disturbance to the subsurface. However, conventional geophysical logging methods typically provide indirect estimates of hydrogeological properties.

Nuclear magnetic resonance (NMR) logging can yield *K* estimates after interpretation of NMR signals (see review by Behroozmand et al., 2015). Though initially used in the petroleum industry, NMR logging has been increasingly applied in shallow unconsolidated sediments with the development of slim-hole NMR logging tools (Perlo et al., 2013; Sucre et al., 2011; Walsh et al., 2013; Zhu et al., 2021). Various petrophysical relationships have been developed to estimate *K* from NMR responses (Coates & Dumanoir, 1973; Dlugosch et al., 2013; Kenyon et al., 1988; Seevers, 1966; Timur, 1968; Walsh et al., 2013). Also, evaluation of these petrophysical relationships (e.g., Crow et al., 2022; Dlubac et al., 2013; Kendrick et al., 2021, 2023; Knight et al., 2016; Maurer & Knight, 2016) reveals that the NMR-derived *K* estimates are very consistent with hydraulic-testing *K* measurements. The optimal model fit in *K* is typically within one order of magnitude, making NMR logging a reliable tool for high-resolution site characterization along boreholes.

Despite the abundant studies aimed at improving *K* estimation using NMR logs, limited efforts have been made to determine 3D *K* distributions. To derive 3D *K* distributions, patterns observed along individual boreholes are correlated using various spatial modeling techniques. Ordinary kriging estimation might be the most popular method for interpolation, which assumes that the random field of targeted properties follows a continuous Gaussian process. To capture different non-Gaussian fields, discrete facies models (de Marsily et al., 2005) are used to better describe discontinuous features, e.g., the indicator estimation and simulation (Deutsch & Journel, 1998) and the Markov chain model (Carle et al., 1998). However, geological structures may not be well reflected in these models, diminishing their predictive capacity of groundwater flow and solute transport (Berg & Illman, 2015; Salamon et al., 2007; Zhao et al., 2023).

The zonation approach (Dietrich & Tronicke, 2009; Doetsch et al., 2010; Dumont et al., 2018; Gueting et al., 2015; Liu et al., 2010; Paasche et al., 2010; Zhang et al., 2022) honors geological information and describes aquifer characteristics by partitioning hydrostratigraphic units, in which the accurate delineation of zone geometry is key to subsurface heterogeneity characterization and its predictive performance (Luo et al., 2017; Zhao & Illman, 2018). The automatic process of zonation is usually performed by clustering. Nevertheless, its performance may not exceed the zonation via visual interpretation in complex geological environments (Zhao & Illman, 2022b). Spatial correlation information (e.g., spatial coordinates) can be incorporated into clustering to provide spatial constraints such that observations are more likely to be partitioned to a cluster in its neighborhood. Studies demonstrated that the use of spatial constraints significantly removes outliers and better reproduces geological structures (Jia et al., 2021; Kawo et al., 2024; Wang & Illman, 2023). Nevertheless, zonation models usually neglect intralayer heterogeneity, potentially biasing hydraulic connections.

In order to validate the inter-well *K* heterogeneity derived using various spatial interpolation techniques (e.g., ordinary kriging, discrete facies, zonation, etc.), spatial *K* models are typically employed

in numerical simulations to predict drawdown responses to aquifer pumping and compare the model simulations with field observations. Previous work has shown that accurate predictions of drawdown responses can be challenging to achieve through spatial modeling of standard in-situ borehole hydraulic measurements (e.g., slug tests, packer tests), particularly in complex environments (Berg & Illman, 2015; Illman et al., 2010; Sun et al., 2024).

In this study, we propose an alternative approach to characterizing 3D heterogeneity based on the spatial correlation of borehole *K* estimates from NMR logs. Spatial *K* models were assessed through a numerical analysis of the drawdown responses to multiple pumping tests. Here, eight NMR logs were collected within a highly heterogeneous, layered geologic deposit consisting of glaciofluvial sediments. Based on high-resolution vertical profiles of NMR data, multiple spatial *K* models were constructed to represent various *K* patterns between boreholes. This includes ordinary kriging estimation, indicator kriging estimation, visually interpreted geological zonation, the spatially associated zonation from clustering, and a multi-level heterogeneity model using zonation and kriging interpolation to interpret interlayer and intralayer heterogeneity with their correlation. Model performance was evaluated by comparing simulated and observed drawdown responses of ten pumping tests conducted across the aquifer-aquitard system. The representativeness of NMR-derived spatial *K* models was further examined by comparing them with permeameter measurements and spatial *K* models based on permeameter tests. Novel aspects of this study include (a) the first use of NMR logging to model high-resolution 3D *K* distribution in a highly heterogeneous environment; (b) the depiction of inter-well *K* variations by modeling multi-level heterogeneity patterns; (c) the evaluation of model utility to predict hydraulic responses of multiple pumping tests through a comparative case study. Our findings suggest that borehole NMR can derive reliable 3D *K* fields, representative of the complex groundwater flow systems.

5.3 Methods

5.3.1 Borehole NMR Logging

The unequal numbers of protons and neutrons in the hydrogen atoms can result in a net nuclear spin. Therefore, water molecules possess a net nuclear magnetic angular momentum, which can be used in the geophysical applications of NMR. For borehole NMR measurements, a permanent magnet installed in the probe generates a static magnetic field B_0. Hydrogen protons precess at the Larmor frequency in this static magnetic field, with the precession frequency and the magnitude of the magnetization vector proportional to the B_0 field strength. A pulsed oscillating magnetic field B_1, tuned to a specific Larmor frequency, is given in the plane transverse to B_0 by the NMR tool to rotate the magnetization of the hydrogen protons at a desired radial distance from the NMR tool. When the B_1 field is removed, the spins return to their equilibrium state in the longitudinal direction, emitting a measurable signal. The decay of this NMR signal in the transverse plane can be described as a multiexponential curve using a characteristic time called the relaxation time, T_2:

$$S(t) = \sum_i A_{0i} e^{-t/T_{2i}} \tag{5.1}$$

where i is the pore index, indicating pores with different T_{2i}. A_{0i} is the initial amplitude for each proportion of T_{2i}, which can be used to estimate the water content when normalized by the initial amplitude in pure water.

It is often assumed that relaxation occurs in the fast diffusion regime, where the magnetization density function is homogeneous across a single pore, resulting in a single relaxation mode. Therefore, T_2 (with a unit of 1/s) can be related to the geometry of the pore in the following manner (Brownstein & Tarr, 1979):

$$T_2^{-1} = \rho \frac{S}{V} \tag{5.2}$$

where S/V is the surface area to volume ratio of the pore space (1/m), and ρ is the material-specific surface relaxivity (m/s), which is governed by the presence of paramagnetic ions on the pore surface, measuring its ability to enhance relaxation. A petrophysical relationship between K and the parameters representative of T_2 resembles the Kozeny-Carman type of equations (Seevers, 1966):

$$K = \frac{\varrho g \phi}{\mu \tau (\frac{S}{V})^2} \tag{5.3}$$

where ϱ is density, g is the gravitational acceleration, μ is dynamic viscosity, τ is tortuosity, and ϕ is porosity, which is indicated by A_0 in saturated materials. Most NMR estimators, including the Schlumberger Doll Research (SDR) (Kenyon et al., 1988), Seevers (Seevers, 1966), Sum-of-Echoes (Sezginer et al., 1999; Walsh et al., 2013), and the Timur-Coates (Coates et al., 1991; Timur, 1968) model, relate the indicators of T_2 (e.g., the mean-log T_2 and the sum-of-echo signal) with S/V to estimate K. In these models, empirically derived constants act as fitting parameters. In other attempts, the material-specific NMR parameters (e.g., ρ and τ) are utilized in the Kozeny-Godefroy model (Dlugosch et al., 2013).

The SDR model is used in this study, which adopts the mean-log T_2 as the main indicator of K:

$$K_{SDR} = C_{SDR} \phi^m T_{2ML}^n \tag{5.4}$$

where T_{2ML} is the mean-log T_2 from the T_2 distribution. C_{SDR}, m, and n are site-specific fitting parameters. The C_{SDR} constant accounts for mineralogical parameters (e.g., surface relaxivity). Previous studies revealed that little changes can occur in C_{SDR} between sites within similar sedimentations (Kendrick et al., 2023; Knight et al., 2016). m is typically assigned an integer value between 0 and 2 for unconsolidated materials. For example, $m = 0$ (Maurer & Knight, 2016), $m = 1$ (Kendrick et al., 2023; Knight et al.,

2016), or $m = 2$ (Dlubac et al., 2013) have been adopted in unconsolidated materials. $n = 2$ is commonly acknowledged to account for the pore geometry in the fast diffusion regime, as in Equation (5.3).

5.3.2 Clustering with Spatial Constraints

The *k*-means clustering algorithm is used to classify hydrostratigraphic units from NMR logs. For a standard case, each sample is assigned to its closest mean vector (i.e., the centroid), and the error sum of squares, *E*, is computed, representing the degree of dissimilarity within a cluster among all samples. The centroids are iteratively updated to find the minimum *E*.

During this process, sample locations are ignored, which produces noise that impacts the method's capacity to constrain geological structures (Wang & Illman, 2023). To reduce outliers and reflect the spatial association of the adjacent samples, spatial coordinates are incorporated into clustering as additional dimensions of the observations. The squared Euclidean distance is assigned as the distance metric with $\| \|_2$ as the Euclidean norm. The objective function *E* with spatial penalty term is:

$$E = \sum_{r=1}^{k} \sum_{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \boldsymbol{C}_r} \|\boldsymbol{x}_i - \boldsymbol{\mu}_r\|_2^2 + \beta \sum_{r=1}^{k} \sum_{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \boldsymbol{C}_r} \|\boldsymbol{y}_i - \boldsymbol{\nu}_r\|_2^2 \tag{5.5}$$

where $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ are the i^{th}-observations of the geophysical and the spatial information, respectively. $\boldsymbol{C}_r$ is set as the r^{th}-cluster ($1 \leq r \leq k$) with $\boldsymbol{\mu}_r$ and $\boldsymbol{\nu}_r$ reflecting the mean vectors of the geophysical and the spatial information. Factor β is a trade-off term for the distances in geophysical and spatial features. A default value of 1 is assigned to β, which can be increased when the medium is well-stratified, and vice versa. Equation (5.5) returns to the standard form when $\beta = 0$.

Different from the standardization of geophysical observations that yields normalized datasets with zero mean and a standard deviation of one, raw spatial coordinates are normalized as follows:

$$y_{i,l} = \frac{Y_{i,l}}{\lambda_l} \tag{5.6}$$

where $Y_{i,l}$ is the l^{th}-spatial coordinate of the i^{th}-observation, and λ_l is the average spatial dimension of hydrostratigraphic units in the l^{th}-direction, which can be inferred from site-specific geology, e.g., correlation length. The incorporation of a spatial term in clustering can significantly improve the delineation of zone geometry (Wang & Illman, 2023).

5.3.3 Multi-Level Heterogeneity Characterization

In zonation models, a representative physical property (e.g., hydraulic conductivity, *K*) is assigned uniformly to a particular zone, generating an internally homogeneous distribution of the concerned parameter for interpretation of interlayer heterogeneity; however, this limits the characterization of finer-scale intralayer heterogeneity. Another potential bias of a zonation model is that hydraulic connections between adjacent layers may not be accurately represented (e.g., sharp versus gradual changes across the zonation interfaces). Hence, a hybrid model is proposed, in which a multi-level heterogeneity pattern is modeled to characterize the spatial heterogeneity of a layered porous medium with intralayer heterogeneity.

Figure 5.1 illustrates how to determine the spatial *K* distribution from NMR logs using a hybrid model, which can be extended to other borehole testing datasets for broader use. First, a raw NMR dataset is used to estimate *K* using petrophysical relationships (e.g., the SDR model). If available, co-located *K* measurements can be used to calibrate a site-specific model fitting parameter. Second, the NMR-based *K* datasets are classified into clusters using *k*-means clustering with spatial constraints, following zone geometry delineated through interpolation and extrapolation of the zonal boundaries along boreholes from the clustered dataset. A commercial program, Leapfrog Geo, is used to delineate zone geometry with the zonal categorical data from the clustering results. A built-in fast radial basis function (Carr et al., 2001) is

adopted to produce first-pass surfaces. The geometric mean of *K* within each zone is then calculated to generate a zonation model of *K*.

To represent intralayer heterogeneity, ordinary kriging interpolation is used to generate inter-well variations within zonal geometric domains based on the clustered *K* assigned to each zone. Here, a multi-level heterogeneity model of *K* is produced, nesting the spatially associated zonation method and the geostatistical interpolation method.

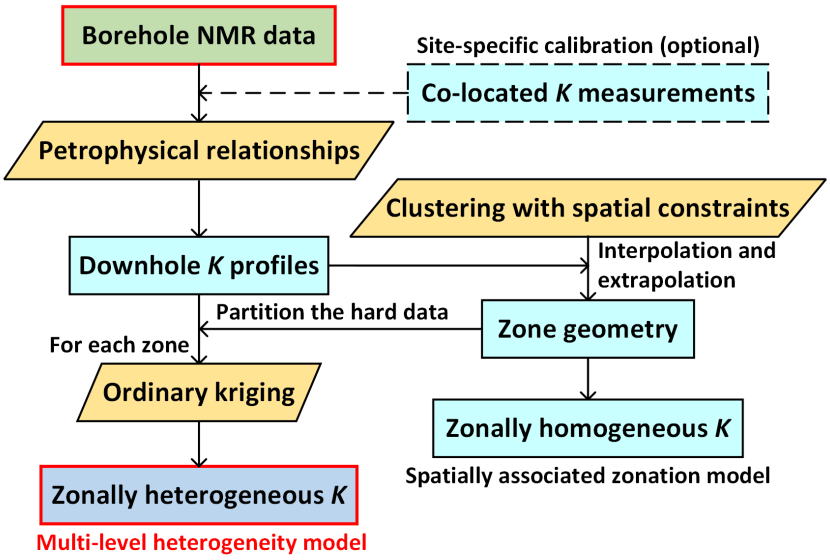


Figure 5.1: A general framework for constructing a multi-level heterogeneity model of *K* from borehole NMR logs for a layered porous medium.

5.4 Site and Data Descriptions

5.4.1 Site Geology and Instrumentation

NMR and hydraulic testing data were collected at the North Campus Research Site (NCRS) on the University of Waterloo campus in Waterloo, Ontario, Canada. The shallow subsurface is comprised of highly heterogeneous glaciofluvial deposits and tills ranging from gravel/sand, silt, and clay, forming alternating layers of aquifer and aquitard units (Alexander et al., 2011). The primary confined aquifer is

located from 8 to 13 m below ground surface (mbgs). A thick regional aquitard begins at 18 mbgs and is comprised of stiff stony silt to sandy silt. Previous studies at the NCRS have treated this layer as the base of groundwater flow models (Berg & Illman, 2011b; Luo et al., 2023; Zhao & Illman, 2018).

The NCRS consists of nine multi-level monitoring wells configured in a square pattern (Figure 5.2a). Monitoring wells include three 4-inch multi-screened pumping wells (PW1, PW3, PW5), two well clusters (PW2s, PW4s), each comprised of three 2-inch single-screened pumping wells, and four Continuous Multichannel Tubing (CMT) wells installed at the midpoint of each side (Figure 5.2b). Each CMT well contains seven monitoring intervals. The pumping wells can be utilized to pump water and monitor pressure changes, while the CMT wells are specifically designed to monitor pressure changes and collect water samples.

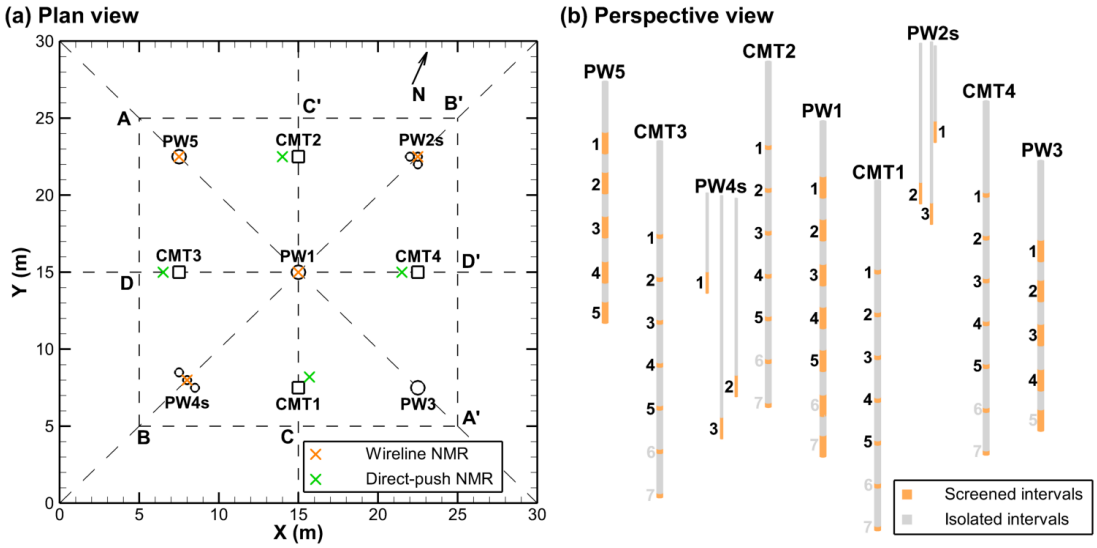


Figure 5.2: Monitoring well and NMR logging locations at the NCRS. a) Plan view of the site instrumented with nine wells, including five pumping wells (PW1, PW2s, PW3, PW4s, PW5) and four observation wells (CMT1, CMT2, CMT3, CMT4). The dashed-line square (A-B'-A'-B) delineates the area for constructing spatial *K* models. b) Perspective view of the nine wells with the screened intervals for pumping or observation. Intervals are numbered in black or light grey. Intervals numbered in black are within the depth range of NMR logs, which are used to evaluate the groundwater flow predictions based on the spatial *K* models.

5.4.2 NMR Logging

To conform to the well configurations at the NCRS, NMR logs were collected using multiple NMR tools from Vista Clara Inc. (Mukilteo, Washington, USA). Eight NMR logs were collected at the NCRS from May 24th to 27th, 2023. Four direct-push NMR logs were performed within 1 m of the CMT wells (e.g., CMT1, CMT2, CMT3, and CMT4) using a Dart Probe (DP140N) specifically designed for use with direct-push drill rigs. The exploration depth of the direct-push logging approaches approximately 12 mbgs. Four wireline NMR logs were conducted in the pumping wells. The JPY350 probe collected NMR signals in two 4" multi-screened wells (i.e., PW1 and PW5), and it was advanced to 15.9 mbgs and 11.2 mbgs in PW1 and PW5. The DP175 probe was used in the deepest wells of the two well clusters (i.e., PW2-3, PW4-3). The bottom of logging depth was 8.2 mbgs and 11.7 mbgs in PW2-3 and PW4-3. Table 5.1 summarizes the basic information on the NMR probes used in the eight NMR logs. In total, 361 NMR measurements were obtained.

To assess whether NMR logs collected using wireline or with the direct-push method coincide, a direct-push log was collected 1 m from the PW4-3 log. A qualitative comparison of these two profiles (Figure B1) proves that the two co-located NMR logs using these different probe advancing methods coincide with only slight differences.

Table 5.1: Information on the NMR probes used in the eight NMR logs.

NMR Probes	Deployment	Sensitive Diameters	Vertical Resolution	Measurement Spacing	Locations
DP140N	Direct-push	12.7 cm/15.2 cm	0.22 m	0.15 m	CMT wells
DP175	Wireline	12.7 cm/15.2 cm	0.23 m	0.15 m	PW2-3 and PW4-3
JPY350	Wireline	33.0 cm/38.1 cm	0.46 m	0.30 m	PW1 and PW5

5.4.3 Hydraulic Tests

Previously reported permeameter and aquifer pumping test results collected at the NCRS (Alexander et al., 2011; Berg & Illman, 2011b; Zhao & Illman, 2017) were examined in this study to

evaluate the representativeness of NMR-derived spatial *K* models. A total of 539 soil samples were collected in all nine wells during the drilling process for falling-head permeameter tests (Alexander et al., 2011; Berg & Illman, 2011b). Here, permeameter measurements co-located with NMR logging were used to calibrate the SDR constant for estimating *K*, while the non-co-located permeameter measurements (i.e., in PW3) were used to evaluate the *K* profiles from the spatial NMR *K* models. The soil samples were collected at a ~ 0.1 m interval along continuous core runs for the four CMT wells and PW1, while selective sampling was conducted in PW2-3, PW3, PW4-3, and PW5.

Berg and Illman (2011b) and Zhao and Illman (2017) reported the results of several aquifer pumping tests at the NCRS. Here, water was pumped/injected at screened intervals at a constant rate. At the same time, drawdown responses were monitored using pressure transducers installed in all available ports within the monitoring network (refer to Figure 5.2b). In total, 247 records of steady-state drawdown data from eight pumping tests (PW1-3, PW1-4, PW3-3, PW5-3, PW1-5, PW2-3, PW5-5, PW4-3) and two injection tests (PW1-1, PW3-1) were used in this study to assess numerically simulated drawdown responses using multiple NMR-derived spatial *K* models. Details of aquifer pumping tests are summarized in Table 5.2.

Table 5.2: Summary of pumping tests for groundwater flow predictions.

Operation Screens	Stress depth* (mbgs)	Stress type	Rates (L/min)	Number of Observations
PW1-1	3.36	Injection	1.89	22
PW3-1	3.69	Injection	0.94	18
PW1-3	7.46	Pumping	10.5	28
PW1-4	9.46	Pumping	6.3	25
PW2-3	8.00	Pumping	1.91	20
PW3-3	7.62	Pumping	2.1	26
PW5-3	7.62	Pumping	7.8	31
PW4-3	11.55	Pumping	30.2	16
PW1-5	11.46	Pumping	4.4	30
PW5-5	11.61	Pumping	8.1	31

* Stress depth indicates the mid-depths of the pumping screens.

5.5 NMR Data Analysis

5.5.1 NMR-Derived K Profiles

Borehole NMR K was estimated using the optimized SDR model. Here, we adopted $n = 2$ and compared the SDR-derived K using $m = 0$ and $m = 1$ with co-located permeameter measurements. Permeameter measurements in the four CMT wells were used to calibrate the C_{SDR} constant in Equation (5.4). Results (see Text B1) show that excluding m in SDR did not reduce the accuracy of NMR K estimates for the NCRS sediments. Therefore, we adopt $m = 0$, implying that K is linearly proportional to T_{2ML} after log transformation. A linear regression of the NMR and permeameter data yielded a $C_{SDR} = 3.86 \times 10^{-3}$ m/s^3, producing optimized K profiles at logging locations (refer to Figure 5.3). The site-specifically optimized SDR model is:

$$K_{SDR} = 3.86 \times 10^{-3} T_{2ML}^2 \tag{5.7}$$

In addition, the geologic logs and the permeameter measurements were incorporated into Figure 5.3. These geologic logs were compiled by Zhao and Illman (2017) based on the layering and soil types of 18 borehole logs in the vicinity of the NCRS. Borehole NMR-derived K estimates range from 1.16×10^{-8} to 1.36×10^{-3} m/s. As a comparison, the permeameter-based K measurements vary from 2.08×10^{-9} to 2.76×10^{-3} m/s within the corresponding depth range.

The NMR-derived K profiles in Figure 5.3 indicate two hydraulically conductive layers. An upper aquifer layer (labelled as “Aquifer 1”) is located at ~ 8 mbgs, and a lower aquifer layer (“Aquifer 2”) is situated at ~ 11 mbgs. Aquifer 2 is not found in CMT2. Less permeable sediments separate these aquifer layers.

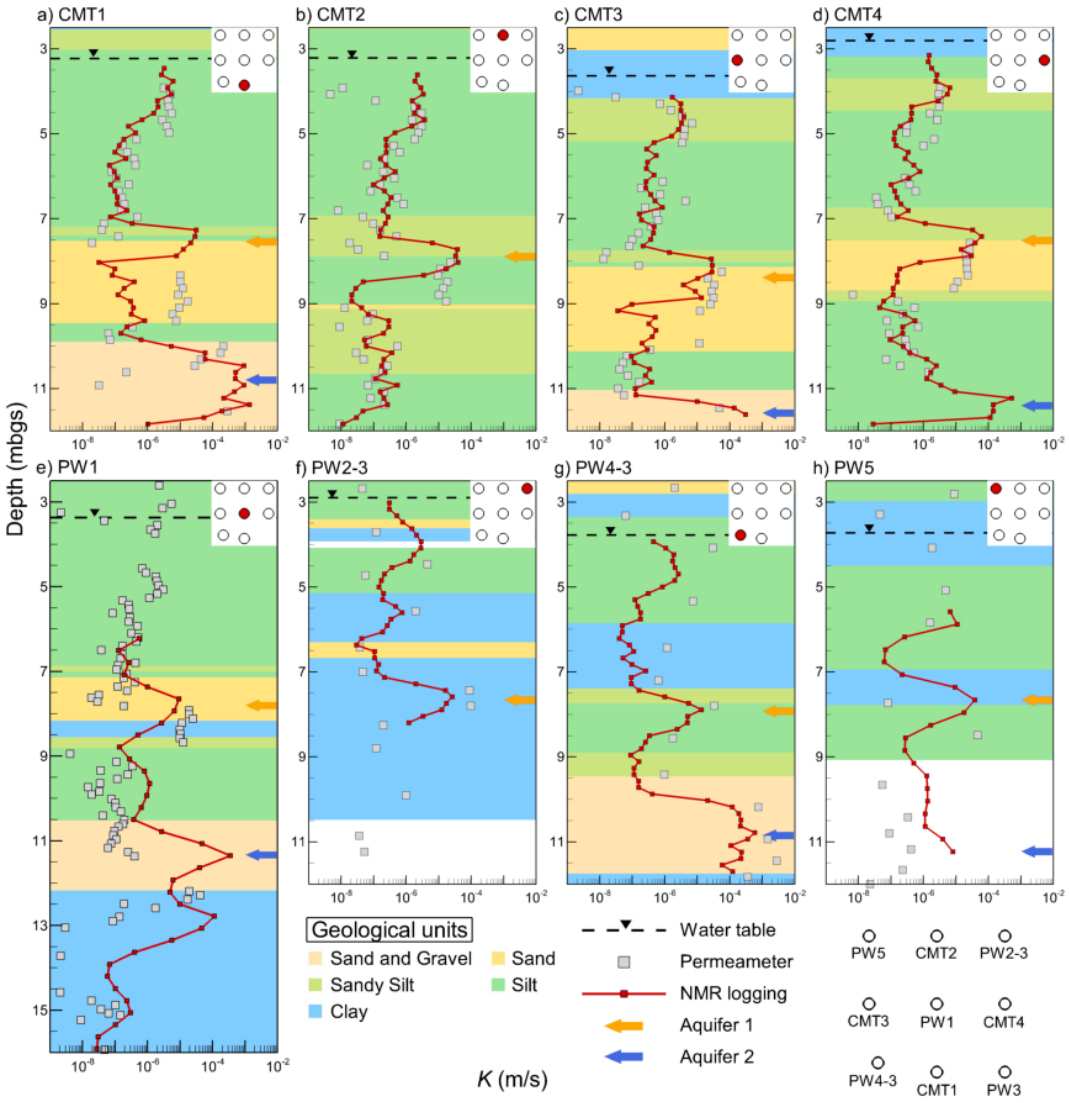


Figure 5.3: The NMR-derived *K* profiles along eight boreholes with co-located permeameter-based *K* measurements and borehole geologic logs. Small maps are added in the upper-right corner of each figure part to highlight the borehole locations. Yellow and blue arrows indicate the locations of Aquifer 1 and Aquifer 2.

5.5.2 Variogram Analysis

The *K* estimates from eight NMR logs were utilized for variogram analyses of the ordinary and indicator kriging estimation. Indicator variogram analysis was performed after coding the input data into a series of Boolean indicator datasets.

The experimental variogram for ordinary kriging is described by:

$$\gamma(h) = \frac{1}{2N(s)} \sum_{i=1}^{N(s)} [Y(x) - Y(x+h)]^2 \tag{5.8}$$

where $\gamma(h)$ is the variogram, h is the lag distance, $Y(x)$ is the ln K observation at location x, and $N(s)$ is the pair of observations in the lag distance. Here, the lag distance is assigned to be the sampling interval of NMR logs (i.e., ~ 0.15 m). The experimental variogram exhibits cyclic changes known as the "hole effect" that results from the layered structure, which was noted in previous variogram analysis of the ln K field at the NCRS (Berg & Illman, 2015) as well as other highly heterogenous porous environments (Salamon et al., 2007).

The experimental variogram is fitted with the exponential variogram model in SGeMS (Remy et al., 2007):

$$\gamma(h) = c(1 - e^{-\frac{3h}{a}}) \tag{5.9}$$

where c is the sill after removing the nugget, and a is the range. Figure B2 illustrates the vertical experimental variogram and the fitted exponential model. The fitted c and a are 7.12 and 2.48, indicating that the variance of ln K is 7.12, and the average correlation length in z is 2.48 m.

Indicator kriging relies on the calibration of a cumulative distribution function (CDF) at each element to compute the expected ln K field. A CDF can provide the probability of ln K as a cumulative histogram based on threshold values. Here, the input ln K data along boreholes was coded into a series of Boolean indicator datasets in terms of nine thresholds that evenly split the full range of the input dataset. Then, the converted indicator datasets were fitted with exponential or spherical models in SGeMS. Table B1 lists the thresholds and model fitting parameters for vertical indicator variograms. In addition, power models were used to fit the lower and upper tails of CDFs with an expected range of K from 1×10^{-9} to 1×10^{-2} m/s, consistent with NCRS glaciofluvial deposits (Alexander et al., 2011).

As insufficient data is available to fit unique variogram models in the horizontal plane, we used the structure and sill (i.e., variance of ln *K*) from the vertical models for the omnidirectional horizontal variogram and indicator variogram. The NMR-derived *K* profiles (refer to Figure 5.3) show that Aquifer 1 is horizontally continuous, and Aquifer 2 is discontinuous but only pinches out in CMT2. Here, the aquifer layers are well extended in the horizontal plane. Therefore, we assign a ratio of 8 between horizontal and vertical correlation lengths, obtaining an estimated horizontal correlation length of 19.82 m. Wang and Illman (2023) investigated the use of various horizontal correlation lengths for spatial constraints in a numerical sandbox experiment. They found that their clustering results are stable if a reliable vertical correlation length can be obtained.

5.5.3 Spatially Associated Clustering

The *k*-means clustering with spatial constraints was used to partition the ln *K* data to interpret hydrostratigraphy. The average layer dimensions were determined from the estimated correlation lengths of 19.82 m, 19.82 m, and 2.48 m in x-, y-, and z-directions, respectively. The "elbow" curve method (Kodinariya & Makwana, 2013) was used to determine an optimal number of clusters (i.e., *k*) from a series of *k*-means clustering outputs. Figure B3a shows the "elbow" curve with an inflection point corresponding to an optimal number of clusters, $k = 6$. Furthermore, the *k*-means ++ algorithm (Arthur & Vassilvitskii, 2007) was used to provide proper seedings of initial mean vectors.

Figure 5.4a shows the NMR-derived *K* along boreholes partitioned into six clusters with spatial association. In contrast, Figure 5.4b shows the clustered *K* profiles using a standard *k*-means clustering regardless of spatial constraints, in which an optimal $k = 5$ is determined using the "elbow" curve method (Figure B3b).

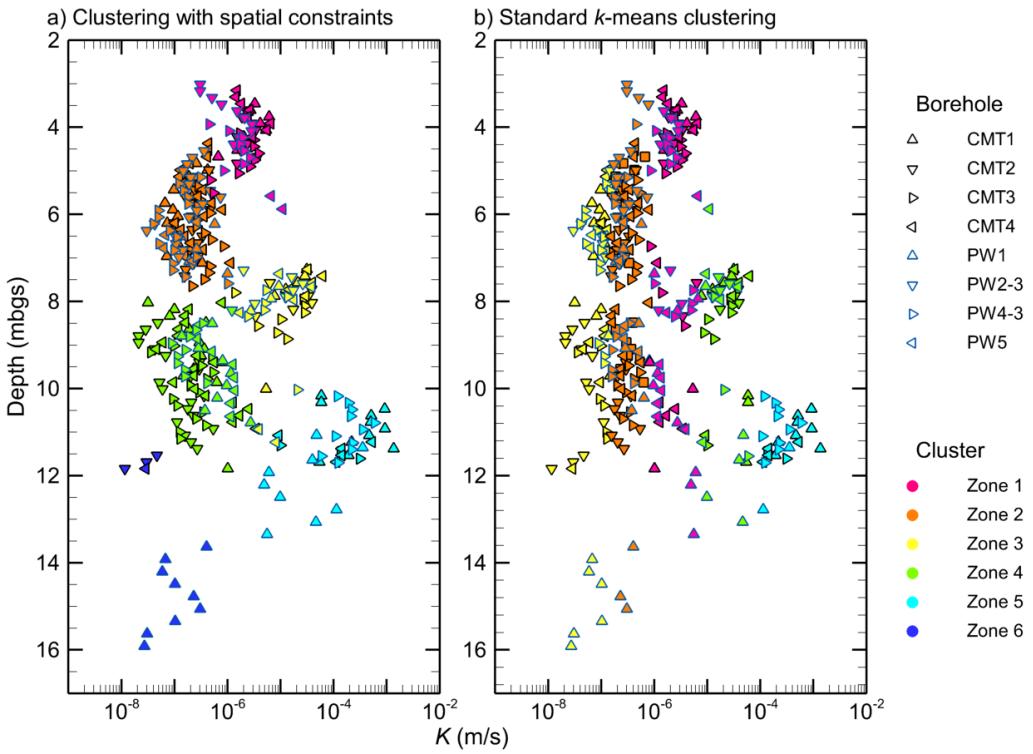


Figure 5.4: Clustered *K* profiles from eight NMR logs a) with and b) without spatial constraints. Various symbols are used to differentiate the borehole locations, which are outlined in black and blue to represent the observations from the CMT and pumping wells, respectively. Clustered *K* observations in different zones are filled in contrasting colors for visualization.

5.6 Results

5.6.1 Comparative Cases of Numerical Models

To assess the representativeness of various spatial *K* models, a numerical groundwater flow simulation was conducted using an integrated hydrologic modeling simulator, HydroGeoSphere (Aquanty Inc., 2025). To minimize the impacts of boundary conditions, the simulation domain is assigned dimensions of 70 m × 70 m × 17 m with PW1 located in the center. The simulation domain is discretized into variably-sized rectangular finite elements with a minimal dimension of 0.15 m in the z-direction (see Figure B4), corresponding to the NMR logging interval. Here, the top and bottom boundaries are impermeable, while constant heads are defined for the lateral boundaries. The spatial *K* models derived

from borehole *K* estimates are gridded within a 20 m × 20 m data domain (refer to Figure 5.2a) encompassing the borehole data. *K* values at the outer edge of our spatial *K* models are extrapolated to the boundary of our groundwater flow models (i.e., 70 m × 70 m simulation domain).

Nine comparative cases of spatial *K* models were designed to examine the influences of multiple factors on representing 3D *K* fields and predicting drawdown responses. Table 5.3 summarizes the input data and spatial interpolation methods of these spatial *K* models. The naming of these cases indicates the spatial interpolation method, the number of borehole logs, and the data source for *K* estimation. For example, *Krig_ordinary_8* represents the *K* field derived by the ordinary kriging of eight NMR *K* profiles, and *Multi-level_9 (permeameter)* stands for the *K* field constructed by the multi-level heterogeneity characterization of nine permeameter *K* profiles.

The first factor is the use of various spatial interpolation methods to obtain 3D *K* fields based on eight NMR logs. Two conventional interpolation methods were considered, including ordinary kriging estimation (Case: *Krig_ordinary_8*) and indicator kriging estimation (Case: *Krig_indicator_8*). Also, the zonation method based on clustering with spatial constraints (Case: *Zonation_clustering_8*) was included. To highlight the performance of spatially associated zonation, we adopted another zonal representation by Zhao and Illman (2017), in which the zone geometry was delineated by visual interpretation of 18 geologic logs, resulting in a 19-layer geological zonation model with zonal *K* estimates calculated using borehole NMR-derived *K* within each layer (Case: *Zonation_geology_8*). Moreover, we included the multi-level heterogeneity model, which nests the spatially associated zonation and the zonally kriged interpolation (Case: *Multi-level_8*).

To investigate the representativeness of spatial *K* models based on varying amounts of NMR *K* data, two NMR datasets were considered to determine 3D *K* fields: one dataset incorporates all available NMR logs for spatial correlation (i.e., eight NMR logs); another dataset incorporates a reduced number of NMR logs to examine the effectiveness of the multi-level heterogeneity model with a limited NMR

dataset. Here, five NMR logs were assigned for the "limited dataset", including PW1, CMT1, CMT2, CMT3, and CMT4. These five NMR logs can form two orthogonal cross-sections intersected at PW1 (refer to C-C' and D-D' in Figure 5.3). Two spatial interpolation methods, including ordinary kriging estimation (Case: *Krig_ordinary_5*) and multi-level heterogeneity characterization (Case: *Multi-level_5*), were applied to derive 3D *K* fields with the limited dataset.

To examine the drawdown prediction performance of NMR-based models, we constructed spatial *K* models using the same spatial modeling procedures based on permeameter measurements. Here, we included two spatial *K* models based on nine permeameter-based *K* profiles, using ordinary kriging estimation (Case: *Krig_ordinary_9 (permeameter)*) and multi-level heterogeneity representation (Case: *Multi-level_9 (permeameter)*). Figure B5 shows variogram fitting for *Krig_ordinary_5*, *Krig_ordinary_9 (permeameter)*, *Multi-level_5*, *Multi-level_8*, and *Multi-level_9 (permeameter)*. Figure B6 presents the clustered *K* profiles for *Multi-level_5* and *Multi-level_9 (permeameter)*.

Table 5.3: Summary of input datasets and spatial interpolation methods for comparative cases of spatial *K* models.

Case name	Number of *K* observations	Borehole testing methods	Spatial interpolation methods
Krig_ordinary_5	254	NMR logging	Ordinary kriging estimation
Krig_ordinary_8	361	NMR logging	Ordinary kriging estimation
Krig_indicator_8	361	NMR logging	Indicator kriging estimation
Krig_ordinary_9 (permeameter)	539	Permeameter	Ordinary kriging estimation
Zonation_geology_8	361	NMR logging	Visually-interpreted geological zonation
Zonation_clustering_8	361	NMR logging	Spatially associated zonation with clustering
Multi-level_5	254	NMR logging	Multi-level heterogeneity characterization
Multi-level_8	361	NMR logging	Multi-level heterogeneity characterization
Multi-level_9 (permeameter)	539	Permeameter	Multi-level heterogeneity characterization

5.6.2 Spatial Distributions of K Fields

Cross-sectional views of spatial *K* models along A-A', B-B', C-C', and D-D' are shown in Figures 5.5 and 5.6. Most NMR logs are available to a depth < 12 mbgs, and the water table is > 3 mbgs. Therefore, spatial *K* representations with a depth range from 3 – 14 mbgs are plotted, which cover the main depth of exploration in the NMR logging.

Figures 5.5a-d and 5.6a-d exhibit the cross-sectional views of spatial *K* models using kriging estimation, including *Krig_ordinary_5*, *Krig_ordinary_8*, *Krig_indicator_8*, and *Krig_ordinary_9 (permeameter)*. Vertical *K* variations with alternating aquifer and aquitard layers can be recognized in *Krig_ordinary_5* (Figures 5.5a and 5.6a). High-*K* materials > 1×10^{-5} m/s are primarily distributed within Aquifer 2, especially in regions adjacent to CMT1, showing a thickened layering with $K > 1 \times 10^{-4}$ m/s. As three additional NMR logs (i.e., PW2-3, PW4-3, and PW5) were involved in *Krig_ordinary_8* (Figures 5.5b and 5.6b), more details of the heterogeneous *K* fields are captured near these well locations.

The 3D *K* field based on indicator kriging (*Krig_indicator*_8, see Figures 5.5c and 5.6c) is consistent with *Krig_ordinary_8*. Nevertheless, the indicator kriging estimation derived slightly more conservative *K* estimates. For instance, the derived *K* field in *Krig_ordinary_8* ranges from 1.86×10^{-8} m/s to 6.66×10^{-4} m/s, while *Krig_indicator_8* yielded *K* estimates from 2.85×10^{-8} m/s to 5.34×10^{-4} m/s. In addition, continuous *K* variations in *Krig_ordinary_8* (e.g., below the bottom of PW2-3 in Figure 5.5b) become truncated in the corresponding region in *Krig_indicator_8* (refer to Figure 5.5c).

Krig_ordinary_9 (permeameter) recognized similar alternating structures of aquifers and aquitards; however, the *K* estimates near well locations show a more contrasting pattern. For example, high-*K* regions in PW4-3 and PW1 exhibit correlated patterns in *Krig_ordinary_8* (see Figure 5.5b). In contrast, the high-*K* region in the B-B' cross-section only exists near PW4-3 (see Figure 5.5d) in *Krig_ordinary_9 (permeameter)*.

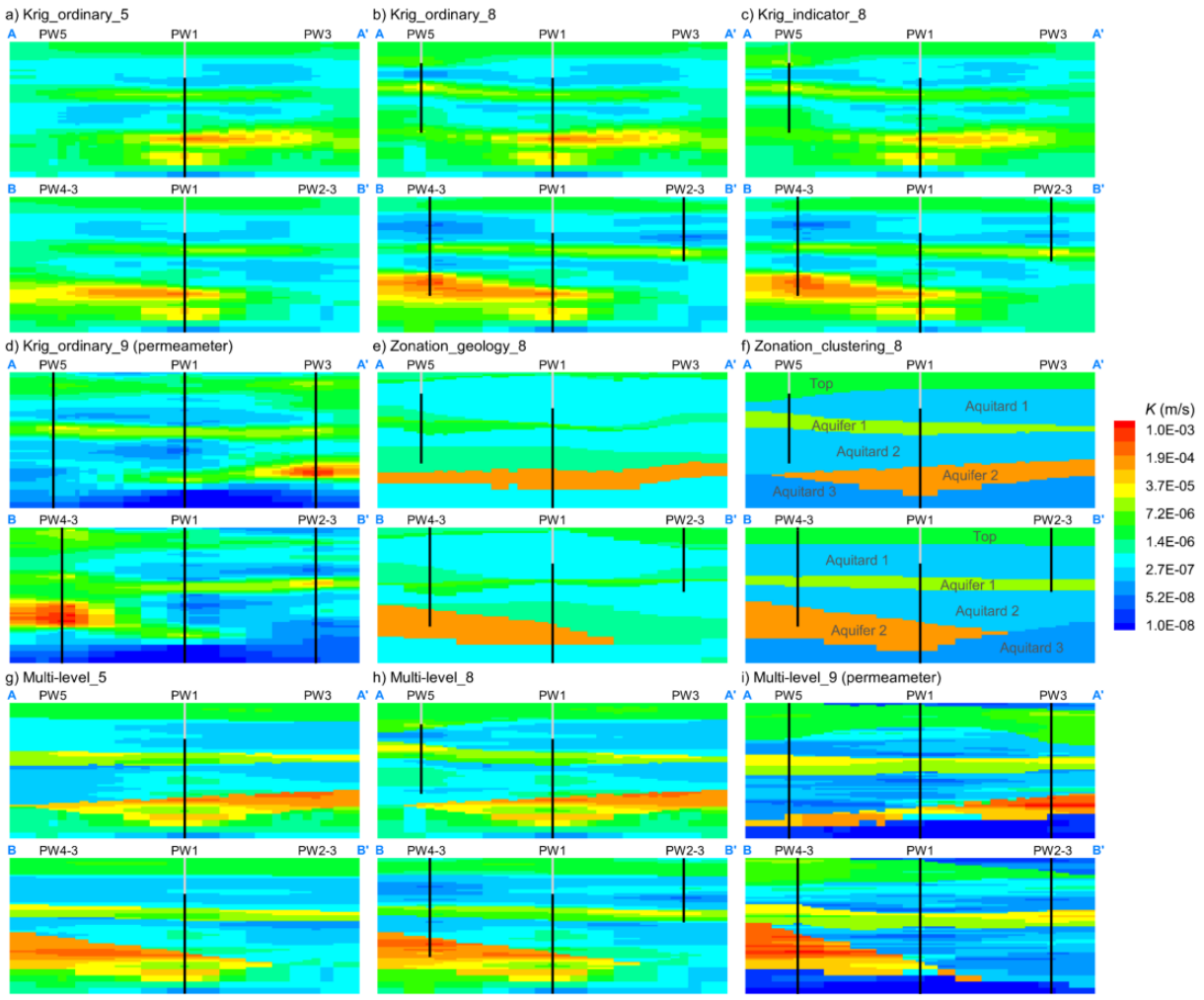


Figure 5.5: Contour plots of *K* distribution in A-A' and B-B' sections based on comparative cases of spatial modeling, including a) *Krig_ordinary_5*: the ordinary kriging using five NMR logs, b) *Krig_ordinary_8*: the ordinary kriging using eight NMR logs, c) *Krig_indicator_8*: the indicator kriging using eight NMR logs, d) *Krig_ordinary_9 (permeameter)*: the ordinary kriging using nine boreholes of permeameter measurements, e) *Zonation_geology_8*: the visually-interpreted geological zonation using eight NMR logs, f) *Zonation_clustering_8*: the spatially associated zonation with clustering using eight NMR logs, g) *Multi-level_5*: the multi-level heterogeneity characterization using five NMR logs, h) *Multi-level_8*: the multi-level heterogeneity characterization using eight NMR logs, and i) *Multi-level_9 (permeameter)*: the multi-level heterogeneity characterization using nine boreholes of permeameter measurements. Six aquifer and aquitard layers obtained by clustering NMR logs are labelled in part f). Logging depths of borehole NMR and core sampling for permeameter tests are indicated in these cross-sections using black lines.

Figures 5.5e-f and 5.6e-f show the four cross-sections of the spatial *K* models for

Zonation_geology_8 and *Zonation_clustering_8*. Six aquifer and aquitard layers were classified by the

spatially associated zonation based on NMR logs (refer to Figures 5.5f and 5.6f), where Aquifer 1 and Aquifer 2 are intersected with three aquitard layers (namely Aquitard 1, Aquitard 2, and Aquitard 3 from the top). Another conductive layer is located near the top boundary. Here, *Zonation_clustering_8* depicts a thin, yet laterally continuous Aquifer 1 extending to the lateral boundaries of the data domain with no preferential dipping direction. Conversely, Aquifer 2 has a larger volume but exhibits a more variable layer thickness. It dips toward the northwest and pinches out in PW5, CMT2, and PW2-3. Aquifer 2 is more conductive than Aquifer 1 by approximately one order of magnitude in *K* (m/s).

Zonation_geology_8 delineates two aquifer layers roughly corresponding to Aquifers 1 and 2 in their geometry and zonal *K* estimates. Slight differences exist in the lateral extension of aquifer layers. For example, Aquifer 2, classified by *Zonation_clustering_8*, extends to CMT4 but pinches out in PW5, while the corresponding aquifer layer delineated by *Zonation_geology_8* has continuous lateral extension in PW5 yet pinches out in CMT4. Compared with *Zonation_geology_8*, lower *K* values were assigned to the three aquitard layers by *Zonation_clustering_8*, especially for Aquitard 3, as more low *K* estimates were classified to these layers with spatially associated zonation.

Figures 5.5g-i and 5.6g-i exhibit the four cross-sections of the 3D *K* fields using multi-level heterogeneity models for *Multi-level_5*, *Multi-level_8*, and *Multi-level_9 (permeameter)*. The *K* distributions of Aquitard 3 in *Multi-level_*5 (see Figures 5.5g and 5.6g) and *Multi-level_8* (Figures 5.5h and 5.6h) were based on the entire dataset rather than the *K* data within this zone due to insufficient *K* estimates for interpolation. Thus, their *K* fields of Aquitard 3 are identical to the corresponding regions in *Krig_ordinary_5* and *Krig_ordinary_8*. In terms of *K* patterns in PW5, *Multi-level_*8 delineated a shallower Aquifer 1 and a less lateral extension of Aquifer 2 than *Multi-level_5*. Still, their general *K* trends are very consistent. Furthermore, the multi-heterogeneity model successfully mapped the heterogeneous *K* distribution internal to aquifer layers. For example, Figure 5.5h exhibits a horizontal *K* variation within Aquifer 2 spanning over two orders of magnitude, where high-*K* materials were found in

CMT1, then decreased around PW1, and faded in CMT2. Moreover, both sharp and gradual changes in layer boundaries were captured. A sharp transition from 1.0×10^{-7} m/s to 1.0×10^{-3} m/s was characterized near the boundary of Aquitard 2 and Aquifer 2 in CMT1 (Figure 5.6h), while a more gradual transition occurred between Aquifer 1 and Aquitard 2 in PW5 (Figure 5.5h).

Similar to the NMR cases, two aquifer layers were identified in *Multi-level_9 (permeameter)* (see Figures 5.5i and 5.6i), with their general aquifer characteristics corresponding to Aquifers 1 and 2 in *Multi-level_8*. A significant difference is that two aquifer layers are connected near PW4-3 and CMT1 in *Multi-level_9 (permeameter)*. In addition, *Multi-level_9 (permeameter)* yielded lower *K* estimates for Aquitard 3. However, NMR logging only reached Aquitard 3 at the top, lacking sufficient *K* observations to clearly characterize this layer.

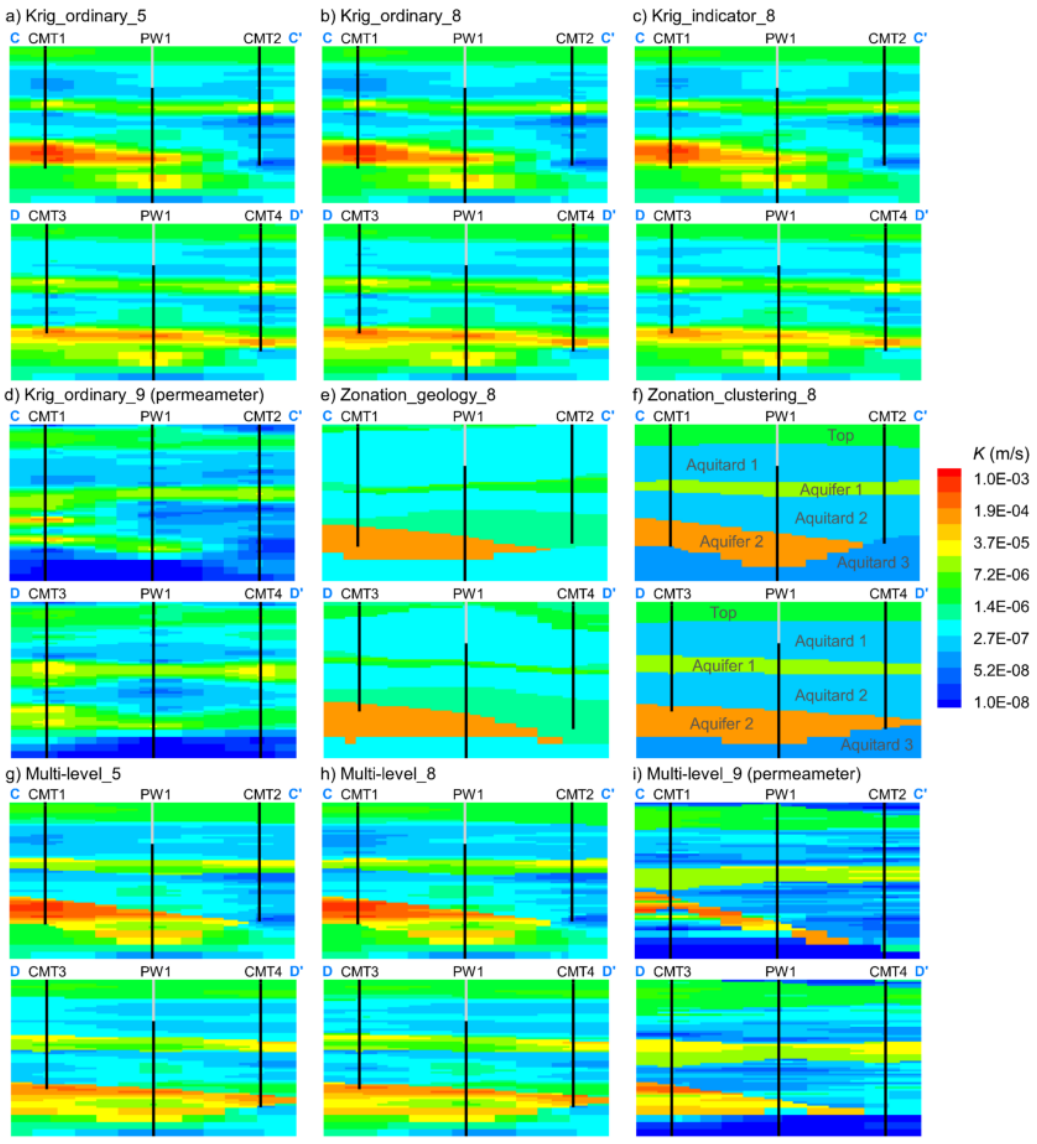


Figure 5.6: Contour plots of *K* distribution in C-C' and D-D' sections. Figure parts a) – i) represent the nine comparative cases of spatial modeling in the same sequence as in Figure 5.5. Six aquifer and aquitard layers obtained by clustering NMR logs are labelled in part f). Logging depths of borehole NMR and core sampling for permeameter tests are indicated in these cross-sections using black lines.

5.6.3 Evaluation of Spatial K Fields

To assess the model's capability of representing spatial heterogeneity, the model-derived *K* based on NMR logs was compared with the permeameter measurements in PW3. PW3 was unsampled in NMR logging, and no NMR logs were conducted within a radial distance of 7.5 m to PW3. To examine various

spatial interpolation methods, we selected five spatial *K* models, including *Krig_ordinary_8*, *Krig_indicator_8*, *Zonation_geology_8*, *Zonation_clustering_8*, and *Multi-level_8*.

Figure 5.7 presents the NMR-derived *K* using various spatial interpolation methods and co-located permeameter *K* measurements. Furthermore, the model-derived *K* estimates and permeameter *K* measurements located within 10 cm were paired to quantitatively calculate their differences. Mean absolute error (MAE) and root-mean-square error (RMSE) were employed to indicate model performance.

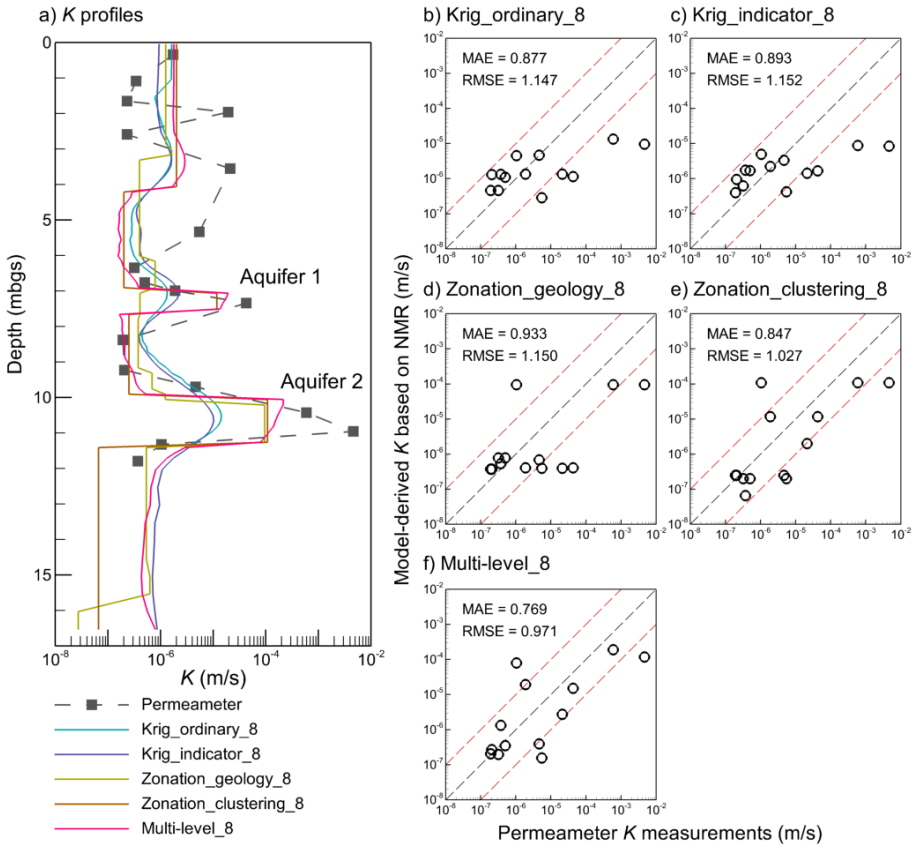


Figure 5.7: Comparison of *K* in PW3 between the model-derived *K* based on NMR logging and the permeameter *K* measurements. a) is the *K* profiles along PW3, in which the locations of Aquifer 1 and Aquifer 2 are labelled. b)-f) are the scatterplots of model-derived *K* versus permeameter *K* measurements using five spatial interpolation methods. The black dashed lines are the 1:1 lines indicating the perfect match, while the red dashed lines indicate the range in which the model fit is within one order of magnitude.

The *K* profile from permeameter measurements indicates the depth ranges of two aquifer layers, which correspond to Aquifer 1 and Aquifer 2. Qualitatively, the *K* profiles of *Krig_ordinary_8* and *Krig_indicator_8* were smoother and underestimated the *K* of the two aquifer layers by over one order of magnitude. The *K* profile from *Zonation_geology_8* matches the permeameter-based *K* in Aquifer 2. However, it does not recognize a high-*K* layer in the corresponding region of Aquifer 1. The *K* profiles from *Zonation_clustering_8* and *Multi-level_8* provide the most consistent *K* values with the permeameter *K* measurements for Aquifer 1, Aquifer 2, and the low-*K* layer (i.e., Aquitard 2) between them.

In terms of the quantitative comparison of the scatterplots (see Figure 5.7b-f), *Krig_ordinary_8* and *Krig_indicator_8* exhibit similar scatter distributions, in which five model-derived *K* estimates are more than one order of magnitude smaller than the permeameter measurements. *Zonation_geology_8* also derived a cluster of *K* that underestimated the permeameter measurements. *Zonation_clustering_8* and *Multi-level_8* obtained better model fits with permeameter measurements than kriging methods and the visually-interpreted zonation method, and their scatter points were distributed evenly on the two sides of the 1:1 line. *Multi-level_8* led the *K* prediction performance among the five spatial interpolation methods.

5.6.4 Simulation of Drawdown Responses

Figure 5.8 shows scatterplots of simulated versus observed drawdown responses in ten pumping/injection tests for nine comparative cases of groundwater flow models. Abundant scatter points were located on the upper-left side of the 1:1 line in the cases using kriging methods (i.e., *Krig_ordinary_5*, *Krig_ordinary_8*, *Krig_indicator_8*, and *Krig_ordinary_9 (permeameter)*). Drawdown fits were clearly improved when the zonation and the multi-level heterogeneity characterization were used for determining 3D *K* fields. For example, the drawdown scatter points in *Zonation_geology_8* and *Zonation_clustering_8* (see Figure 5.8e-f) are distributed evenly about the 1:1 line, with their misfits in a reasonable range. When the number of NMR logs was reduced, the drawdown predictions based on the

kriged *K* field (*Krig_ordinary_5* in Figure 5.8a) significantly deteriorated. In contrast, the drawdown misfits remained at the same degree for *Multi-level_5* (see Figure 5.8g). Lastly, we noticed that drawdown prediction performance also varied for the simulation of hydraulic stress at different locations. For example, the simulated drawdown in the PW1-3 test did not fit satisfactorily with the observations for all cases using NMR logs, though the drawdown fits were slightly improved in *Multi-level_5* and *Multi-level_8*.

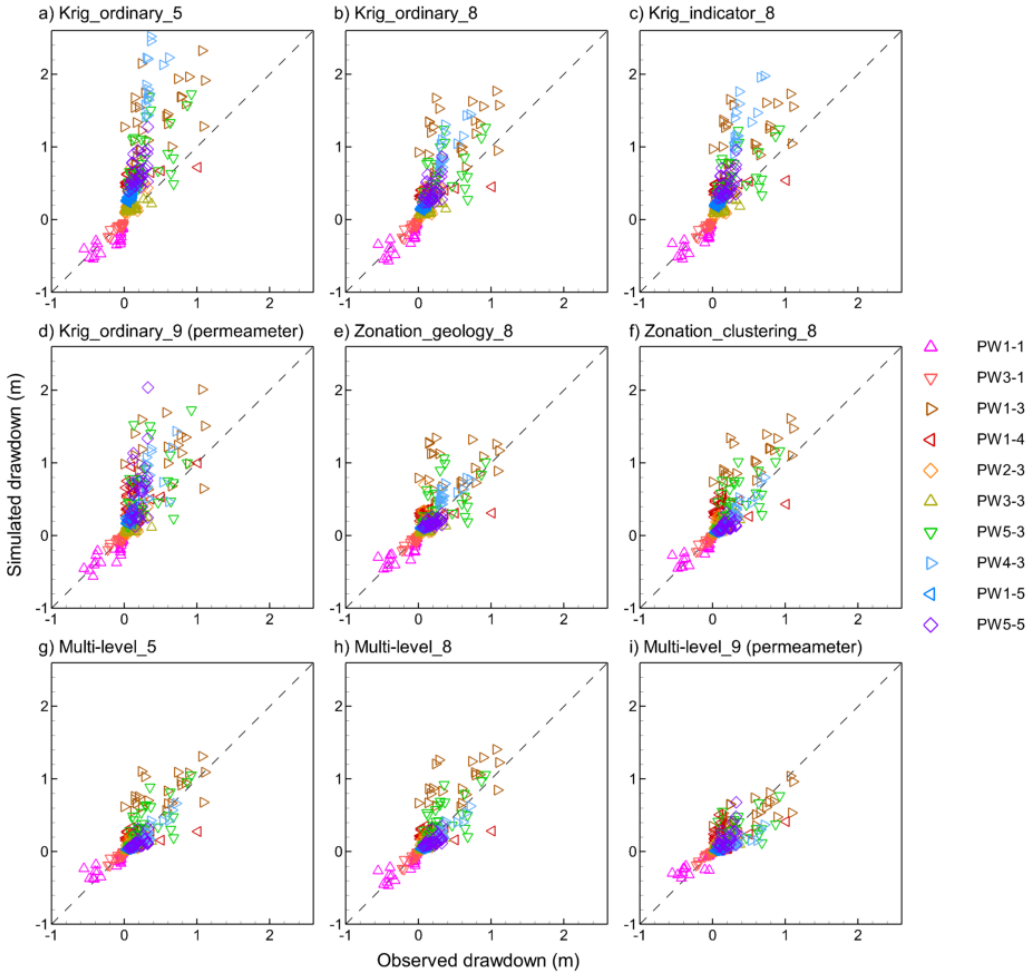


Figure 5.8: Comparison of simulated and observed drawdown. (a-j) indicate the drawdown comparisons for aquifer hydraulic tests at PW1-1, PW3-1, PW1-3, PW2-3, PW3-3, PW5-3, PW1-4, PW1-5, PW4-3, and PW5-5, respectively. Note: Negative drawdown signals in PW1-1 and PW3-1 represent injection tests.

To fairly compare the degrees of drawdown misfits among different pumping tests, we employed MAE and RMSE as performance indicators and normalized them by the average observed drawdown of each pumping test:

$$(NMAE)_i = \frac{\sum_{j=1}^{n_i} |s_{ij} - s_{ij}^*|}{\left|\sum_{j=1}^{n_i} s_{ij}^*\right|} \tag{5.10}$$

$$(NRMSE)_i = \frac{\sqrt{\frac{1}{n_i}\sum_{j=1}^{n_i}(s_{ij} - s_{ij}^*)^2}}{\frac{1}{n_i}\left|\sum_{j=1}^{n_i} s_{ij}^*\right|} \tag{5.11}$$

where $(NMAE)_i$ and $(NRMSE)_i$ are the normalized MAE and RMSE for the *i*th pumping test. n_i is the number of drawdown data in the *i*th pumping test. s_{ij} and s_{ij}^* indicate the *j*th simulated and observed drawdown data of the *i*th pumping test. Figure 5.9 summarizes NMAE and NRMSE and codes them in contrasting colors to better visualize their rankings.

Krig_ordinary_5 had the largest misfits for almost all pumping tests. NMAE and NRMSE were significantly reduced when three more NMR logs were involved for spatial correlation. However, NMAE and NRMSE were not further decreased by applying indicator kriging or using permeameter-based *K* for conditioning. Compared with the kriging methods, groundwater flow models based on the zonation methods yielded improved drawdown fits, especially by examining NRMSE, which was more sensitive to large misfits. Drawdown prediction performances were further improved after characterizing the finer-scale heterogeneity internal to each layer. Here, drawdown fits of *Multi-level_5* were comparable to *Multi-level_8*, and their prediction performances were also at the same level as *Multi-level_9 (permeameter)*.

NMAE	PW1-1	PW3-1	PW1-3	PW1-4	PW2-3	PW3-3	PW5-3	PW4-3	PW1-5	PW5-5	Overall
Krig_ordinary_5	0.74	0.63	2.11	2.61	1.72	0.74	1.72	4.10	2.71	2.40	0.58
Krig_ordinary_8	0.59	0.50	1.36	1.46	1.23	0.55	0.93	1.50	1.08	1.07	0.36
Krig_indicator_8	0.59	0.56	1.47	1.87	1.12	0.65	1.03	2.35	1.71	1.44	0.42
Krig_ordinary_9 (permeameter)	0.51	0.45	1.27	2.04	1.26	0.51	1.17	1.17	1.84	2.06	0.41
Zonation_geology_8	0.52	0.41	0.87	1.13	1.21	0.56	0.65	0.48	0.35	0.30	0.25
Zonation_clustering_8	0.39	0.36	0.97	1.66	0.71	0.42	0.68	0.18	0.23	0.38	0.25
Multi-level_5	0.38	0.35	0.62	1.02	0.53	0.39	0.54	0.26	0.33	0.45	0.19
Multi-level_8	0.36	0.37	0.79	1.13	0.59	0.42	0.58	0.25	0.30	0.36	0.21
Multi-level_9 (permeameter)	0.40	0.30	0.43	1.26	0.57	0.53	0.48	0.61	0.68	0.45	0.14

NRMSE	PW1-1	PW3-1	PW1-3	PW1-4	PW2-3	PW3-3	PW5-3	PW4-3	PW1-5	PW5-5	Overall
Krig_ordinary_5	0.83	0.72	2.26	2.68	2.27	0.86	2.01	4.20	2.74	2.56	2.30
Krig_ordinary_8	0.69	0.59	1.58	1.62	1.55	0.70	1.18	1.56	1.10	1.26	1.24
Krig_indicator_8	0.68	0.63	1.66	1.97	1.38	0.80	1.24	2.43	1.74	1.62	1.51
Krig_ordinary_9 (permeameter)	0.63	0.55	1.48	2.50	1.75	0.72	1.63	1.31	2.13	2.80	1.79
Zonation_geology_8	0.64	0.51	1.20	1.40	1.89	0.73	0.91	0.55	0.44	0.36	0.96
Zonation_clustering_8	0.53	0.45	1.16	1.86	0.92	0.55	0.90	0.20	0.29	0.50	0.89
Multi-level_5	0.56	0.49	0.83	1.29	0.71	0.58	0.70	0.29	0.40	0.56	0.70
Multi-level_8	0.52	0.48	1.01	1.37	0.77	0.58	0.75	0.28	0.36	0.46	0.74
Multi-level_9 (permeameter)	0.59	0.40	0.53	1.41	0.73	0.74	0.69	0.65	0.75	0.60	0.76

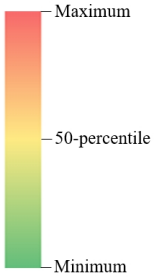


Figure 5.9: Summary of the color-coded NMAE and NRMSE for the comparison between the simulated and observed drawdown responses. Color is coded vertically by each pumping test, in which maximum and minimum metric values are labeled in red and green. The change in color is linear to the change in metric values. Note that the colors are given independently for each column, not for comparison between different pumping tests.

5.7 Discussion

5.7.1 Comparison of Drawdown Prediction Performance

Different locations within Aquifer 1, Aquifer 2, and the Top layer were stressed through independent pumping tests to examine the predictive capability of various spatial *K* models implemented in groundwater flow models. Such a thorough examination also reduces the possibility of biasing advantageous results. Among various spatial *K* models, the ordinary kriged models usually overpredict drawdowns, suggesting that *K* connectivity is not adequately represented. It is caused by incorporating all neighboring observations as conditioning data for interpolation. Therefore, *K* estimation in regions away from the logging boreholes approaches the geometric mean of the entire data domain. For example, the 3D *K* field of *Multi-level_8* (Figure 5.5h) exhibits 1 m of Aquifer 2 in PW3 with its *K* in 1×10^{-4} m/s, while the corresponding region in *Krig_ordinary_8* only presents a trace of discontinuous high-*K* materials. Furthermore, the drawdown prediction performance of the ordinary kriged models decreases

considerably when the number of conditioning data reduces (i.e., *Krig_ordinary_5*). This fact suggests that a representative ordinary-kriged model requires abundant observations for conditioning. In addition, the predictive capability of the kriging-based model is not improved when other assumptions of spatial connectivity are adopted (e.g., discrete facies in indicator kriging). Conversely, the prediction performance of *Krig_indicator_8* lags slightly behind *Krig_ordinary_8*. This may result from the expected *K* field in indicator kriging estimation, which considers the full range of possible values in the estimated cumulative distribution function to calculate *K*; however, the incorporation of less relevant distributions also results in biased *K* estimation. Therefore, the local geological information was not accurately depicted using indicator kriging in this layered porous medium.

The improved drawdown predictions in the zonation cases support the idea that the general *K* trends can be captured well using zonation methods. Thus, it is feasible to adequately reproduce the layer delineation based on the interpolation of layer boundaries along boreholes. Figure 5.9 shows that *Zonation_clustering_8* has comparable NMAE values, but a smaller NRMSE than *Zonation_geology_8*. It suggests that the 3D *K* field of *Zonation_geology_8* presents quite a few large discrepancies in layering. A potential bias of the zonation based on the visual interpretation of geologic logs is that core samples may be lost or compressed during the drilling process, causing depth misalignment of layer boundaries. Thus, the spatially associated clustering using high-resolution vertical *K* estimates (e.g., from borehole NMR) may be more reliable in delineating zone geometry for a complex layered porous medium.

The incorporation of intralayer heterogeneity further improves drawdown predictions. Here, we summarize the geometric mean of *K*, the variance of ln *K* ($\sigma^2_{\ln K}$), maximum and minimum *K* of each aquifer and aquitard layer of *Multilevel_8* in Table 5.4. Table 5.4 shows that Aquifer 2 has gridded *K* values varying by almost two orders of magnitude, with $\sigma^2_{\ln K} > 1$. Therefore, we argue that it may be necessary to represent the heterogeneity in zonal *K* for highly heterogeneous, layered geologic deposits.

Moreover, it is noticeable that the drawdown prediction performances based on multi-level heterogeneity characterization can remain at the same level, though reducing the number of NMR logs.

The two groundwater flow models based on the permeameter measurements yield no better drawdown predictions than the optimal NMR-based models using the same spatial interpolation methods. Considering more input data used in permeameter-based models, their performances were supposed to be slightly improved compared to NMR-based models. Again, the error may result from the soil loss and compression during core sampling. For example, sample loss can be found between 7.5 – 8.5 mbgs in permeameter measurements of CMT1 (refer to Figure 5.3a). The depth misalignment is detrimental to drawdown predictions particularly in complex layered porous media such as the NCRS. Additionally, the permeameter *K* may be biased towards the vertical component of *K* due to vertical sample collection. In contrast, NMR-derived *K* does not assume any potential directional component of *K*. Furthermore, we notice that large drawdown misfits occur in the PW1-3 test for all NMR-based models, while the observed drawdown was satisfactorily predicted in *Multi-level_9 (permeameter)*.

The relatively large misfits (e.g., in the PW1-3 test) in the NMR-based models may originate from the different local sorting in the multi-screened wells (e.g., PW1). Bimodal distributions of relaxation time were presented in the PW1 and PW5 logs, which were not observed in the four CMT wells and the 2" wells (e.g., PW2-3, PW4-3). Figure B7 presents the relaxation time distributions of eight NMR logs. This bimodality exhibits the existence of abundant low-*K* materials, which may come from the fine content deposition near the well domain. Such bimodal distributions of relaxation time caused the underestimation of *K* due to the left shift of the mean-log T_2, resulting in enlarged simulated drawdowns for the PW1-3 test (see Figure 5.8).

Table 5.4: Statistics of *K* estimates within each layer of the spatial *K* model derived from the multi-level heterogeneity characterization of eight NMR logs (Case: *Multi-level_8*).

Layer ID	Mean *K* (m/s)	$\sigma^2_{\ln K}$	Max *K* (m/s)	Min *K* (m/s)

Top	1.96×10^{-6}	0.150	9.88×10^{-6}	3.66×10^{-7}
Aquitard 1	2.06×10^{-7}	0.244	8.97×10^{-7}	3.61×10^{-8}
Aquifer 1	1.04×10^{-5}	0.353	4.88×10^{-5}	2.18×10^{-6}
Aquitard 2	2.88×10^{-7}	0.467	2.76×10^{-6}	2.48×10^{-8}
Aquifer 2	8.12×10^{-5}	1.031	6.47×10^{-4}	5.88×10^{-6}
Aquitard 3	5.25×10^{-7}	0.945	1.65×10^{-5}	1.86×10^{-8}
Overall	1.07×10^{-6}	4.049	6.47×10^{-4}	1.86×10^{-8}

5.7.2 Usefulness of NMR Logging for Characterizing Spatial Heterogeneity

So far, we have demonstrated that NMR logging is useful in determining 3D *K* distributions, and the most representative *K* field is derived from the multi-level heterogeneity characterization. The glacial aquifer system at our site is highly heterogeneous, comprised of six alternating layers of aquifers and aquitards with substantial *K* variations in both horizontal and vertical directions. Still, the spatial *K* model based on five NMR logs can adequately represent the spatial heterogeneity for groundwater flow simulations. We advocate that such a framework should be used for other hydrogeologic environments to determine reliable 3D *K* distributions for groundwater flow simulation.

Based on our study, the broader use of NMR logging can be envisioned. For example, NMR logging also yields estimates of total porosity and effective porosity in the saturated zone. The 3D porosity distributions could be readily determined from similar spatial modeling procedures. The accurate representations of porosity and *K* are critical to the numerical simulations of solute transport within an advective flow regime. Therefore, the use of NMR logging has the potential to accurately predict solute transport behaviors. Furthermore, new borehole NMR tools (e.g., DP140J from Vista Clara Inc., see Morozov et al., 2024 for details) with a vertical resolution of 4 cm can be employed to characterize spatial heterogeneity at a finer scale, which should facilitate the recognition of preferential flow paths in solute transport.

6 Integrating NMR Logging with Three-Dimensional Steady-State Hydraulic Tomography for Improved Mapping of Hydraulic Conductivity Distribution

Citation:

Wang, C., & Illman, W. A. (2025). Integrating NMR Logging with Three-Dimensional Steady-State Hydraulic Tomography for Improved Mapping of Hydraulic Conductivity Distribution. Submitted to *Journal of Hydrology* (Under review).

6.1 Overview

Hydraulic tomography (HT) analysis is a robust approach to mapping the spatial distribution of hydraulic conductivity (K). However, it typically requires a dense monitoring network for acquiring multiple head responses to pumping. Smooth tomograms can result after decreasing the number of pumping tests or the observation interval density. Nuclear magnetic resonance (NMR) logging has been applied in the last decade for subsurface characterization of unconsolidated sediments. In this study, the integrated models of NMR logging and HT analysis were constructed for improved K estimation in a highly heterogeneous glaciofluvial deposit. Moreover, the impact of limited calibration data on the HT performance was investigated, in which a reduced number of pumping tests and fewer observation points were considered. Results show that improved drawdown predictions of independent pumping tests can be obtained by incorporating the spatial K models from NMR logging as the initial guess for geostatistical inverse modeling of HT analysis. Various NMR K models can be successfully calibrated with refined inter-well spatial K variability. These models include an ordinary kriging model, a clustering-based zonation model, and a hybrid multi-level heterogeneity model nesting the intralayer kriged K into the

zonal representation. For HT analysis with a limited dataset, the integration of NMR logging can significantly improve the resolution of K tomograms, resulting in more reliable drawdown predictions compared to the K model solely based on HT analysis. This study advocates that the joint use of HT analysis and NMR logging is more efficient and robust for subsurface heterogeneity characterization in complex aquifer systems.

6.2 Introduction

Groundwater flow and solute transport investigations require accurate knowledge of hydraulic properties (e.g., hydraulic conductivity (K), porosity (θ), and specific storage (S_S). Therefore, hydrogeologists have used various techniques (e.g., hydraulic tests, geophysical surveys, and tracer tests) to characterize the subsurface and determine hydraulic parameters. Traditional hydraulic tests (e.g., slug tests, pumping tests) usually provide direct estimates of hydraulic properties. However, these methods typically assume that the subsurface is homogeneous, regardless of its heterogeneous nature. Wu et al. (2005) found that conventional analysis of pumping tests (i.e., analytical solutions) is insufficient in predicting accurate drawdown behaviors, and the estimated K may not be representative of the aquifer.

To evaluate the representativeness of 3D K distributions, an effective approach is to predict drawdown responses at multiple observation intervals during aquifer pumping tests and compare them to field observations. Previous work has shown that accurate predictions of drawdown responses can be difficult to achieve based on standard in-situ borehole hydraulic measurements (e.g., slug tests, packer tests), particularly in complex environments (Berg & Illman, 2015; Illman et al., 2010; Sun et al., 2024). Research has shown that more accurate depictions of 3D K distributions for reliable numerical groundwater flow predictions can be achieved using hydraulic tomography (HT) analysis (Berg & Illman, 2011b; Cardiff et al., 2013; Hochstetler et al., 2016; Tiedeman & Barrash, 2020).

HT was proposed and developed over the past three decades. During the HT survey, targeted aquifers are stressed by the process of injecting/pumping water at the screened intervals. Meanwhile, a monitoring network with multiple observation intervals is installed to collect drawdown signals over large volumes. The analysis of HT survey data relies on the joint inverse modeling of hydraulic head data from multiple pumping tests to interpret the spatial heterogeneity of hydraulic parameters. Therefore, HT is a cost-effective method to collect non-redundant drawdown datasets for accurate characterization of subsurface heterogeneity (Hou et al., 2023; Zha et al., 2015). So far, HT has been demonstrated as a mature technique to provide reliable estimates of spatial distributions of hydraulic parameters through synthetic studies (e.g., Cardiff & Barrash, 2011; Gottlieb & Dietrich, 1995; Luo et al., 2020; Tso et al., 2015; Wang & Illman, 2023; Yeh & Liu, 2000; Zhu & Yeh, 2005, 2006), laboratory experiments (e.g., Berg & Illman, 2011a; Illman et al., 2007, 2008, 2010, 2012; Liu et al., 2002; Liu et al., 2007a; Luo et al., 2017; Zhao & Illman, 2021), field studies (e.g., Berg & Illman, 2011b, 2013, 2015; Bohling et al., 2007; Brauchler et al., 2013; Cardiff et al., 2013; Huang et al., 2011; Illman et al., 2009; Klepikova et al., 2020; Luo et al., 2023; Ning et al., 2023; Tiedeman & Barrash, 2020; Zha et al., 2016; Zhao et al., 2023; Zhao & Illman, 2017, 2018, 2022a, 2022b) and basin-scale aquifer characterization (e.g., Daranond et al., 2020; Liu et al., 2020a; Liu et al., 2020b; Luo et al., 2022; Xia et al., 2023; Yeh et al., 2009).

Geostatistical and stochastic inverse methods have been developed to interpret the spatial K variability of the aquifer-aquitard systems (e.g., Gutjahr & Wilson, 1989; Harvey & Gorelick, 1995; Hoeksema & Kitanidis, 1984, 1989; Kitanidis, 1995; Kitanidis & Vomvoris, 1983; Rubin & Dagan, 1987; T. C. J. Yeh et al., 1996; Yeh et al., 1995). Geostatistical parameters of the aquifer systems were incorporated into the inverse modeling, in which the most likely spatial distribution of hydraulic parameters and their uncertainty are estimated. Among various geostatistical inversion codes, Successive Linear Estimator (SLE) (Yeh & Liu, 2000; see also Xiang et al., 2009; Zhu & Yeh, 2005) has been extensively applied in HT analysis. The SLE code originates from the linear cokriging estimator. To

resolve the nonlinear relationship between hydraulic parameters and observations (e.g., drawdown responses), the SLE code iteratively updates the estimated parameter fields and their conditional covariances during the inversion process. Xiang et al. (2009) proposed the Simultaneous Successive Linear Estimator (SimSLE), in which head data from all pumping tests were simultaneously interpreted to calibrate the spatial distribution of K and S_S.

Challenges remain despite the robust performance of the geostatistical inverse models for HT analysis. A main issue with the practical application of HT is that observation interval density can be sparse. When fewer pumping tests and observation intervals are applied in a field HT survey, smooth K tomograms (i.e., spatial K distributions) can result (e.g., Berg & Illman, 2011b; Cardiff et al., 2013; Illman et al., 2009). Furthermore, stratigraphic boundaries of the layered aquifer systems are still smoothed even when a sufficient number of drawdown data are used for HT analysis (e.g., Wang & Illman, 2023; Zhao et al., 2019), which may result in inaccurate mapping of preferential flow pathways (Bianchi et al., 2011b; Zheng et al., 2011) for predictions of advective transport.

Researchers considered incorporating stratigraphic information from other sources (e.g., geological information, borehole hydraulic testing data, and geophysical survey data) into the HT analysis to improve the HT performance. Based on the interpretation of stratigraphic information along boreholes, Zhao et al. (2016) delineated zone geometry for a 2D sandbox aquifer system and constructed geological zonation models with internally homogeneous K values for each zone. The geological zonation models served as the initial K distribution for the geostatistical inversion of the HT analysis. The results of Zhao et al. (2016) and subsequent field studies (e.g., Luo et al., 2017; Zhao & Illman, 2017, 2018, 2022a) demonstrated that the HT estimates were more representative of the geological conditions and can yield improved drawdown prediction performance after incorporating reliable stratigraphic information into HT analysis. Other attempts to construct initial K models utilized different spatial interpolation approaches to generate initial K distributions based on the borehole testing data. For example, Zhao et al. (2023)

generated spatial *K* distributions based on HPT-estimated *K* along boreholes using ordinary kriging interpolation. The HPT-based spatial *K* model was further calibrated using HT analysis, resulting in improved drawdown prediction performance. To date, many efforts have been made to enhance the characterization of site heterogeneity by integrating different types of prior information, including geology logs (e.g., Zhao & Illman, 2017, 2018), flowmeter data (e.g., Luo et al., 2023), borehole hydraulic testing methods (e.g., Zhao et al., 2023), and geophysical data (Wang & Illman, 2023).

Geophysical surveys, especially borehole geophysical logging techniques, have been widely used to characterize subsurface heterogeneity, which are usually cost-effective and cause little to no disturbance to the subsurface. However, few attempts were made to jointly use geophysical and HT surveys to determine the spatial distribution of *K*. A possible reason is that few petrophysical relationships have been built directly with *K*. Therefore, only hydrostratigraphic information can be inferred from the geophysical survey. For example, Wang and Illman (2023) investigated the integration of borehole logging data of electrical conductivity (EC) and gamma ray (GR) radiation with HT analysis in a numerical sandbox study. They found that improved HT estimates can result when more reliable zone geometry can be delineated from the clustering of multiple geophysical logging data; however, zonal *K* values can only be derived using hydraulic data (e.g., permeameter measurements and the interpretation of pumping test results).

Borehole nuclear magnetic resonance (NMR) is a novel geophysical logging approach for hydrogeophysical characterization. Conventional geophysical logging approaches (e.g., EC and GR logging) typically yield indirect information about *K*. In comparison, the geophysical applications of NMR use the nuclear spin of hydrogen protons in water molecules. The nuclear spin decay yields a temporal distribution, indicating that relaxation processes interact with disparate pore structures within the porous medium (see review by Behroozmand et al., 2015). As a result, it can be related to the pore characteristics and be further used to interpret *K*. With the development of slim-hole NMR logging tools

(Perlo et al., 2013; Sucre et al., 2011; Walsh et al., 2013; Zhu et al., 2021), NMR's utility is enabled for fast application in shallow unconsolidated sediments. Various petrophysical relationships have been developed to estimate *K* from NMR signals (Coates & Dumanoir, 1973; Dlugosch et al., 2013; Kenyon et al., 1988; Seevers, 1966; Timur, 1968; Walsh et al., 2013). Evaluation of these petrophysical relationships (e.g., Crow et al., 2022; Dlubac et al., 2013; Kendrick et al., 2021, 2023; Knight et al., 2016; Maurer & Knight, 2016; Wang et al., 2025a) has revealed that the NMR-derived *K* estimates are very consistent with *K* measurements from hydraulic tests. The optimal model fit in *K* is typically within one order of magnitude (with a unit of m/s), making the NMR technique a reliable tool for high-resolution site characterization along boreholes. It is important to keep in mind that existing NMR logging tools yield information on *K* around boreholes with a small support volume with a radius of several tens of centimeters, and not between boreholes.

To better map spatial *K* variability between boreholes, Wang et al. (2025b) correlated *K* patterns observed along multiple NMR logging locations using various spatial interpolation approaches, including ordinary kriging estimation, clustering-based zonation, and a hybrid multi-level heterogeneity characterization approach nesting the zonally kriged *K* into each zone. The results from Wang et al. (2025b) showed that the optimal NMR-derived spatial *K* model can adequately replicate the *K* profiles from permeameter measurements with a wide range of *K* estimates, accounting for the glacial soils from gravel, sand, silt, and clay. Meanwhile, the numerical simulations based on the NMR *K* models can reasonably reproduce the observed drawdown in multiple pumping tests. Following Wang et al. (2025b), we wonder whether the integration of NMR-derived *K* models with HT-based *K* models can further improve the representation of spatial *K* fields in complex aquifer systems, particularly when only a limited HT dataset is available for calibration of the geostatistical inverse models, which has been considered to yield smooth *K* tomograms. Another research topic is which spatial interpolation approach

(e.g., ordinary kriging, zonal representation, multi-level heterogeneity characterization) can lead to an optimal initial *K* model for HT analysis.

In this study, we first established two HT calibration datasets with different numbers of steady-state head observations. The full HT dataset contains the observed head data from all nine wells induced by eight pumping tests, corresponding to a HT survey with sufficient data density for mapping the spatial *K* variability in this complex porous medium. The limited HT dataset incorporates head data from four observation wells and a central pumping well utilized in four pumping tests. Such a relatively low-density HT survey only has 1/3 of the head records for model calibration efforts compared to the full dataset, resulting in a smooth *K* tomogram. Then, three NMR-derived spatial *K* models using various spatial correlation approaches, as previously reported by Wang et al. (2025b), as well as the homogeneous *K* field with no prior *K* information, were calibrated using the steady-state head data via HT analysis. The drawdown prediction performance of all spatial *K* models, including the NMR-derived, the HT-estimated, and their integrated models, was evaluated by comparing the simulated with the observed drawdown from five independent pumping tests. In addition, the estimated *K* profiles at well locations were compared to permeameter *K* estimates. The innovative aspects of this study include: 1) its first field application of integrating borehole NMR data with HT analysis for improved mapping of *K* heterogeneity; 2) examining various spatially interpolated *K* models as initial *K* realizations for improved HT results; and 3) exploring the usefulness of integrating NMR logging with the HT analysis for reliable *K* estimates using a more cost-effective and practical design of the HT survey.

6.3 Site and Data Descriptions

6.3.1 Site Description

The NMR logging surveys and pumping tests used in this study were performed at the North Campus Research Site (NCRS) located on the University of Waterloo, Ontario, Canada. The

hydrogeology for the shallow surface of the NCRS has been extensively studied over the past two decades, yielding fruitful outcomes of subsurface characterization of hydraulic properties (e.g., Alexander et al., 2011; Berg & Illman, 2011b, 2013, 2015; Luo et al., 2023; Wang et al., 2025b; Zhao et al., 2023; Zhao & Illman, 2017, 2018). The shallow surface is comprised of highly heterogeneous glaciofluvial deposits and till, with soils from gravel, sand, silt, to clay. Based on permeameter measurements, Alexander et al. (2011) reported that the *K* values (m/s) within the NCRS span over six orders of magnitude. Specifically, a dense and stiff Catfish Creek Till located at approximately 16 meters below ground surface (mbgs) is regarded as a hydraulic barrier at the bottom of the aquifer system of interest. A sandy-to-clayey silt sequence, namely the Maryhill Till, is deposited over the Catfish Creek Till. Thin erosional remnants of the Tavistock Till overlie the Maryhill Till, which is below the top organic soil. Alternating layers of aquifer and aquitard units were found based on borehole geology logs (Alexander et al., 2011; Berg & Illman, 2011b; Zhao & Illman, 2017). The primary "aquifer zone" is situated between 8 and 13 mbgs, with low *K* materials separating two aquifer layers. The upper aquifer layer, namely "Aquifer 1", consists of sand to sandy silt, and the lower aquifer layer ("Aquifer 2") is composed of sandy gravel. Previous pumping records indicated that the "aquifer zone" is confined.

Nine multi-level monitoring wells were instrumented in an area of 15 × 15 m at the NCRS. A well map and the locations of their screens are provided in Figure 6.1. Three 4-inch multi-screened pumping wells and two well clusters were located at the four corners and the center of this well area. The multi-screened wells, including PW1, PW3, and PW5, contained 1-m screens separated from adjacent screens by 1-m PVC pipe. PW2 and PW4 are well clusters, each comprised of three 2-inch single-screened pumping wells, numbered from shallow to deep. For example, PW2-3 is the third interval to the ground surface in PW2. The bottom 1-m tubing of each single-screened well was screened. In addition, four Continuous Multichannel Tubing (CMT) wells (Solinst Canada Ltd.) were installed at the midpoint of

each side. Each well contains seven channels for observing hydraulic head, which are evenly spaced at a 2-m distance.

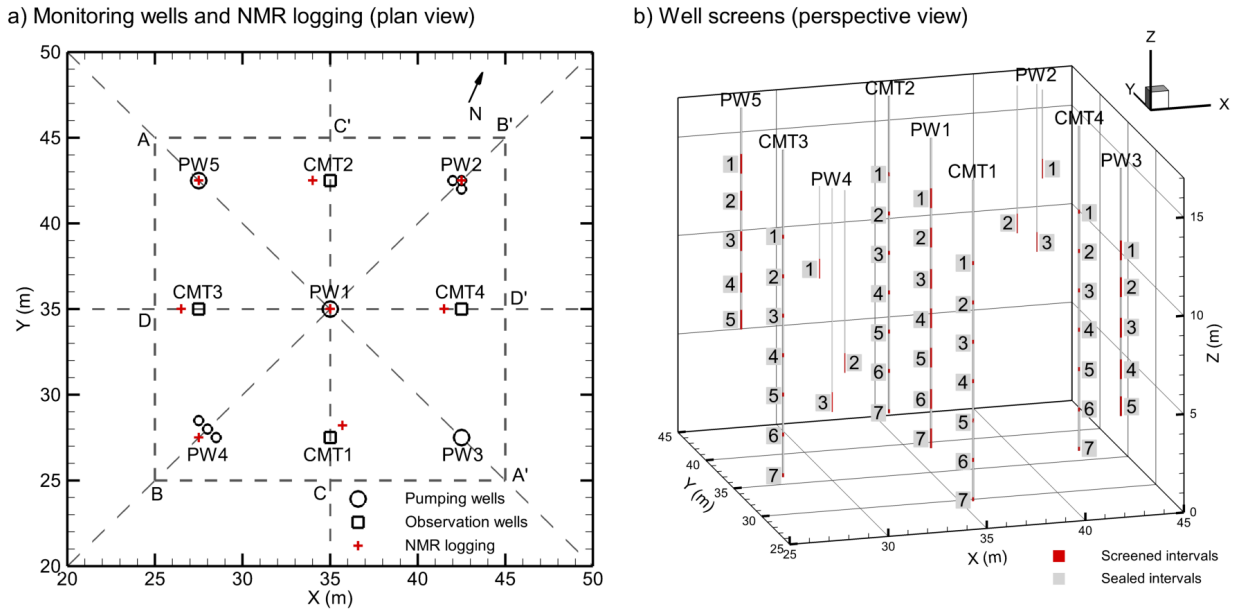


Figure 6.1: Locations of monitoring wells for pumping tests and NMR logging. a) Plan view of the locations of the nine monitoring wells and NMR logging. The dashed-line square (A-B'-A'-B) delineates the area of the data domain in 20 m by 20 m, and the coordinates are consistent with the simulation domain of the numerical modeling of groundwater flow. b) Perspective view of the well screens. The CMT wells have a screened interval in each channel, which is conceptualized as well screens in this schematic plot.

6.3.2 Steady-State HT Survey

Previously reported by Berg and Illman (2011b) and Zhao and Illman (2017), 13 pumping tests were conducted at the NCRS, yielding significant drawdown data reaching steady-state flow conditions. Table 6.1 summarizes the 13 pumping tests. For this study, in total, 411 steady-state head measurements were used for model calibration and/or validation of the HT analysis. Two calibration datasets were selected to investigate the impacts of limited head data for model calibration. The "full" HT dataset leveraged the full capabilities of the monitoring network at the NCRS, incorporating eight pumping tests that stressed the highly heterogeneous glaciofluvial deposit at multiple depths and various well locations, along with head records from all nine observation wells. Therefore, four pumping tests were selected in

the screens spaced equally in PW1, and another four pumping tests were chosen from the other four pumping wells located at the four corners of the square pumping zone (refer to Figure 6.1a). In summary, the full HT dataset (referred to as *HT_8*) comprises 242 head measurements from eight pumping tests conducted in PW1-1, PW1-3, PW1-5, PW1-7, PW2-3, PW3-3, PW4-3, and PW5-5.

Table 6.1: Summary of the pumping tests for model calibration of the *HT_4* and *HT_8* scenarios and their validation.

Operation screens	Stress depth* (mbgs)	Stress type	Rates (L/min)	Total number of heads	Data use
PW1-3	7.46	Pumping	10.5	33(19†)	*HT_8* and *HT_4*
PW3-3	7.62	Pumping	2.1	31(21†)	*HT_8* and *HT_4*
PW1-5	11.46	Pumping	4.4	28(14†)	*HT_8* and *HT_4*
PW5-5	11.61	Pumping	8.1	41(30†)	*HT_8* and *HT_4*
PW1-1	3.36	Injection	1.89	22	*HT_8*
PW2-3	8.00	Pumping	1.91	25	*HT_8*
PW4-3	11.55	Pumping	30.2	28	*HT_8*
PW1-7	15.56	Pumping	1.05	34	*HT_8*
PW3-1	3.69	Injection	0.94	20	Validation
PW5-3	7.62	Pumping	7.8	42	Validation
PW1-4	9.46	Pumping	6.3	37	Validation
PW5-4	9.73	Pumping	7.8	39	Validation
PW1-6	13.56	Pumping	0.95	31	Validation

* Stress depth indicates the mid-depths of the pumping/injection screens.

† The number of heads within the brackets indicates the heads for the limited calibration dataset, *HT_4*.

On the other hand, the limited HT dataset utilized PW1-3, PW1-5, PW3-3, and PW5-5 for model calibration. Different from the *HT_8* scenarios, only observation data from PW1 and four CMT wells were incorporated here. As a result, the limited HT dataset (referred to as *HT_4*) included 83 steady-state head measurements.

6.3.3 NMR Logging Survey

Eight borehole NMR logs were performed by Wang et al. (2025b) at multiple well locations of the NCRS. Four NMR logs were obtained in PW1, PW2-3, PW4-3, and PW5, using a wireline to lower

the NMR probe. Another four NMR logs were obtained within 1 m of the CMT wells (CMT1, CMT2, CMT3, CMT4), which were deployed using a direct-push rig. Various probes from Vista Clara Inc. (Mukilteo, Washington, USA) were used to facilitate logging at different locations. As the auger used to drill the 4" pumping wells had an approximate diameter of 0.26 m, the JPY350 probe was used to measure NMR signals in PW1 and PW5, which sampled cylindrical shells of soil with diameters of 0.33 m and 0.38 m, corresponding to different frequencies. The exploration depths were 15.9 mbgs and 11.2 mbgs in PW1 and PW5, and the vertical resolution of the JPY350 was 0.41 m. The DP175 probe was used in the two deepest wells (i.e., PW2-3 and PW4-3) of the well clusters, with logging depths reaching 8.2 mbgs and 11.7 mbgs, respectively. The vertical resolution of the DP175 was 0.23 m. The DP140N was designed for use with direct-push drill rigs. The DP140N probe reached approximately 12 mbgs for all logging locations (i.e., four CMT wells). The vertical resolution of the DP140N was 0.22 m. During the field survey, the logging interval in PW1 and PW5 was 0.3 m, while the logging interval in the other locations was 0.15 m. In total, 361 NMR measurements were obtained at eight borehole locations.

The following form of the Schlumberger Doll Research (SDR) model (Kenyon et al., 1988) was used to estimate *K* based on the NMR signals collected at the NCRS:

$$K_{SDR} = C_{SDR} T_{2ML}^2 \tag{6.1}$$

where K_{SDR} is the hydraulic conductivity (m/s) estimated using the SDR model, C_{SDR} is a model fitting parameter, which accounts for mineralogical parameters (e.g., surface relaxivity) and was reported to be stable for similar sedimentations (Crow et al., 2022; Kendrick et al., 2023; Knight et al., 2016; Wang et al., 2025a), and T_{2ML} is the mean-log time of the relaxation time distribution. Wang et al. (2025a) calibrated C_{SDR} using the permeameter measurements at the NCRS. Therefore, an optimal C_{SDR} for glaciofluvial deposits used in this study was 3.86×10^{-3} m/s^3. Figure 6.2 exhibits the NMR-derived *K* profiles at eight borehole locations.

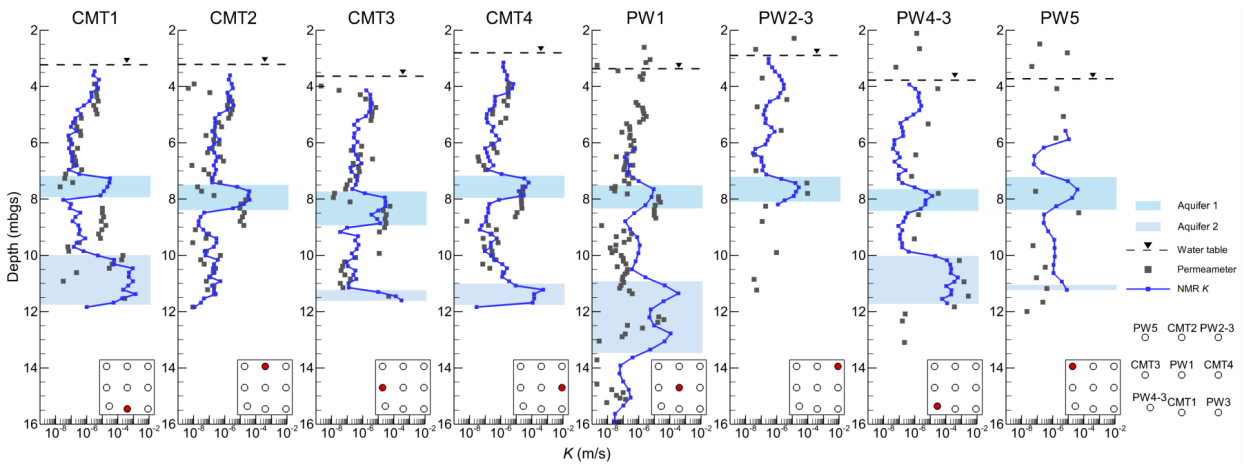


Figure 6.2: The NMR-derived *K* profiles along eight boreholes. Co-located permeameter measurements are provided for comparison. The blue-shaded zones are Aquifers 1 and 2 based on the clustering of NMR *K* data.

Three spatial interpolation approaches were used to derive spatial *K* models based on these NMR logs. The detailed processes of spatial correlation were documented in Wang et al. (2025b). Briefly, the first approach was the ordinary kriging estimation. The variogram analysis of the NMR-derived ln *K* profiles yielded a variance of ln *K* and a vertical correlation length of 7.12 and 2.48 m, respectively, corresponding to the sill and range values for an exponential model of vertical variogram. A ratio of 8 was assigned between the horizontal and vertical correlation lengths to obtain an omnidirectional horizontal variogram model. After ordinary kriging interpolation, the corresponding spatial *K* model was defined as *NMR_krig*. The second approach was a clustering-based zonation, in which spatial constraints were incorporated into a *k*-means clustering (Wang & Illman, 2023) to facilitate the accurate interpretation of layer boundaries along boreholes. The zone geometry was delineated through interpolation and extrapolation of the layer boundaries along boreholes, and the zonal *K* estimates were obtained by calculating the geometric mean of the NMR *K* data within each zone. Hence, an internally homogeneous zonal *K* representation was obtained, defined as *NMR_zonal*. The third approach built a hybrid model of multi-level heterogeneity patterns (defined as *NMR_multi*). Instead of assigning an internally homogeneous *K* value to each zone, this approach used ordinary kriging estimation to populate *K*

variations within each clustering-based zone, where the clustered NMR *K* data served as hard data for kriging. The spatial correlation of NMR *K* profiles was performed within a 20 m × 20 m area that encompassed the NMR logs (see the dashed-line square in Figure 6.1a).

6.4 Hydraulic Tomography Analysis

6.4.1 Numerical Models of Groundwater Flow

Numerical simulations of groundwater flow were conducted using the integrated hydrologic modeling code, HydroGeoSphere (Aquanty Inc., 2025). To mitigate the effects of lateral boundary conditions, the simulation domain was assigned dimensions of 70 m × 70 m × 17 m, with PW1 positioned at its center. This simulation domain was discretized into variably-sized rectangular finite elements, which were gradually refined from 5 m × 5 m × 0.5 m near the lateral boundaries to 0.5 m × 0.5 m × 0.3 m in the vicinity of the well locations (see Figure C1 for the model grid). The top and bottom boundaries were assigned as impermeable, and constant-head conditions were assigned for the lateral boundaries. To highlight the differences in *K* estimates from multiple cases of HT analyses, the comparison of *K* fields focused on a 20 m × 20 m area of the "data domain".

Three NMR-derived spatial *K* models in the data domain were gridded and shown in Figure 6.3. Figure 6.3 illustrates that the *NMR_zonal* model identified two aquifer layers, Aquifers 1 and 2, through clustering of NMR *K* profiles. These two aquifer layers were bounded or intersected by three aquitard layers: Aquitards 1, 2, and 3. An additional layer (i.e., the Top layer) was identified near the top boundary. The *K* values at the outer edge of these NMR-derived *K* models (i.e., 20 m × 20 m) were extrapolated to the lateral boundaries of the 70 m × 70 m simulation domain. Figure C2 exhibits the NMR *K* fields for the entire simulation domain. These NMR models served as initial *K* distributions for HT analyses of the integrated models.

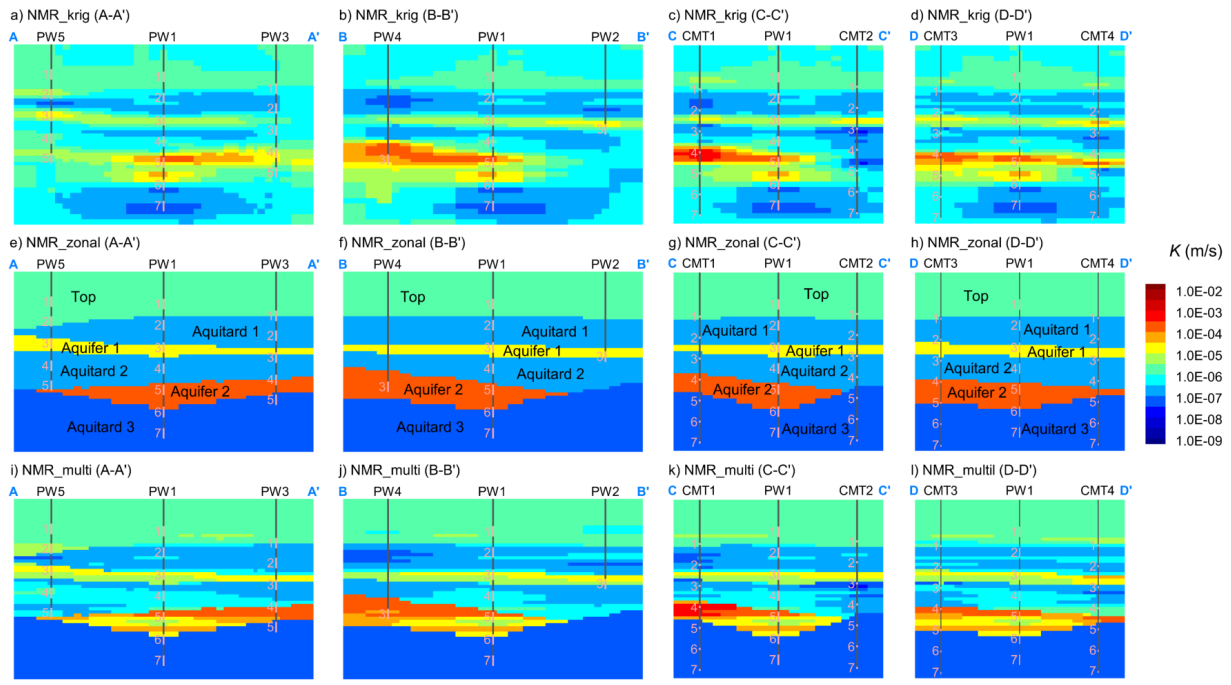


Figure 6.3: Contour plots of NMR-derived *K* distributions in four 2D cross-sections based on the spatial correlation of NMR logging data. a)-d), e)-h), and i)-l) represent the A-A', B-B', C-C', and D-D' cross-sections of the *NMR_krig*, *NMR_zonal*, and *NMR_multi* cases of spatial *K* models, respectively. The well screens are labelled and highlighted. Six zones, including Aquifers 1 and 2, Aquitards 1, 2, and 3, and the Top layer, are labelled on the *NMR_zonal* model in the figure parts e)-h).

Four comparative cases of HT analyses were designed to examine the drawdown prediction performance using HT-estimated *K* with different initial *K* distributions for geostatistical inverse modeling. When NMR logs were not integrated, an effective *K* field was used as the initial *K* distribution for the HT analysis, which treated the simulation domain as homogeneous and isotropic. This initial homogeneous *K* value was determined using inverse modeling with the full HT calibration dataset. Furthermore, three NMR *K* models were applied as the initial guess for HT analyses in different cases. For brevity, the spatial *K* models solely relying on HT analysis for *K* estimation were defined as the "HT models", while the models that jointly used NMR logging data and HT analysis for *K* estimation were defined as the "integrated models" in this study. Another factor is the number of steady-state head measurements used for inverse modeling. Two scenarios were assigned for each case, with *HT_8* and *HT_4* representing the HT analysis using eight and four pumping tests for calibration. Again, the *HT_8*

scenario comprised head observations from all monitoring wells, whereas the *HT_4* scenario only included heads from four CMT wells and PW1. Table 6.2 is a summary of the comparative HT cases.

Table 6.2: Information on the initial *K* distributions and number of head data for comparative cases of HT models.

HT case name	Initial *K* distribution	Number of heads
HT_4-Homo	HT-estimated homogeneous *K*	83
HT_8-Homo		242
HT_4-NMR_krig	NMR *K* using ordinary kriging estimation	83
HT_8-NMR_krig		242
HT_4-NMR_zonal	NMR *K* using clustering-based zonation	83
HT_8-NMR_zonal		242
HT_4-NMR_multi	NMR *K* using a hybrid model of multi-level heterogeneity	83
HT_8-NMR_multi		242

6.4.2 Calibrated K Fields via HT Analysis

An initial effective *K* of 1×10^{-5} m/s was assigned during the optimization of the initial *K* fields in the geostatistical inverse modeling of *HT_4-Homo* and *HT-8-Homo*. The model-independent parameter estimation code, PEST (Doherty, 2015), was used for optimization of the effective *K* field. The *HT_4-Homo* case yielded an estimated *K* of 7.02×10^{-6} m/s, and a corresponding uncertainty was indicated by the 95% confidence interval, which has upper and lower limits of 8.32×10^{-6} m/s and 5.93×10^{-6} m/s, respectively. Likewise, the effective field optimization for the *HT_8-Homo* model yielded a *K* of 1.17×10^{-5} m/s, with upper and lower limits of 1.34×10^{-5} m/s and 1.02×10^{-5} m/s, respectively.

The geostatistical inverse modeling of all comparative cases employed the Simultaneous Successive Linear Estimator (SimSLE) code (Xiang et al., 2009) to calibrate the initial *K* fields (e.g., the effective *K* and the NMR *K* models). To circumvent the non-linearity between ln *K* and heads, SimSLE adopts a linear relationship to condition HT estimates during an iterative process. An optimal iteration step was determined when both the biggest head misfit and the overall mean squared error were decreased to relatively small prescribed values. The calibrated *K* fields of the optimal iteration steps within the data

domain are shown in Figures 6.4 and 6.5, for the cross-sections of A-A' and B-B', as well as C-C' and D-D', respectively. Figures C3 and C4 show the calibration scatterplots between simulated and observed drawdowns for the models using the *HT_4* and *HT-8* calibration datasets. Figures C5 and C6 exhibit the perspective views of the calibrated *K* fields using the *HT_4* and *HT-8* datasets, respectively.

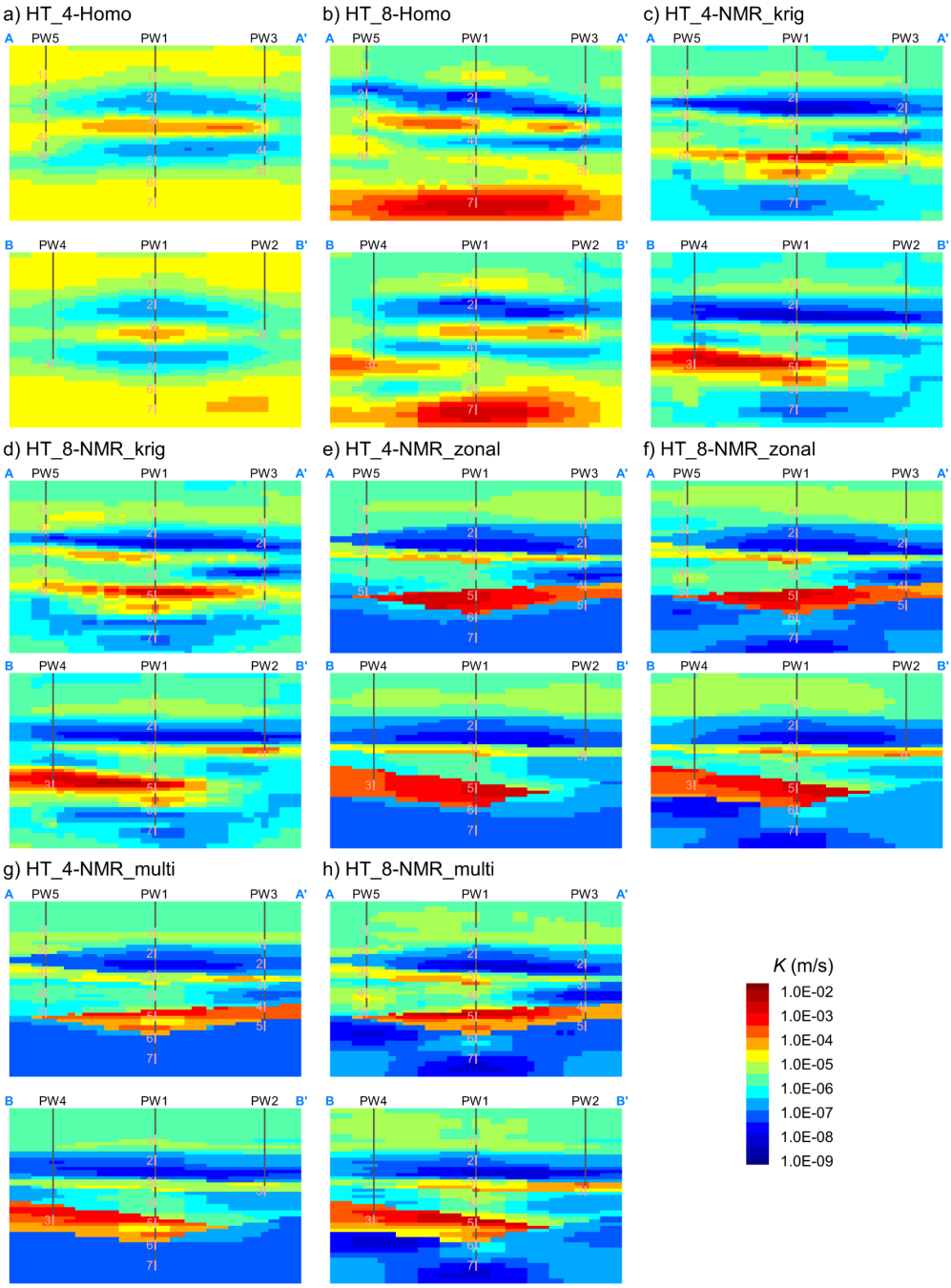


Figure 6.4: Contour plots of *K* distributions in A-A' and B-B' cross-sections based on comparative cases of HT analysis, including a) *HT_4-Homo*, b) *HT_8-Homo*, c) *HT_4-NMR_krig*, d) *HT_8-NMR_krig*, e) *HT_4-NMR_zonal*, f) *HT_8-NMR_zonal*, g) *HT_4-NMR_multi*, and h) *HT_8-NMR_multi*.

In the *HT_4-Homo* model, a relatively smooth *K* tomogram was derived, exhibiting horizontally extended layering. Aquifer 1 was identified by the *HT_4-Homo* model at a similar depth to the "-3" intervals of the CMT wells (see Figure 6.5a), and this layer reached the vicinity of PW3 and four CMT wells, with its *K* values within the range between 1×10^{-5} m/s and 1×10^{-4} m/s. In addition, Aquifer 1 was bounded by two aquitard layers with *K* values on the order of magnitude of 10^{-7} m/s. The bottom 5 m of the simulation domain was estimated to be relatively homogeneous with *K* values on the order of magnitude of 10^{-5} m/s. When more pumping tests with additional observation heads were used for HT analysis, the *HT_8-Homo* model captured more heterogeneity patterns. Compared to Aquifer 1 in the *HT_4-Homo* model that horizontally extended towards PW3 and all CMT wells, Aquifer 1 in the *HT_8-Homo* model also extended to PW2 and PW5 (see Figure 6.4b). More importantly, Aquifer 2 was successfully identified at a depth between the "-4" and "-5" intervals of the CMT wells (see Figure 6.5b). It exhibited variable *K* values between 1×10^{-5} m/s and 1×10^{-4} m/s, which were in a similar range to Aquifer 1. A thick high-*K* zone is shown in Aquifer 2 near PW4 (see Figure 6.4b) and CMT3 (see Figure 6.5b). Additionally, Aquifer 2 gradually got thinner towards the northeast (i.e., the direction of PW2) and pinched out before reaching CMT2, PW2, and CMT4. Furthermore, the *K* values for regions above Aquifer 1 by the *HT_8-Homo* model were estimated to be lower than those by the *HT_4-Homo* model, while higher *K* estimates with a maximum value of $> 1 \times 10^{-3}$ m/s were obtained for the bottom three meters of the simulation domain.

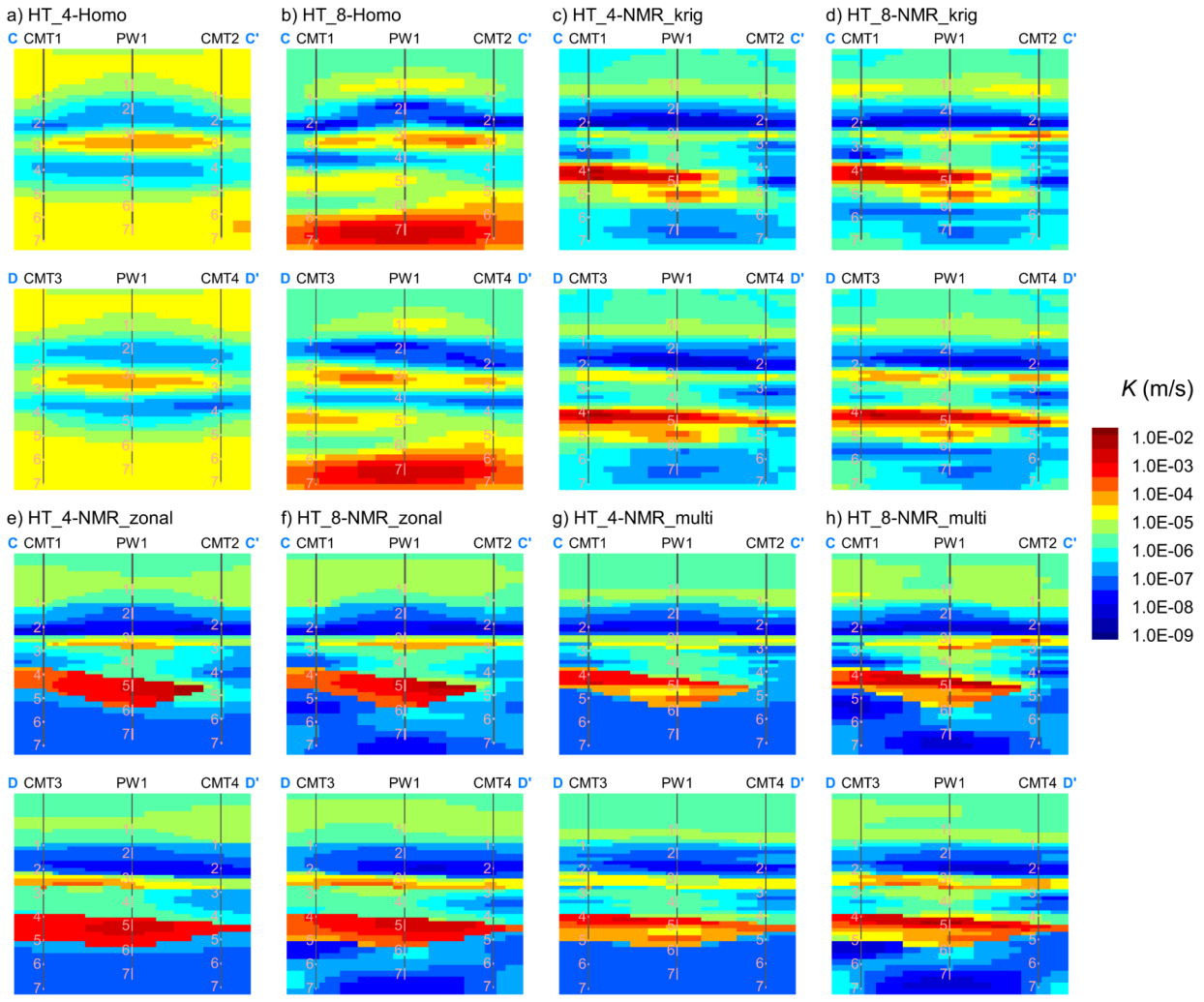


Figure 6.5: Contour plots of *K* distributions in C-C' and D-D' cross-sections based on comparative cases of HT analysis, including a) *HT_4-Homo*, b) *HT_8-Homo*, c) *HT_4-NMR_krig*, d) *HT_8-NMR_krig*, e) *HT_4-NMR_zonal*, f) *HT_8-NMR_zonal*, g) *HT_4-NMR_multi*, and h) *HT_8-NMR_multi*.

Compared to the HT models, the most significant difference in the *K* fields of the integrated models is that the patterns of spatial *K* variability were altered for the lower half of the simulation domain (see Figures 6.4c-h and 6.5c-h). High-*K* regions exist in the bottom several meters of the simulation domain in both HT models. In contrast, the integrated models followed the general *K* patterns of their initial NMR models, characterizing the bottom region of the simulation domain as low-*K* materials. Moreover, Aquifer 2 was estimated to be more conductive in the integrated models, with average *K* values exceeding 1×10^{-4} m/s. Compared to the NMR models, the integrated models yielded more contrasting *K*

variations after calibration. For example, the extent of Aquifers 1 and 2 did not significantly change in models *HT_4-NMR_krig* (Figures 6.4c and 6.5c) and *HT_8-NMR_krig* (Figures 6.4d and 6.5d) after calibrating the initial *NMR_krig* model (Figures 6.3a-d), while the magnitudes of *K* in both Aquifers 1 and 2 increased with decreasing *K* in adjacent aquitard layers, after more head observations were used for calibration in the integrated models initialized with the *NMR_krig* model. In the *HT_8-NMR_krig* model, the *K* estimates within Aquifer 2 reached as high as 1×10^{-3} m/s (see Figures 6.4d and 6.5d). Another difference is that more details of *K* heterogeneity were observed in Aquitard 3 when more pumping tests were progressively used for model calibration in models *HT_4-NMR_krig* (Figures 6.4c and 6.5c) and *HT_8-NMR_krig* (Figures 6.4d and 6.5d).

Intralayer heterogeneity was derived by models *HT_4-NMR_zonal* and *HT_8-NMR_zonal* after calibrating the internally homogeneous *NMR_zonal* model. For example, the *K* values of Aquitard 2 (see Figures 6.4e-f and 6.5e-f) in the vicinity of PW1 by models *HT_4-NMR_zonal* and *HT_8-NMR_zonal* were approximately one order of magnitude larger than those by the *NMR_zonal* model (see Figures 6.3e-h for its location), potentially enabling weak hydraulic connections between Aquifers 1 and 2. Additionally, the patterns of spatial *K* variability by models *HT_4-NMR_multi* and *HT_8-NMR_multi* (see Figures 6.4g-h and 6.5g-h) were in good correspondence with the patterns by models *HT_4-NMR_zonal* and *HT_8-NMR_zonal*. A noticeable difference was on the *K* estimation within Aquifer 2, where the two integrated models initialized with the *NMR_multi* model revealed more significant vertical *K* variations (Figures 6.4g-h and 6.5g-h), whereas the two models initialized with the *NMR_zonal* model tended to yield more gradual *K* variations radially towards both horizontal and vertical directions (Figures 6.4e-f and 6.5e-f).

Figure 6.6 summarizes the basic statistics of *K* distributions within the data domain for all NMR, HT, and integrated models. Generally, the geometric mean, maximum, and minimum *K* values for the HT-calibrated *K* fields were stable, except for the *HT_4-Homo* model, which yielded a narrower *K* range over

four orders of magnitude. In contrast, the other calibrated *K* fields typically exhibited *K* ranges spanning over six orders of magnitude, which was consistent with the *K* statistics based on permeameter measurements (refer to Alexander et al., 2011). Moreover, variable variances of ln *K* were obtained for different cases. With more spatial heterogeneity patterns being captured, the variance of ln *K* for the *HT_8-Homo* model ($\sigma^2_{\ln K}$ = 5.45) significantly increased compared to the *HT_4-Homo* model ($\sigma^2_{\ln K}$ = 2.25). When the NMR models were calibrated using HT analysis, larger variances of ln *K* were also obtained. It is worth noting that the variances of ln *K* remain relatively stable for the integrated models based on the same initial *K* model, even when using different HT calibration datasets (*HT_4-NMR-krig* and *HT_8-NMR_krig*).

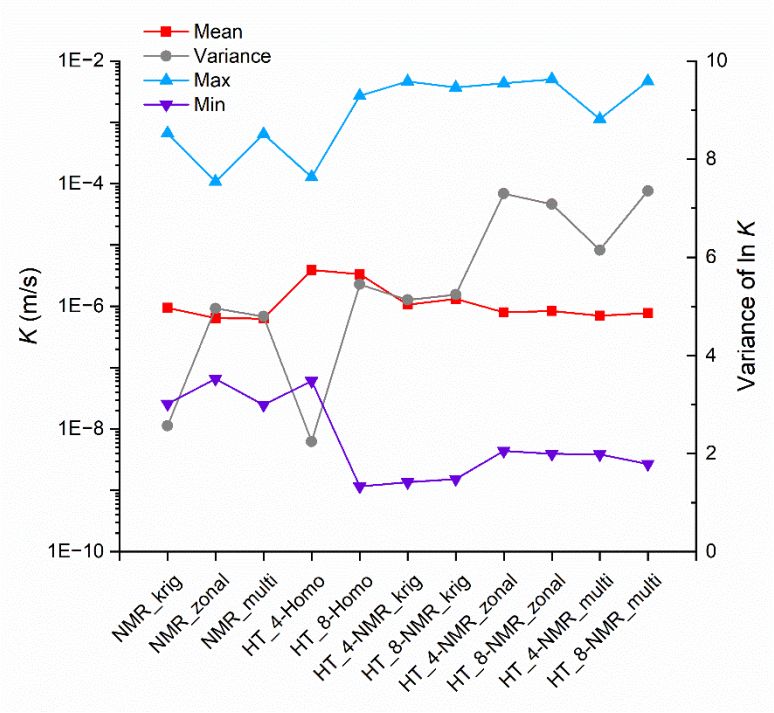


Figure 6.6: Statistics of the spatial *K* fields for all NMR, HT, and integrated models. The geometric mean of *K*, the maximum and the minimum *K*, and the variance of ln *K* within the data domain of the spatial *K* models are provided.

6.5 Evaluation of *K* Fields

6.5.1 Drawdown Simulations of Groundwater Flow Models

Drawdown data from five independent pumping tests not used in the model calibration efforts were incorporated to evaluate the drawdown prediction performance of the groundwater flow models based on different *K* fields. Validation scatterplots between observed and simulated drawdowns of comparative cases are provided in Figure 6.7. Scatter points of drawdown records are represented using various colors and symbols for each pumping test. To quantitatively evaluate the drawdown prediction performance, two performance indicators were employed to calculate the discrepancy between the simulated and observed drawdown responses, including the mean absolute error (L_1) and the mean squared error (L_2).

$$L_1 = \frac{1}{n}\sum_{i=1}^{n}|s_i - s_i^*| \tag{6.2}$$

$$L_2 = \frac{1}{n}\sum_{i=1}^{n}(s_i - s_i^*)^2 \tag{6.3}$$

Figure 6.8 provides a summary of the color-coded L_1 and L_2 for the drawdown fits of all pumping tests, including those used in the model calibration efforts. The last two columns ("Overall" and "Validation") provide the L_1 and L_2 of the drawdown fits for the overall 13 pumping tests and the five independent pumping tests for validation purposes, respectively.

Generally, the observed drawdown responses in the five independent pumping tests were reasonably reproduced in the numerical simulations of all comparative cases. The NMR models yielded more dispersive scatter points, especially for PW5-3. Nevertheless, these scatter points were generally distributed evenly on the two sides of the 1:1 line (see Figures 6.7a-c) between the model simulations and field observations. Among the three NMR models, models *NMR_zonal* and *NMR_multi* yielded somewhat

more reliable drawdown predictions than the *NMR_krig* model. The HT models and the integrated models yielded improved drawdown predictions compared to the NMR models, as more scatter points were aligned closely with the 1:1 line, with fewer large misfits.

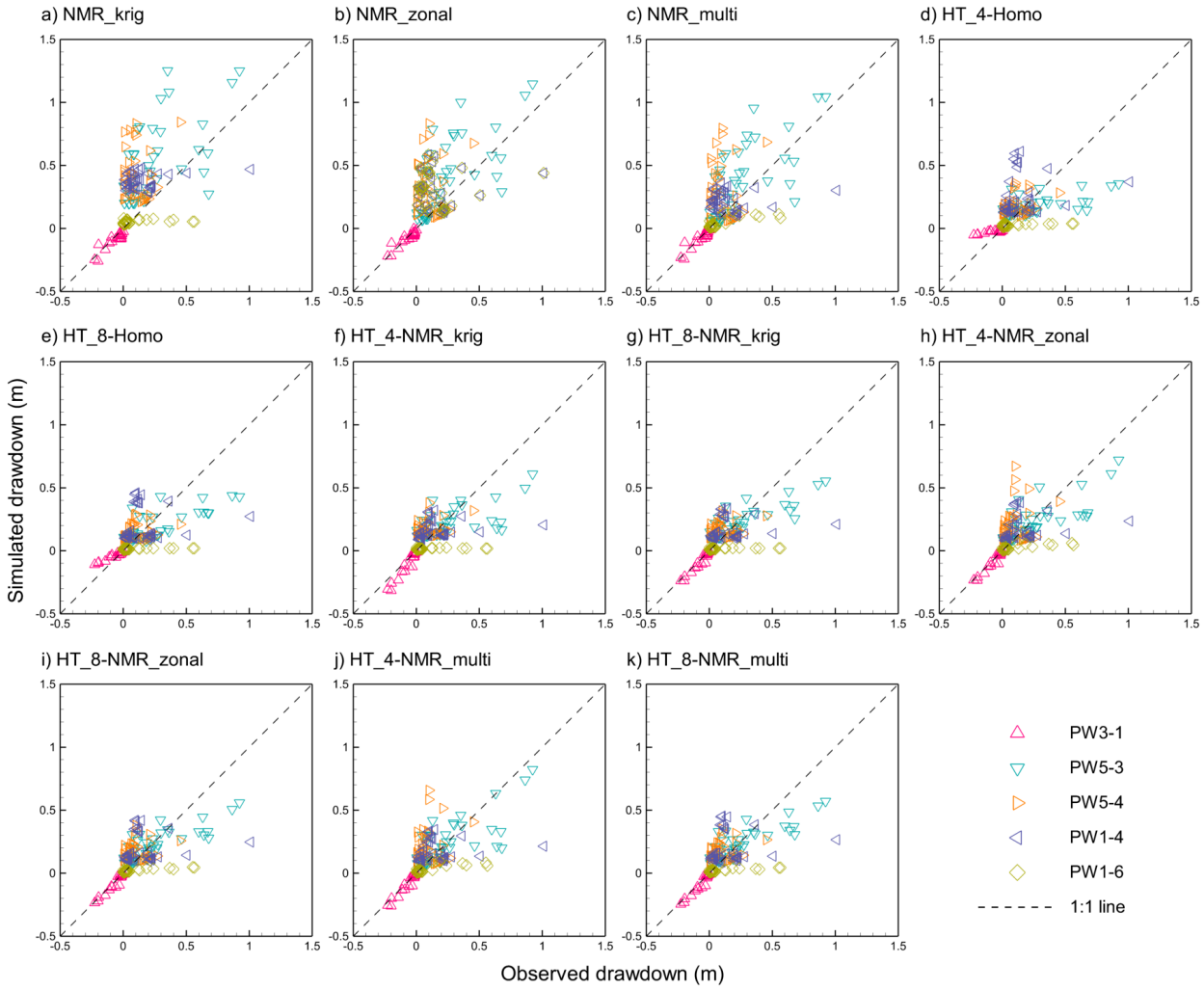


Figure 6.7: Validation scatterplots between simulated and observed drawdowns of different spatial *K* models, including a) *NMR_krig*, b) *NMR_zonal*, c) *NMR_multi*, d) *HT_4-Homo*, e) *HT_8-Homo*, f) *HT_4-NMR_krig*, g) *HT_8-NMR_krig*, h) *HT_4-NMR_zonal*, i) *HT_8-NMR_zonal*, j) *HT_4-NMR_multi*, and k) *HT_8-NMR_multi*. The drawdown responses to five pumping tests are used, including the PW3-1, PW5-3, PW5-4, PW1-4, and PW1-6 tests. A 1:1 line is used to indicate the perfect model fit with observations in each figure part.

Based on the L_1 and L_2 for the validation drawdown fits of five independent pumping tests and the overall drawdown fits of 13 pumping tests, the HT models and the integrated models yielded significantly

more reliable drawdown simulations than the NMR models. In terms of the NMR models, the *NMR_multi* model led the drawdown predictions among these NMR models. In addition, large discrepancies in the drawdown fits occurred for the simulations of some pumping tests (e.g., PW1-3) using the NMR models.

Examination of Figures 6.7d-k revealed that the differences in drawdown fits among the HT models and the integrated models were relatively small. For the HT models, the drawdown prediction performance of the *HT_8-Homo* model was improved compared to the *HT_4-Homo* model, as indicated by the L_1 and L_2 for the validation drawdown comparisons. Furthermore, the drawdown fits were considerably improved for the simulations of pumping tests that were only used in the *HT_8* calibration dataset (PW1-1, PW1-7, PW2-3, and PW4-3). For example, the L_2 for the drawdown fits of the PW1-1, PW2-3, and PW4-3 tests decreased from 0.043, 0.016, and 0.185 by the *HT_4-Homo* model to 0.005, 0.001, and 0.019 by the *HT_8-Homo* model. Based on Figure 6.8, the overall drawdown replications of the four pumping tests involved in both calibration datasets (PW1-3, PW1-5, PW3-3, and PW5-5) were also slightly improved by the *HT_8-Homo* model, compared to the *HT_4-Homo* model.

The integrated models using a limited HT dataset (*HT_4*) for model calibration yielded reliable drawdown predictions, comparable to the predictions of the *HT_8-Homo* model. When more pumping tests were used for model calibration, the drawdown prediction performance of all integrated models further improved, surpassing that of the *HT_8-Homo* model. Moreover, the integrated models using different NMR models yielded comparable overall drawdown predictions, though the drawdown predictions of their initial NMR models (*NMR_krig*, *NMR_zonal*, and *NMR_multi*) varied.

L_1	PW1-3	PW1-5	PW3-3	PW5-5	PW1-1	PW1-7	PW2-3	PW4-3	PW3-1	PW5-3	PW1-4	PW5-4	PW1-6	Overall	Validation
NMR_krig	0.609	0.113	0.096	0.259	0.108	0.052	0.161	0.726	0.033	0.295	0.254	0.308	0.085	0.249	0.220
NMR_zonal	0.407	0.024	0.088	0.078	0.069	0.151	0.086	0.160	0.025	0.194	0.220	0.241	0.069	0.150	0.168
NMR_multi	0.339	0.027	0.093	0.100	0.064	0.151	0.073	0.143	0.026	0.176	0.158	0.237	0.068	0.137	0.149
HT_4-Homo	0.158*	0.025*	0.096*	0.052*	0.132	0.062	0.069	0.328	0.048	0.150	0.174	0.109	0.084	0.116	0.122
HT_8-Homo	0.065*	0.026*	0.078*	0.050*	0.043*	0.070*	0.027*	0.101*	0.033	0.133	0.152	0.098	0.087	0.079	0.109
HT_4-NMR_krig	0.100*	0.025*	0.070*	0.064*	0.078	0.064	0.170	0.146	0.039	0.130	0.118	0.102	0.086	0.093	0.102
HT_8-NMR_krig	0.081*	0.030*	0.063*	0.060*	0.026*	0.053*	0.026*	0.126*	0.019	0.114	0.127	0.106	0.086	0.076	0.099
HT_4-NMR_zonal	0.102*	0.036*	0.068*	0.051*	0.032	0.088	0.085	0.158	0.018	0.109	0.125	0.155	0.080	0.090	0.107
HT_8-NMR_zonal	0.075*	0.030*	0.069*	0.056*	0.025*	0.050*	0.023*	0.132*	0.017	0.108	0.143	0.116	0.082	0.077	0.102
HT_4-NMR_multi	0.108*	0.034*	0.083*	0.067*	0.038	0.115	0.073	0.150	0.022	0.106	0.129	0.154	0.075	0.094	0.107
HT_8-NMR_multi	0.074*	0.029*	0.062*	0.051*	0.025*	0.049*	0.020*	0.118*	0.020	0.106	0.148	0.112	0.083	0.074	0.102

L_2	PW1-3	PW1-5	PW3-3	PW5-5	PW1-1	PW1-7	PW2-3	PW4-3	PW3-1	PW5-3	PW1-4	PW5-4	PW1-6	Overall	Validation
NMR_krig	0.505	0.014	0.018	0.107	0.016	0.007	0.084	0.589	0.001	0.134	0.077	0.139	0.025	0.136	0.087
NMR_zonal	0.264	0.001	0.016	0.013	0.009	0.075	0.023	0.058	0.001	0.072	0.070	0.093	0.020	0.060	0.058
NMR_multi	0.203	0.001	0.019	0.021	0.008	0.074	0.018	0.036	0.001	0.058	0.039	0.091	0.020	0.050	0.048
HT_4-Homo	0.035*	0.001*	0.017*	0.004*	0.043	0.009	0.016	0.185	0.005	0.046	0.055	0.015	0.028	0.035	0.033
HT_8-Homo	0.007*	0.001*	0.010*	0.004*	0.005*	0.011*	0.001*	0.019*	0.002	0.032	0.044	0.012	0.031	0.015	0.026
HT_4-NMR_krig	0.018*	0.001*	0.009*	0.007*	0.014	0.009	0.333	0.039	0.003	0.035	0.031	0.014	0.031	0.037	0.025
HT_8-NMR_krig	0.012*	0.001*	0.007*	0.006*	0.002*	0.005*	0.001*	0.023*	0.001	0.024	0.034	0.014	0.030	0.013	0.022
HT_4-NMR_zonal	0.026*	0.002*	0.008*	0.004*	0.003	0.021	0.073	0.035	0.001	0.024	0.034	0.036	0.026	0.023	0.027
HT_8-NMR_zonal	0.011*	0.001*	0.007*	0.005*	0.001*	0.004*	0.001*	0.025*	0.001	0.024	0.039	0.017	0.028	0.014	0.024
HT_4-NMR_multi	0.020*	0.001*	0.011*	0.008*	0.003	0.040	0.049	0.034	0.001	0.025	0.034	0.037	0.024	0.023	0.027
HT_8-NMR_multi	0.011*	0.001*	0.006*	0.004*	0.001*	0.004*	0.001*	0.019*	0.001	0.022	0.042	0.015	0.028	0.013	0.023

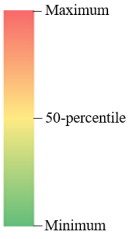


Figure 6.8: Summarized mean absolute error (L_1) and mean squared error (L_2) for the drawdown fits between model simulations and field observations for each pumping test. The "Overall" column represents the drawdown fits for all 13 pumping tests, and the "Validation" column exhibits the drawdown fits for the five pumping tests (i.e., PW3-1, PW5-3, PW1-4, PW5-4, and PW1-6) solely used for model validation. The numbers followed by a superscripted asterisk (*) mean this pumping test was used for model calibration in this case. Color is coded vertically by each pumping test, in which maximum and minimum values are labeled in red and green. The change in color is linear to the change in values.

6.5.2 Comparison with Borehole K Measurements

The *K* profiles extracted from different spatial *K* models were also compared with co-located permeameter measurements at locations of nine observation wells (see Figure 6.9). For clarity, the integrated models calibrated using the *HT_4* calibration dataset were removed from the comparison in Figure 6.9, as the vertical *K* trends of their extracted *K* profiles were generally consistent with those using the *HT_8* calibration dataset. Correspondingly, Figure C7 presents the *K* profiles from the HT and the integrated models using the *HT_4* dataset, along with the permeameter measurements.

First, models *NMR_zonal* and *NMR_multi* identified that the *K* values of Aquifers 1 and 2 were typically on the order of magnitude of 10^{-5} and 10^{-4} m/s, respectively, while the *K* values of these two aquifer layers were estimated to be slightly lower by the *NMR_krig* model. Note that the permeameter

measurements were discontinuous over several depth intervals due to soil compression and loss, as reported by Alexander et al. (2011). Hence, the depth ranges of Aquifers 1 and 2 identified by NMR models were occasionally misaligned with the high-*K* layers determined by permeameter measurements.

The *HT_4-Homo* model yielded smooth downhole *K* profiles. Only Aquifer 1 can be reliably captured at most borehole locations; however, its depth and thickness deviated from those identified by the NMR models. When more pumping tests were used for calibration, the resolution of *K* profiles by the *HT_8-Homo* model became significantly improved. Compared to the *HT_4-Homo* model, Aquifer 1 mapped by the *HT_8-Homo* model had a reduced layer thickness in PW1, PW3, PW4-3, PW5, and CMT1, and its estimated *K* in PW2-3 was increased. The *HT_8-Homo* model and the NMR models showed good correspondence for the *K* magnitude and the depth range of Aquifer 1 at most borehole locations. In addition, Aquifer 2 was also captured in PW1, PW3, PW4-3, CMT1, and CMT3. However, high *K* values were found in the bottom three meters of the data domain by the *HT_8-Homo* model, compared to permeameter measurements.

The three integrated models resembled their initial NMR models in the vertical *K* trends, with more contrasting *K* variations, especially in the lower half of these *K* profiles. In terms of the *K* comparison among the integrated models, the upper half of the *K* profiles (i.e., Aquifer 1 and above) by these integrated models were very consistent. For the mapping of Aquifer 2, models *HT_8-NMR_krig* and *HT_8-NMR_multi* exhibited more significant vertical *K* variations than the *HT_8-NMR_zonal* model (e.g., in CMT3). Compared to the *HT_8-Homo* model, these integrated models yielded approximately one order of magnitude higher *K* values of Aquifer 2 (e.g., in the vicinity of PW1, PW3, CMT1, and CMT3). Additionally, Aquifer 2 did not extend to PW2-3 and CMT2 based on the NMR logs (see Figure 6.2), and the integrated models also recognized these regions as low-*K* materials (see Figure 6.9). Moreover, the integrated models using the *HT_4* dataset for calibration yielded *K* profiles (see Figure C7) with *K* trends

similar to and confined between the corresponding NMR models and the integrated model calibrated using the *HT_8* dataset.

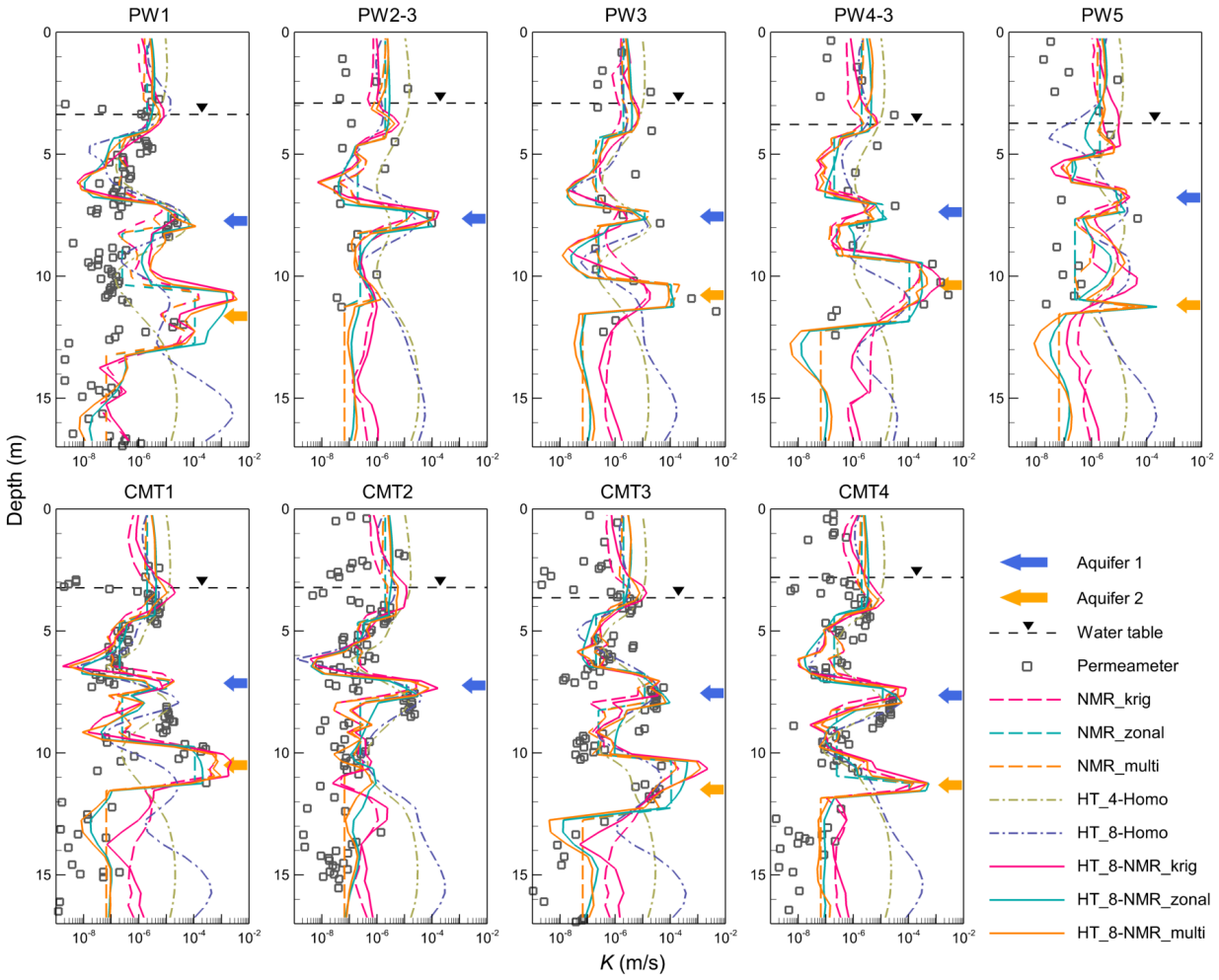


Figure 6.9: Comparison of *K* profiles at nine well locations between the model-derived *K* and the permeameter *K* measurements. The NMR, the HT, and the integrated models using the full calibration dataset are included in the comparison. Aquifers 1 and 2 are denoted using colored arrows.

6.6 Discussion

6.6.1 Drawdown Prediction Capabilities of Groundwater Flow Models

Based on comparisons with field observations, the NMR *K* models can yield reasonable overall drawdown predictions for the pumping tests conducted in this complex groundwater flow system.

Compared to the drawdown predictions (refer to Sun et al., 2024) based on *K* fields from other borehole hydraulic testing methods (e.g., slug tests, permeameter tests, and hydraulic profiling tool survey), the improved drawdown predictions of the groundwater flow models based on borehole NMR *K* estimates can be attributed to its ability to characterize subsurface heterogeneity at high resolution for soils ranging from gravel, sand, silt to clay. However, these NMR *K* models based on the interpolation of borehole *K* data cannot accurately represent the hydraulic connections between boreholes. For example, the dispersive distribution of scatter points for the PW5-3 test (Figures 6.7a-c) indicated that the complex patterns of hydraulic connectivity in different flow paths, as observed in multiple ports, were not faithfully described. Particularly, the connected network of high *K* was not adequately reproduced using the NMR models. For example, the drawdowns for the PW5-4 test were overestimated.

The insufficient mapping of spatial *K* variability can be somewhat anticipated, as the inter-well heterogeneity was generated based solely on existing borehole information, thereby depending on the density of hard data. Moreover, limits existed in the borehole logging techniques. For example, the direct-push NMR logging can only explore to a depth of ~12 mbgs at the NCRS, and measurement noise occurred in the PW1 and PW5 logs, as bimodal distributions of T_2 signals were recorded by Wang et al. (2025b).

On the other hand, HT analysis is a robust approach that makes reliable drawdown predictions. In this study, the HT models yielded substantially improved drawdown predictions compared to the NMR models, even though a smooth *K* tomogram was obtained by the *HT_4-Homo* model. It advocated the use of HT analysis to enhance the understanding of hydraulic connections for *K* estimation. Furthermore, integrating NMR logs with HT analysis can further lead to refined drawdown predictions. The integrated models depicted layered characteristics of this aquifer system at a higher resolution compared to the HT models, especially for regions with fewer observation data (e.g., the lower half of the aquifer system). For example, the limited HT dataset (i.e., *HT_4*) includes only seven steady-state drawdown records from the

"-6" and "-7" observation intervals of the four CMT wells and PW1, which were situated near the bottom boundary of the simulation domain. The full HT dataset (i.e., *HT_8*) had 16 more drawdown records available from the "-6" and "-7" intervals compared to the limited dataset. Table C1 summarizes the drawdown records from the "-6" and "-7" observation intervals. Additionally, only the PW1-7 test stressed the aquifer system at a depth of 15.56 mbgs, resulting in drawdown responses near the bottom boundary, while the other pumping tests involved in the calibration can only stress the aquifer at a location as deep as Aquifer 2 (e.g., the stress depth of the PW4-3 test was 11.55 mbgs). Therefore, the lower half of the aquifer system, particularly the bottom three meters of the simulation domain, was not adequately calibrated solely using the HT analysis due to the insufficient drawdown data. On the other hand, the NMR models provided initial guesses of low K values for the region below Aquifer 2 (i.e., Aquitard 3), facilitating the model calibration in the geostatistical inverse modeling. Hence, the drawdown prediction capability of the HT model using the full calibration dataset can still be improved by integrating it with NMR logging data.

When the observation data were sparse (the *HT_4* dataset), various integrated models still produced reliable drawdown predictions. Based on the L_1 and L_2 of the overall and the validation drawdown fits, these integrated models exhibited similar drawdown prediction capabilities when different spatial interpolation approaches were employed for initial K distributions. One noticeable difference was that large drawdown misfits occurred for the simulation of the PW2-3 test by the *HT_4-NMR_krig* model when the calibration data were sparse. Nevertheless, the integrated models using a limited HT calibration dataset yielded comparable drawdown predictions to those by the HT models using a denser calibration dataset (the *HT_8* dataset).

6.6.2 Implications for Characterization of Spatial Heterogeneity

To directly illustrate the spatial connections of high K materials, the isosurfaces at specified elemental K values of 1×10^{-5} m/s and 1×10^{-4} m/s are plotted using different colors in Figure 6.10,

indicating the locations and dimensions of Aquifers 1 and 2 identified by each spatial *K* model within the data domain. When more pumping data were incorporated (see Figures 6.10a and b), two aquifer layers, especially Aquifer 2, can be more readily characterized using the HT models. Compared with *HT_4-Homo*, the improvements in drawdown predictions and aquifer characterization in *HT_8-Homo* again demonstrated that the ability of HT analysis to interpret the *K* field is highly dependent on data density.

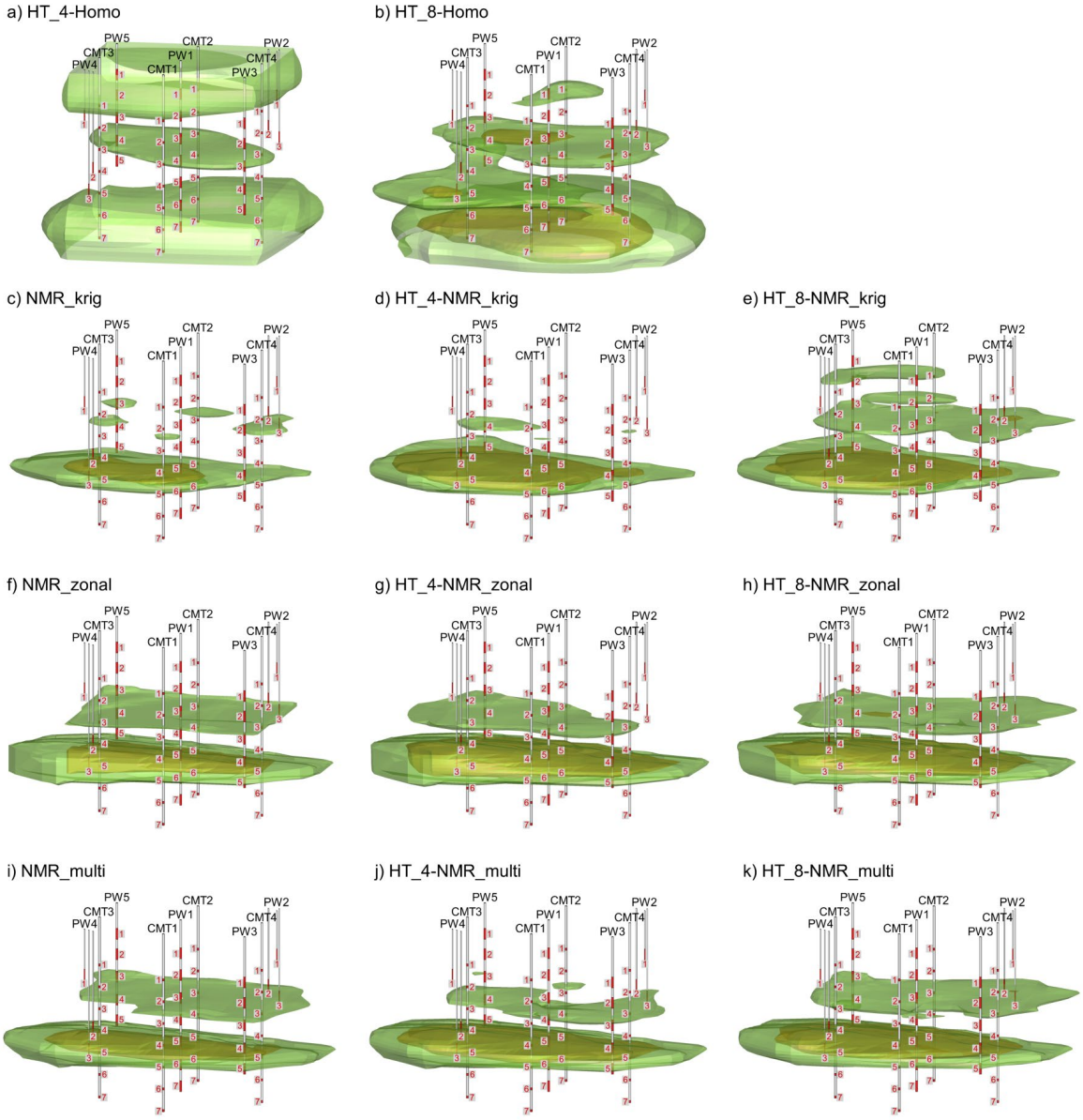


Figure 6.10: Locations and dimensions of Aquifers 1 and 2 identified by different spatial *K* models. The aquifers are illustrated using isosurfaces at two specified *K* values, 1×10^{-5} m/s and 1×10^{-4} m/s, corresponding to the green and yellow isosurfaces.

As suggested by Yeh and Liu (2000), an optimal monitoring network requires that the horizontal spacing between two observation wells be half of the horizontal correlation length and that the vertical spacing between two observation intervals be at least half of the vertical correlation length. However, such an optimal monitoring network is not very practical for complex groundwater environments (e.g., the highly heterogeneous glaciofluvial deposit at the NCRS). For example, the estimated horizontal and vertical correlation lengths of the NMR-derived *K* were 19.84 and 2.48 m, respectively. Corresponding to these correlation lengths, more than 100 observation intervals are required to develop an optimal monitoring network within the data domain, which is impractical to implement in the field work.

Comparatively, the full HT calibration dataset in this study had an average of 30 observation points for eight pumping tests. The *HT_8-Homo* model yielded reliable drawdown predictions; however, its *K* representation, particularly in mapping high-*K* zones (Figure 6.10b), remained biased compared to *K* presentations that utilized NMR logging to locate aquifers. Based on the vertical *K* replication in Figure 6.9 and the aquifer identification in Figure 6.10, Aquifer 1 was accurately characterized by the *HT_8-Homo* model. Still, the shape and *K* magnitude of Aquifer 2 by the *HT_8-Homo* model were not consistent with those interpreted by the NMR and the integrated models. It can be anticipated that the resolution of *K* tomograms is further increased when more pumping tests and a denser monitoring network are involved for a refined HT analysis, especially for the lower half of the simulation domain.

A limited HT calibration dataset was used in this study, with an average of 20 observation points for four pumping tests. Such a low-density HT survey was not sufficient for mapping this double-layer aquifer system (see Figure 6.10a). After incorporating the NMR models as initial *K* distributions, two aquifer layers can be better reproduced with appropriate dimensions and *K* magnitudes by the integrated models using the limited calibration dataset.

This study also compared three spatial interpolation approaches to derive initial *K* models for HT analysis. Two aquifer layers delineated by multiple NMR models showed different horizontal *K* trends. For example, the *NMR_krig* model delineated discontinuous high-*K* lenses for Aquifer 1; in contrast, models *NMR_zonal* and *NMR_multi* yielded continuous horizontal layering of Aquifer 1 with varying geometries. After calibration using HT analysis, the integrated models optimized the layer geometry and dimensions, yielding more consistent downhole *K* estimates and comparable drawdown prediction performances among them. Based on the comparison between the spatially correlated NMR *K* models and the integrated models optimized using HT analysis, reliable spatial *K* models based on borehole testing data can reasonably reproduce the patterns of spatial *K* variability. Such models entail high-resolution borehole *K* estimation (e.g., from NMR logging) and spatial correlation suitable for the geological environment of the test site (e.g., using the multi-level heterogeneity model). However, accurate imaging of the spatial connectivity of high-*K* materials for complex aquifer systems can only be achieved by the methods that directly yield 3D heterogeneous *K* estimates (e.g., HT).

Other than the large drawdown misfit that occurred in one pumping test of the *HT_4-NMR_krig* model result, differences among the three integrated models existed in the spatial *K* variability that the initial models *NMR_krig* and *NMR_multi* resulted in vertically more contrasting *K* variations within Aquifer 2 after calibration, compared to the integrated models initialized with the *NMR_zonal* model. Such “preferential flow paths” probably better reproduce the glaciofluvial sedimentation at the NCRS. Based on a comprehensive consideration of drawdown prediction capabilities and the mapping of spatial *K* variability, the *NMR_multi* model was perhaps the most optimal initial model for HT estimation in this layered porous medium; however, either spatial interpolation approach was applicable for constructing initial *K* distributions for geostatistical inverse modeling of HT analysis.

7 Three-Dimensional Analysis of Heat Tracer Transport with Hydraulic Conductivity Distributions from Various Approaches in a Complex Aquifer System

Citation:

Wang C., Ning Z., Luo N., Zhao, Z., Berg S. J., & Illman, W. A. (2025). Three-Dimensional Analysis of Heat Tracer Transport with Hydraulic Conductivity Distributions from Various Approaches in a Complex Aquifer System. Submitted to *Water Resources Research* (Under review).

7.1 Overview

Accurate predictions of tracer transport in complex aquifer systems remain a formidable challenge due to the spatial heterogeneity of hydraulic conductivity (K). Here, the feasibility of predicting heat transport within highly heterogeneous glaciofluvial deposits was explored based on the classical advection-dispersion equation with K distributions from various characterization methods. In this experiment, a heat tracer was used as an alternative to the conservative tracer, in which groundwater was chilled and reinjected into the subsurface under a divergent flow condition. Temperature signals were monitored at depth-specific observation ports, revealing highly non-uniform migration and spreading of the heat tracer plume. Comparative K representations were used for transport simulations, including homogeneous, kriged, and zonal K fields, derived from direct-push K logging, permeameter measurements, and pumping test data, as well as hydraulic tomography (HT) results. HT has been demonstrated to be robust for characterizing subsurface heterogeneity. However, the effectiveness of using K tomograms for predicting heat transport behavior has yet to be rigorously investigated in the field. Moreover, the heterogeneous porosity field was represented using effective porosity estimates from

nuclear magnetic resonance (NMR) logging. Simulated and observed temperature breakthrough curves were quantitatively compared using temporal moment analysis. Results showed that the simulation using the HT *K* field can best reproduce the spatiotemporal evolution of the heat plume. Furthermore, the impacts of transport and thermal parameters were examined, including porosity, dispersivities, and thermal conductivity. This study suggests that the advection-dispersion equation using HT-based *K* can adequately predict the transport behavior in a complex aquifer system.

7.2 Introduction

The accurate prediction of tracer transport behavior in porous media has consistently been a critical issue over the last half-century. Traditionally, the classical advection-dispersion equation (ADE) has been applied to conceptualize transport models, with a homogeneous hydraulic conductivity (*K*) assigned to the entire aquifer. Based on the classical theory, mechanical dispersion led to the temporal spreading and evolution of the tracer plume. Thus, dispersivity was employed to quantify the hydraulic mixing process, which occurs either longitudinally or transversely in the horizontal and vertical directions to the migration paths.

An early-time landmark field experiment for investigating subsurface contaminant transport was conducted at the Canadian Forces Base (CFB) Borden site, where a natural-gradient tracer test was performed (see Mackay et al., 1986). Sudicky et al. (1983) found that macrodispersivity increases with more heterogeneities encountered in the migration path. Therefore, a unique dispersivity value cannot be assigned to the medium to describe the evolution of the dispersed zone of the tracer. Moreover, the importance of subsurface heterogeneity was also acknowledged, which led to irregular concentration profiles and an enhancement of longitudinal dispersion. Sudicky (1986) investigated the influence of *K* variability on macrodispersivity and found that the observed tracer spreading increased with time.

The sandy aquifer at the Borden site exhibits only mild heterogeneity, yielding Gaussian forms of tracer patterns after its spatiotemporal evolution. Similar studies have been performed at Cape Cod, Massachusetts (Garabedian et al., 1991; Leblanc et al., 1991) and the North Bay landfill site (Sudicky et al., 2010). For more complex aquifer systems with highly heterogeneous *K* fields, accurately representing the spatial heterogeneity of *K* is essential for transport simulations.

Since the 2000s, a series of tracer tests (see review by Zheng et al., 2011) have been conducted at the Macrodispersion Experiment (MADE) site in Mississippi, USA. Different from the previous transport studies, the highly heterogeneous nature of this porous medium led to non-Fickian behavior of the tracer plume migration. To explain the non-Fickian behavior, a rate-limited mass transfer process between the mobile and the immobile domain of the medium is introduced to solute transport (Harvey & Gorelick, 2000), forming a dual-domain mass transfer (DDMT) model. Initially, Feehley et al. (2000) found that using the DDMT model can better reproduce the observed tracer plume than the ADE, particularly in terms of the longitudinal spreading of the tracer plume. In a subsequent study by Liu et al. (2010), the use of the DDMT model was found to adequately reproduce the primary characteristics of the breakthrough curves (BTCs) when geological features were captured. However, the DDMT model was employed as an alternative conceptualization approach for predicting transport behavior, due to the insufficient characterization of *K* heterogeneity. Liu et al. (2007b) demonstrated that the mass transfer process is governed by both diffusion and slow advection between the mobile and immobile domains. Zheng et al. (2011) stated that the actual hydrodynamic mixing caused by subsurface heterogeneity is accounted for to a large extent by mass transfer between the mobile and immobile domains.

Researchers advocated that transport simulations should be based solely on advection and molecular diffusion, provided that subsurface heterogeneity can be precisely characterized. In the MADE studies, Salamon et al. (2007) incorporated the hole effect structure into stochastic geostatistical simulations and found that the longitudinal mass distribution profiles can be reproduced using the

classical ADE with a subset of *K* realizations. Ronayne et al. (2010) represented submeter-scale, highly connected channels within a background matrix based on a multivariate Gaussian *K* field. They found that the ADE can reasonably reproduce key characteristics of the observed BTC. Bianchi et al. (2011b) confirmed that a high degree of flow and transport connectivity exists due to preferential flow paths, a concept initially suggested by Feehley et al. (2000) and Harvey and Gorelick (2000). Dogan et al. (2014) produced 3D *K* fields based on high-resolution direct-push *K* data and found that the ADE can adequately reproduce the solute transport behavior. Bianchi and Zheng (2016) developed 3D *K* fields based on the simulation of lithofacies and yielded reasonable concentration fits with observations, concluding that the transport at the MADE site is controlled by a network of high *K* sediments (preferential flow paths) embedded in a less permeable matrix. In summary, it is possible to accurately predict transport behavior in complex geological environments when hydraulic connectivity patterns are adequately described.

HT has been demonstrated to be a robust approach for characterizing subsurface heterogeneity. Most studies (e.g., Berg & Illman, 2011b; Cardiff et al., 2013; Hochstetler et al., 2016; Tiedeman & Barrash, 2020) worked on predicting drawdown responses to independent pumping tests, while limited studies have focused on predicting tracer transport behavior with HT-derived *K* distributions. Ni et al. (2009) performed a 2D HT analysis using the geostatistical inverse approach in a synthetic mildly-heterogeneous aquifer and found that the HT model can yield the most reliable predictions of BTC peaks and phases. Illman et al. (2012) designed a sandbox experiment of forced-gradient tracer tests within a mildly heterogeneous sandy aquifer. The temporal moment analysis between simulated and observed BTCs demonstrated that utilizing HT analysis for mapping heterogeneous *K* can facilitate high-resolution predictions of solute transport. Zhao et al. (2022) extended the work of Illman et al. (2012) and concluded that the tracer transport predictions using HT-estimated *K* can be further refined when reliable borehole lithofacies information is incorporated. Wang and Illman (2023) integrated geophysical logging data with HT analysis for improved *K* representations and adequately reproduced the main characteristics of

observed BTCs for two tracer tests within a highly heterogeneous numerical sandbox aquifer system. All these synthetic or laboratory studies employed the classical ADE for simulations.

Only a few field tracer experiments have been conducted to evaluate the HT-estimated *K*. Jimenez et al. (2015) applied a sequential procedure of HT inversion to image the 3D *K* distribution for a mildly heterogeneous porous medium at the Stegemühle Site located south of the city of Göttingen, Lower Saxony, Germany. They found that the main characteristics of the BTCs can be recovered. However, the tracer BTCs can also be approximated by a simplified effective *K* model with higher dispersivity. Sanchez-León et al. (2016) conducted two forced-gradient tracer tests in a more heterogeneous alluvial aquifer at the Hydrogeological Research Site Lauswiesen in the Neckar valley close to Tübingen, Germany. The pilot-point inversion method was used to yield a HT *K* model. Additionally, one tracer dataset was used to calibrate the parameters for the DDMT model (e.g., mobile-domain porosity). Results revealed that using the HT-estimated *K* can successfully predict the BTC for another independent tracer test. Nevertheless, the mobile-domain porosity after calibration (0.0195 for sandy materials) is unreasonably low. Therefore, a tracer test has yet to be conducted in a highly heterogeneous geological deposit, in which HT-estimated *K* is used for numerical predictions of transport behaviors, and the site-specific calibration to tracer test data is not performed for optimized model fit with observations.

In this study, a heat tracer test was designed with a dense monitoring system for recording temperature BTCs at high resolution. The heat tracer was used as a proxy (Anderson, 2005) for conservative tracer migration, which has been extensively used to study the advective transport process (e.g., Doro et al., 2015; Giambastiani et al., 2013; Klepikova et al., 2016; Sarris et al., 2018; Schwede et al., 2014; Somogyvari & Bayer, 2017; Tarnawski et al., 2015; Vandenbohede et al., 2009; Wagner et al., 2014). Ma et al. (2012) found that the temperature depth profiles and temperature peak arrivals simulated by the heat transport model were in reasonable agreement with observations when the *K* distribution was obtained from the calibration of a bromide transport model. However, a temperature-related density effect

may exist in heat tracer tests, causing upward flow when heated water moves through the porous medium. Therefore, to avoid this temperature-related density effect, a heat tracer test was performed by chilling groundwater and reinjecting it into the subsurface to establish divergent flow through multiple advective paths. Based on our test data, the density difference between the chilled and the background groundwater is small, leading to significantly reduced density-driven vertical flow. The spatially variable temperature data were monitored at 24 observation ports through a highly heterogeneous glaciofluvial deposit at the North Campus Research Site of the University of Waterloo.

Innovative aspects of this study include: 1) a heat tracer test is performed by injecting chilled groundwater, and a dense monitoring network is used to track plume migration and spreading; 2) nuclear magnetic resonance (NMR) logging allows for the estimation of a heterogeneous effective porosity field, which was previously hard to quantify using other in situ testing methods; and 3) the use of HT analysis provide reliable K estimates for reasonable predictions of the spatiotemporal evolution of the heat tracer plume based on the classical ADE.

The workflow of this study can be divided into four parts. First, this new heat tracer test is described in detail, including its experimental design, sensor configuration, and data collection procedures. Second, comparative cases of K representations and the NMR-derived effective porosity field are constructed for numerical simulations of heat transport. Third, the simulated temperature BTCs are compared with field observations. The temporal moment analysis is employed to quantify the misfits. Additionally, the sensitivity of transport and thermal parameters to heat tracer transport is assessed. These parameters include porosity, dispersivities, and thermal conductivity. Our primary goal is to reproduce the main characteristics of the observed BTCs within the 3D monitoring network, assuming advection and local dispersion to be the dominant transport processes. Meanwhile, the heterogeneous K and porosity distributions are derived from the analyses of field measurements without parameter calibration (e.g., porosity) using tracer data for optimized temperature fits.

7.3 Experimental Design

7.3.1 The North Campus Research Site

The North Campus Research Site (NCRS) is located on the University of Waterloo campus in Waterloo, Ontario, Canada. The heat tracer test was performed in the shallow subsurface of the NCRS, which consists of glaciofluvial deposits and till ranging from gravel, sand, silt, to clay. A dense and stiff Catfish Creek Till is regarded as the hydraulic barrier at the bottom of the research zone of interest, which is located at approximately 16 meters below ground surface (mbgs). A sandy-to-clayey silt sequence, the Maryhill Till, overlies the Catfish Creek Till. Above the Maryhill Till are the thin erosional remnants of the Tavistock Till, which is situated below the contemporary organic soil. Borehole geology logs (Alexander et al., 2011; Berg & Illman, 2011b; Zhao & Illman, 2017, 2018) revealed alternating layers of aquifer and aquitard units. The main "aquifer zone" is located from approximately 8 – 13 mbgs, in which low *K* materials separate two discontinuous aquifer layers. The upper aquifer ("Aquifer 1") is composed of sand to sandy silt, and the lower aquifer ("Aquifer 2") is composed of sandy gravel. Based on previous pumping tests, these aquifer layers have been found to behave as confined or semi-confined aquifers (Berg & Illman, 2011b).

The NCRS is instrumented with nine wells in a 15 m × 15 m square, as shown in Figure 7.1. It includes three 4-inch multi-screened wells (PW1, PW3, and PW5), two well clusters (PW2s and PW4s), and four Continuous Multichannel Tubing (CMT) wells (CMT 1 through 4). Each well consists of multiple screens completed at different depths, with the larger number following their IDs indicating increased depths from the ground surface (e.g., PW1-1 through PW1-7). The five pumping wells (PW1 through PW5) can be utilized to pump water and monitor responses to pumping and tracer tests (e.g., drawdown and temperature), whereas the four CMT wells are specifically designed to monitor hydraulic head and temperature changes as well as to collect water samples.

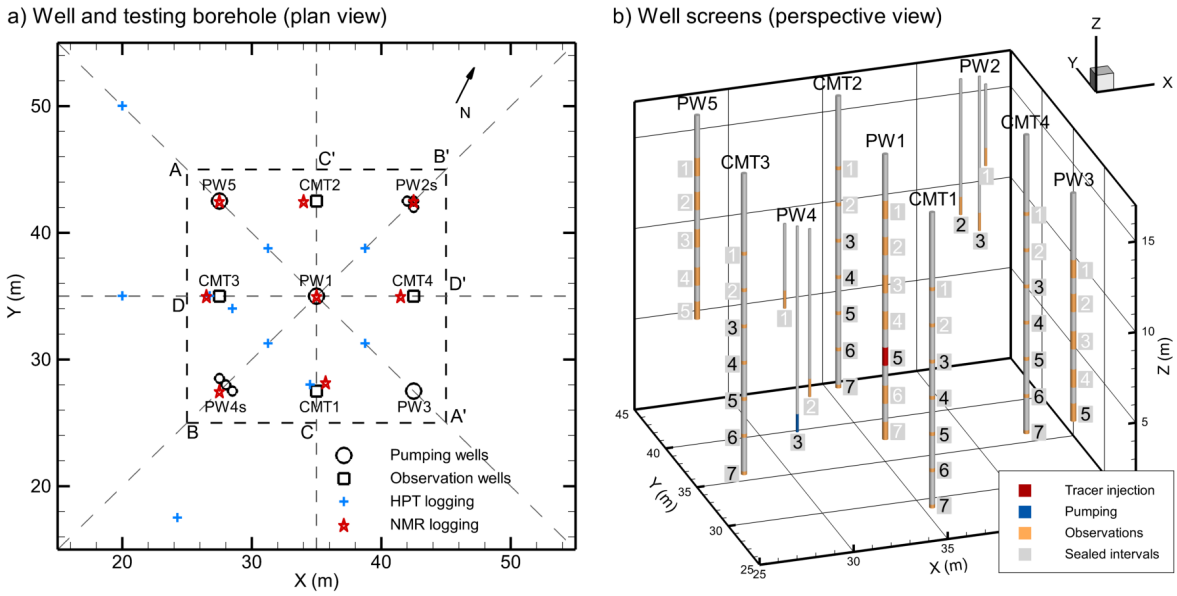


Figure 7.1: Locations of monitoring wells and testing boreholes. a) Plan view of the locations of the monitoring wells, HPT logging, and NMR logging. The dashed-line square (A-B'-A'-B) delineates the area of the data domain (20 m × 20 m) for subsequent visualization of *K* and porosity distributions, which encompasses the wells (within a 15 m × 15 m extent). The coordinates are consistent with the simulation domain of the numerical model of heat transport. b) Perspective view of the well screens, in which the screens for tracer injection, pumping, observation, and sealed intervals are highlighted in different colors. Screens numbered in black were used for observation of head and temperature changes during the tracer test.

7.3.2 The Heat Tracer Test

Since the shallow subsurface at the NCRS has been contaminated by infiltrated road salt applied during winter, it is difficult to conduct a solute transport experiment with salt as a tracer, given the high background concentration of chloride. Based on the groundwater sampling data by Williamson (2016), the chloride concentration ranged from 88.8 mg/L to 1556.7 mg/L in the upper four channels of CMT2 and CMT3 during the sampling event in April 2016.

As a result, heat was used as a tracer instead of a conservative salt tracer. For an indicator of the advective transport process, a common issue with the heat tracer is that the heated water may induce a significant density effect, resulting in enhanced vertical movement of the heat tracer plume. Here, a cost-effective design was considered to mitigate the density effect, which utilized the temperature difference

between the groundwater and air in early winter in Southern Ontario. Here, the groundwater was pumped from the subsurface and stored in surface reservoirs. The pumped water was naturally chilled in the air until the desired temperature difference from the background groundwater was reached. Afterwards, the chilled water was injected into the targeted screen (PW1-5). The daily-average temperature at the NCRS during the tracer injection period was from -5.4 to 1.2 °C (data sources: The UW Weather Station), while the groundwater temperature for the targeted aquifer was ~10 °C. For example, assuming that a 9 °C temperature difference can be established by chilling the groundwater at 10 °C, the water density can vary at most 0.027%. In contrast, achieving the same temperature difference by heating the groundwater at 20 °C (a more common design) leads to a density difference that is more than a factor of 8 larger than the design of this tracer test. Therefore, the density effect can be effectively reduced by chilling the groundwater instead of raising the temperature at the NCRS. In addition, to prevent the intraborehole vertical mixing of chilled water, the tracer injection was limited within a 1-m screen using straddle packers for isolation.

The heat tracer test was conducted from November 28, 2024, to January 16, 2025. The chilled water was injected at PW1-5 (refer to Figure 7.1). To continuously obtain a groundwater supply, PW4-3 was pumped, and the pumped groundwater was stored in five water reservoirs to chill the groundwater. Each water reservoir has a capacity of 8.43 m^3 (4.5 m length × 2.23 m width × 0.84 m height). Afterwards, the chilled groundwater was pumped from the water reservoirs and injected into PW1-5. PW4-3 is a highly productive well, with a maximum pumping rate exceeding 30 L/min. The injection and pumping rates were controlled to be as consistent as possible, with the rate being set close to the maximum injection rate at PW1-5.

The groundwater temperatures in the monitoring network were recorded for 48 hours prior to the tracer test, allowing for a stable background temperature distribution to be obtained. The heat tracer test can be divided into three stages. The first stage, known as “Forced-gradient flow”, involved establishing

steady-state flow conditions by pumping groundwater at PW4-3 and directly injecting the groundwater via a single water hose at PW1-5. The average flow rate in Stage 1 was determined to be 8.01 L/min. Then, the tracer injection began during Stage 2 ("Heat injection"), with chilled water at 1 °C being injected into PW1-5. At the same time, groundwater was pumped at PW4-3 and stored in other reservoirs. In total, the tracer injection during Stage 2 lasted over a period of 13 days, with the average injection and pumping rates being 7.43 and 7.72 L/min, respectively. After observing clear temperature decreases at the monitoring points, the heat tracer test proceeded to Stage 3 ("Heat dissipation"), in which the pumped water from PW4-3 was directly injected into PW1-5 using a single water hose, as in Stage 1. Stage 3 lasted for ~35 days, and the average pumping/injection flow rate was 8.19 L/min. Table 7.1 summarizes the basic information for each stage of the heat tracer test. During the experiment, pumping and injection were occasionally interrupted for brief periods due to power outages. Detailed records of flow rates, including interruptions during the heat tracer test, are provided in Table D1.

Table 7.1: Three stages of the heat tracer test.

Stages	Start date/time	Durations (h)	Injection rates (L/min)	Temperature of injected water (°C)	Pumping rates (L/min)
Forced-gradient flow	2024-11-28 15:00	15	8.01	10.19	8.01
Heat injection[(1)]	2024-11-29 9:00	216	7.43	1.11	7.72
Heat injection[(2)]	2024-12-09 9:00	96	7.43	3.19	7.72
Heat dissipation	2024-12-12 10:00	837	8.19	10.27	8.19

[(1)] and [(2)] indicate two periods of the heat injection with different temperatures of chilled water for injection in PW1-5.

A surface-lift jet pump (Model 506518SS, Burcam) was used to pump water at PW4-3. To accurately adjust the pumping rate, a variable frequency drive (Model EM15-SP, ATO Inc.) was connected in series with the surface lift pump, allowing for a fine adjustment of up to 1/500 of the pumping rate. Two submersible pumps (Model 415, Solinst Canada Ltd., and MP1, Grundfos Canada Inc.) were used in turn to pump water from the chilled water reservoirs and to inject the chilled water into

PW1-5. Dials were equipped for linear adjustments of the flow rate. During Stages 1 and 3, PW1-5 and PW4-3 were connected using a water hose, and the jet pump drove the water circuit. During Stage 2, the jet pump extracted groundwater from PW4-3, while the submersible pumps injected the chilled water into PW1-5.

Temperature data were collected at 24 monitoring points within the data domain (refer to the dashed square in Figure 7.1). To record the natural changes of groundwater temperature during the test period, the temperature of the ambient groundwater was recorded at a well (NC17) located beyond the data domain, which was 18.8 m away from the nearest observation well, PW4-3. Two types of temperature sensors were used. To fit in the narrow channels (~1 cm of inner diameter) of the CMT wells, the lower five screens of the four CMT wells (CMT1, CMT2, CMT3, and CMT4) were monitored using a micro platinum resistance temperature detector (RTD) (HC-28, Beijing Heng Rui Chang Tai Technology Co., Ltd.). HC-28 can reliably detect temperature within the range of 0-35 °C, and its accuracy is ± 0.2% FS. In total, 20 units of HC-28 were installed in CMT wells. Furthermore, HC-28 was also installed in five reservoirs to monitor the temperature of the chilled water. The pumping wells (2” or 4” screened wells) were monitored using a temperature-pressure dual sensor (Model 3001 Levellogger, Solinst Canada Ltd.) with a built-in RTD. These Levelloggers have a measurement range of 0-50 °C. The accuracy of the Levelloggers is ± 0.05 °C. Levelloggers were installed in PW1-5, PW2-2, PW2-3, PW3-5, PW4-3, and NC-17.

Some shallow observation points, including PW2-1, PW4-1, and the upper two intervals of the CMT wells, were not equipped with temperature sensors because they were considered to be affected by air temperature, as determined through preliminary temperature observations without tracer injection, where the groundwater temperatures of these intervals were variable while the temperatures observed at deeper ports were stable. In addition, PW4-2 was found to be silted, so this observation port was also

removed from the dataset. For the multi-screen wells, a packer was used to install a Levellogger in PW3-5. Moreover, a blank liner was used to temporarily seal PW5, preventing its intraborehole vertical flow.

Two turbine flowmeters (DN25, Anhui Jujie Automation Technology Co., Ltd) were installed on the water hose to monitor pumping and injection rates. Additionally, the pumping rate was manually measured using a 2-Litre graduated cylinder, and the injection rate was checked with a paddlewheel flowmeter (F2000 RT, Blue-White Industries, Ltd.), which also provided in-situ temperature readings.

The temperature data in CMT wells were automatically acquired at an interval of 5 seconds, using a data acquisition system (cDAQ-9174 with NI-9205 modules, National Instruments Corp.). The temperature and pressure data from PW1-5, PW2-2, PW2-3, PW3-5, and PW4-3 were recorded every 5 minutes, which could only be read after retrieving the Levelloggers from the monitoring intervals.

The temperature data of all observation points were processed by averaging the temperature records over every hour. Figure 7.2 shows the observed temperature curves over time for all observation points, in which the zero-time point corresponds to the beginning of tracer injection (Stage 2). The ambient groundwater temperature distribution observed at different locations generally exhibited a downward temperature decrease along each borehole. Due to air temperature variations (detailed records can be acquired from Eric D. Soulis Memorial Weather Station), the temperature of chilled water in PW1-5 increased over the last several days for tracer injection. Thus, during Stage 2, the temperature of chilled water remained at 1.11 °C over the first 216 hours and then increased to 3.19 °C over the last 96 hours of the tracer injection. Table 7.1 summarizes the temperature in PW1-5 during each stage.

For comparison with numerical simulations, points were manually selected to represent the main characteristics of the temperature curves. Twenty points were assigned for six temperature curves (CMT1-3, CMT1-4, CMT1-5, CMT1-6, CMT3-4, and CMT3-5) with a maximum temperature decrease of > 0.3 °C. In contrast, five points can sufficiently represent the main features of the other 18 temperature curves with a maximum temperature decrease of < 0.3 °C. It was noticed that the temperature subtly

increased at most 0.1 °C at PW2-2 and PW2-3 (see Figure 7.2 parts xxi and xxii). Hence, the temperatures at these two locations were treated as unchanged. In summary, a total of 210 representative points were obtained from the temperature curves.

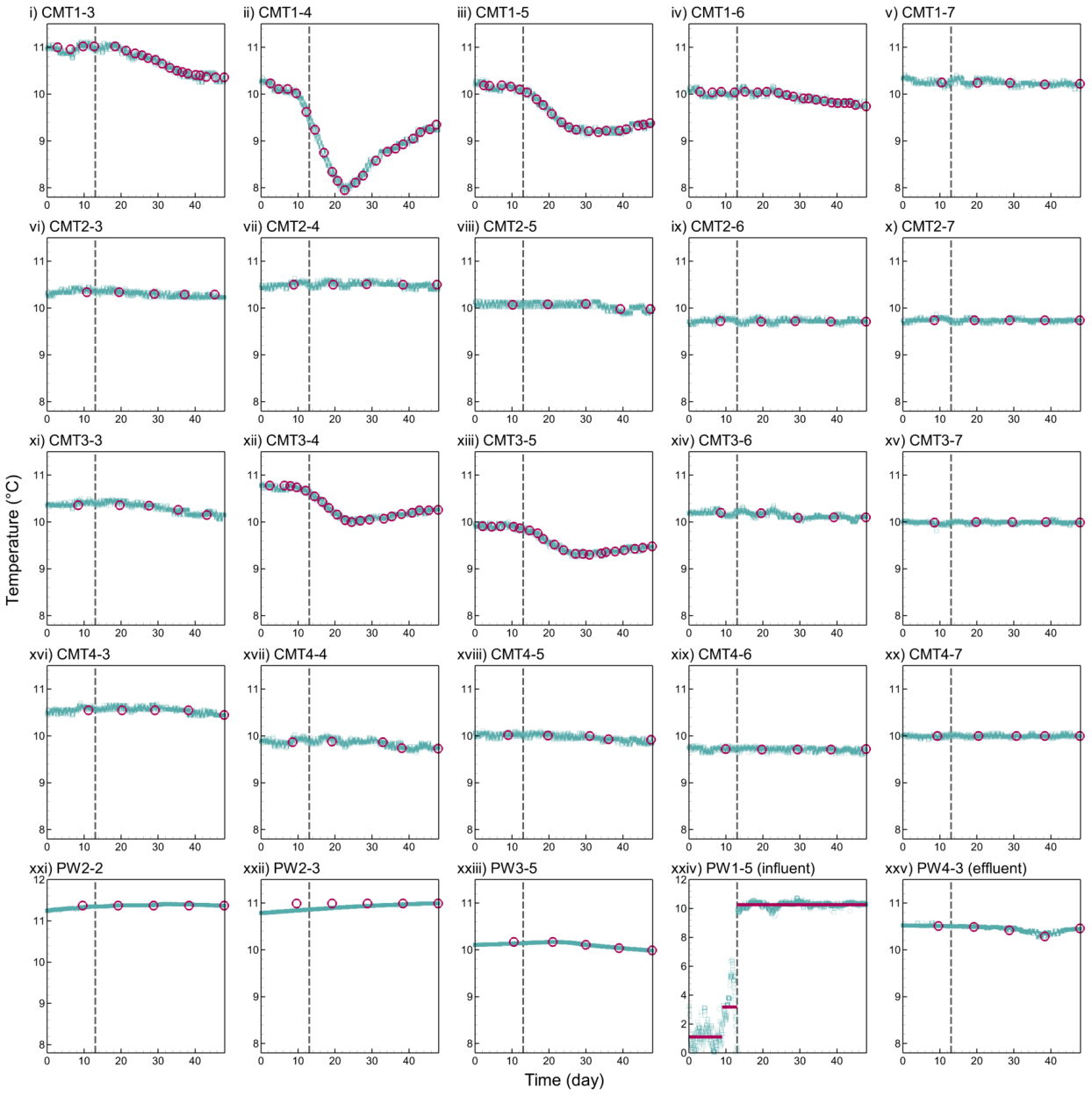


Figure 7.2: Observed temperature curves over time for all observation points during the heat tracer test. The continuous temperature records were provided in green hollow dots. Red hollow circles are selected to outline the main characteristics of temperature curves. The black dashed line indicates the beginning of

the heat dissipation phase of the experiment (Stage 3). The red solid horizontal lines in xxiv) represent the average temperatures of injected water over these periods, measured in PW1-5.

7.3.3 Borehole Testing Data

Borehole testing methods incorporated into this study include downhole nuclear magnetic resonance (NMR), and hydraulic profiling tool (HPT) logging surveys, as well as permeameter tests on core samples. Figure 7.3 displays the *K* profiles from permeameter measurements and two co-located HPT-derived *K* profiles, as well as the total and effective porosity profiles from NMR logging. Meanwhile, borehole geology logs compiled by Zhao and Illman (2017) are provided for comparison.

NMR logs were collected using multiple NMR tools from Vista Clara Inc. (Mukilteo, Washington, USA) from May 24th to 27th, 2023. Four direct-push NMR logs were performed by Wang et al. (2025a) within 1 m of the four CMT wells (CMT1, CMT2, CMT3, and CMT4) using a Dart probe (DP140N) specifically designed for use with a direct-push drill rig. The exploration depth of these direct-push logs was ~12 mbgs. Four wireline NMR logs were performed by Wang et al. (2025b) in the four pumping wells (PW1, PW2-3, PW4-3, and PW5), using the JPY350 and DP175 probes to collect NMR signals in 4” and 2” wells, respectively. The NMR measurements using DP140N and DP175 have a vertical spacing of ~0.15 m, while the JPY350 probe yielded measurements at a spacing of ~0.30 m. In total, eight NMR logs were collected at the NCRS, yielding 361 NMR measurements. NMR signals can be used to estimate multiple hydraulic parameters (Behroozmand et al., 2015), including *K*, total water content, and mobile water content. Here, the total and mobile water content provided by NMR logging were incorporated in this study for the estimation of total porosity and effective porosity in the saturated zone. The NMR-derived effective porosity was determined by distinguishing the mobile fluid volume of the relaxation time (T_2) distribution from the immobile fluid volume, in which a cutoff time, typically 33 ms for unconsolidated materials, was used. Additionally, a bimodal distribution of relaxation time was visible in the PW1 and PW5 logs, which came from the fine content deposition near the well screen and

was previously reported by Wang et al. (2025b). Therefore, the total porosity logs in PW1 and PW5 were excluded from the dataset, while the effective porosity logs were not affected by the increased fine content.

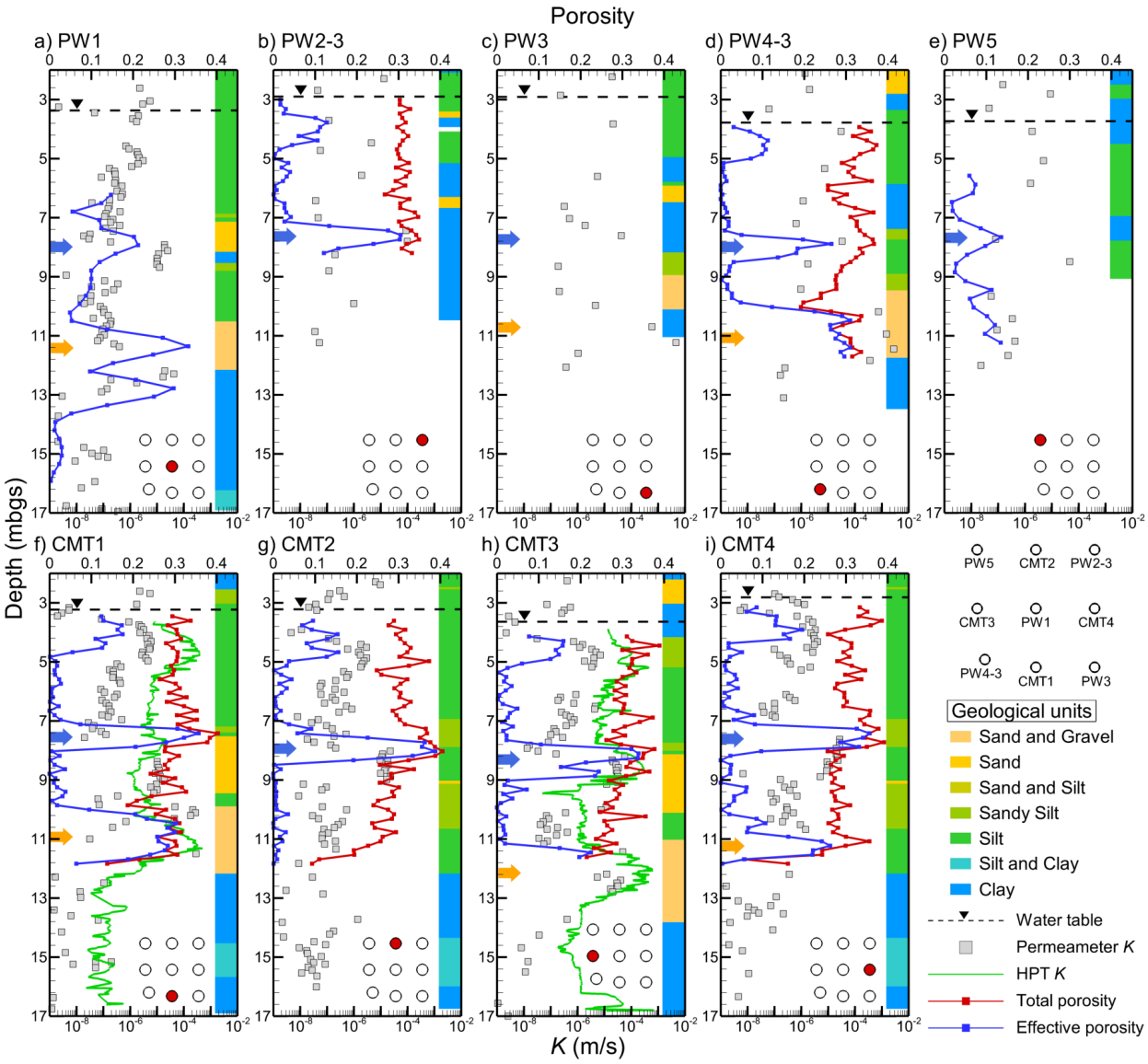


Figure 7.3: Downhole profiles of *K* and porosity at nine observation wells of the NCRS. *K* profiles include permeameter measurements at all nine wells and their co-located HPT-derived *K* values at CMT1 and CMT3. Porosity profiles for both total and effective porosity were determined by NMR logging. The total porosity profiles at PW1 and PW5 are excluded. Compiled borehole geology logs are provided for comparison, in which seven soil types were defined, including Sand and Gravel, Sand, Sand and Silt, Sandy Silt, Silt and Clay, and Clay. The locations of Aquifers 1 and 2 are marked with blue and yellow arrows. Water level measurements were conducted during the NMR logging period.

HPT surveys were conducted at 11 locations from July 30th to August 1st, 2019 by Sun et al. (2022). The HPT boreholes were distributed inside and outside the nine-well square, as shown in Figure 7.1a. An HPT probe (Model K6040, Geoprobe) was used for all HPT surveys, which advanced to a depth range of 14.0-17.1 mbgs during the logging process. It is worth noting that three HPT logs were collected near two monitoring wells (see Figure D1 for the naming and estimated *K* values of all HPT logs). Two HPT logs, HPT6 and HPT6-2, were < 1 m distance away from CMT3, and another HPT log (HPT3) was collected adjacent to CMT1. The data from the HPT probe, including water injection pressure, flow rate, and electrical conductivity, were recorded at a 1.5-cm vertical interval. A site-specific equation to estimate *K* was developed by Zhao and Illman (2022b), utilizing a power-law relationship:

$$\begin{aligned} \log K &= 1.41 \log\left(\frac{Q_{inj}}{P_c}\right) + 4.25, \frac{Q_{inj}}{P_c} < 5.5 \times 10^{-6}\ m^2/s \\ K &= 6.9 \times 10^{-4}\ m/s, \frac{Q_{inj}}{P_c} \geq 5.5 \times 10^{-6}\ m^2/s \end{aligned} \tag{7.1}$$

where Q_{inj} is the injected flow rate, and P_c is the correlated water injection pressure after removing the hydrostatic and atmospheric pressures. Other equations to estimate *K* using the HPT data include a model developed by Borden et al. (2021). The site-specific Equation (7.1) enabled more reliable drawdown predictions at the NCRS based on the numerical simulations by Zhao and Illman (2022b). In total, 8602 *K* measurements were obtained from 11 HPT logs.

For permeameter measurements, a 10-cm split-spoon sampler was used by Alexander et al. (2011) to collect continuous vertical core samples at an interval of ~0.1 m during the drilling of boreholes at CMT1, CMT2, CMT3, CMT4, and PW1. In addition, selective sampling was conducted in PW2-3, PW3, PW4-3, and PW5 by Zhao and Illman (2017). These core samples were delivered to a laboratory on the UW campus for falling-head permeameter tests. As a result, a total of 539 soil samples were collected from nine wells. The locations of two aquifer layers, identified as Aquifers 1 and 2, were marked on these

downhole profiles (see Figure 7.3) based on the borehole testing data (permeameter measurements, HPT-derived *K* profiles, and NMR-derived effective porosity profiles).

7.3.4 Hydraulic Tomography Survey

Previously reported by Berg and Illman (2011b) and Zhao and Illman (2017), 13 pumping tests were conducted at the NCRS, which yielded drawdown-time data until achieving steady-state flow conditions. As the chilled water was injected into Aquifer 2 during this heat tracer test, this dedicated HT analysis aimed to map the *K* heterogeneity for the potential tracer migration zone (Aquifer 2 and its adjacent layers). Therefore, steady-state head observations from eight pumping tests that stressed the porous medium in the vicinity of Aquifer 2 were incorporated into this HT analysis.

Two calibration datasets were created to investigate the HT performance with varying degrees of head information: the full and limited datasets. The "full" HT dataset leveraged the full capabilities of the monitoring network at the NCRS. It incorporated eight pumping tests stressed at PW1-4, PW1-5, PW1-6, PW1-7, PW3-3, PW4-3, PW5-3, and PW5-5, with observation data from all nine wells. In comparison, the limited HT dataset used pumping tests conducted at PW1-4, PW1-5, PW4-3, and PW5-5 for model calibration, with observation data only from PW1 and four CMT wells. In total, the full HT dataset incorporated 272 steady-state head measurements for model calibration, while the limited dataset scenario utilized 99 head measurements. Table D2 provides the information on all pumping tests used for model calibration.

7.4 Heat Transport Models

7.4.1 Numerical Modeling

The groundwater flow and heat transport simulations were conducted using HydroGeoSphere (Aquanty Inc., 2025). To minimize the impacts of boundary conditions, the simulation domain was

assigned dimensions of 70 m × 70 m × 17 m with PW1 located in the center (refer to Figure 7.1a). The simulation domain was discretized into variably sized rectangular finite elements, with a minimum dimension of 0.3 m in the z-direction (see Figure D2), corresponding to the NMR logging interval. Here, the top and bottom boundaries were assigned to be impermeable, while constant heads were defined for the lateral boundaries.

Four approaches were adopted to derive spatial *K* distributions at the NCRS based on different datasets. The first approach used the homogeneous and anisotropic representation of the *K* field (*Uniform_anisotropic*). The second approach employed the ordinary kriging estimation for the interpolation and extrapolation of borehole *K* data. Two types of downhole *K* profiles were applied for kriging, including HPT logging (*Krig_HPT*) and permeameter measurements (*Krig_permeameter*). The third approach used a 19-layer geologically interpreted zonation model (Figure D3) (Zhao & Illman, 2017) to represent the zone geometry, with internally homogeneous *K* values derived from permeameter tests without calibration (*GeoZone_permeameter*) and *K* values estimated for each layer by calibration to eight pumping tests (*GeoZone_pumping*). The last approach is the highly parameterized HT analysis, in which the limited (four pumping tests) and the full (eight pumping tests) HT calibration datasets were used in *HT_4* and *HT_8*, respectively, to characterize *K* heterogeneity. In total, seven cases of different spatial *K* models were performed for groundwater flow and heat transport simulations, as summarized in Table 7.2.

Table 7.2: Summary of seven comparative cases of spatial *K* models.

Case ID	Data source	Number of data observations
Uniform_anisotropic	Pumping tests	272 steady-state heads in 8 pumping tests
Krig_HPT	HPT logging survey	8602 HPT-derived *K* in 11 boreholes
Krig_permeameter	Permeameter tests	539 falling-head permeameter *K* in 9 boreholes
GeoZone_permeameter	Permeameter tests	539 falling-head permeameter *K* in 9 boreholes
GeoZone_pumping	Pumping tests	272 steady-state heads in 8 pumping tests
HT_4	Pumping tests	99 steady-state heads in 4 pumping tests
HT_8	Pumping tests	272 steady-state heads in 8 pumping tests

Transport and thermal parameters in numerical modeling were determined either using field measurements or were based on literature values. A heterogeneous porosity field was derived by populating the NMR-derived effective porosity estimates into the 19-layer geological zonation model. As this study assumed that the hydraulic mixing process can be represented using an accurate mapping of K heterogeneity, only local dispersion was considered in numerical models. Correspondingly, a longitudinal dispersivity of 0.5 m was assigned to the entire simulation domain, which is consistent with the minimum dimension of the rectangular elements in the xy-plane. Meanwhile, transverse dispersivities in the horizontal and vertical directions were assumed to be 1/5 (0.1 m) and 1/10 (0.05 m) of the longitudinal dispersivity, respectively. The bulk thermal conductivity, ρ_b, is comprised of solid-phase thermal conductivity, ρ_s and water thermal conductivity, ρ_w:

$$\rho_b = (1-\theta)\rho_s + \theta\rho_w \tag{7.2}$$

where θ is the total porosity. A laboratory sample analysis of Ontario field soil by Tarnawski et al. (2015) measured seven samples and yielded an average ρ_s of 2.25 $W \cdot m^{-1} \cdot K^{-1}$, with a range from 1.74 to 2.44 $W \cdot m^{-1} \cdot K^{-1}$. This average ρ_s was applied in numerical simulations. The specific heat capacity of the solid only varies slightly. Therefore, the value of the "Soil, dry" class in the Engineering Toolbox dataset (ETB, 2025) was used, which is 800 $J \cdot kg^{-1} \cdot K^{-1}$. These transport and thermal parameters are summarized in Table 7.3.

Table 7.3: Transport and thermal parameters used in heat transport simulations, including dispersivities, thermal conductivity, and specific heat capacity of the solid.

Parameters	Values
Longitudinal dispersivity	0.5 m
Horizontal-transverse dispersivity	0.1 m
Vertical-transverse dispersivity	0.05 m
Solid thermal conductivity	2.25 $W \cdot m^{-1} \cdot K^{-1}$
Solid specific heat capacity	800 $J \cdot kg^{-1} \cdot K^{-1}$

7.4.2 Representations of K and Porosity

Comparative cases of heterogeneous *K* representations are shown in Figures 7.4 and 7.5 for the cross-sectional views of A-A' and B-B', and of C-C' and D-D', respectively. Figure D4 exhibits the perspective views of all *K* fields within the entire simulation domain.

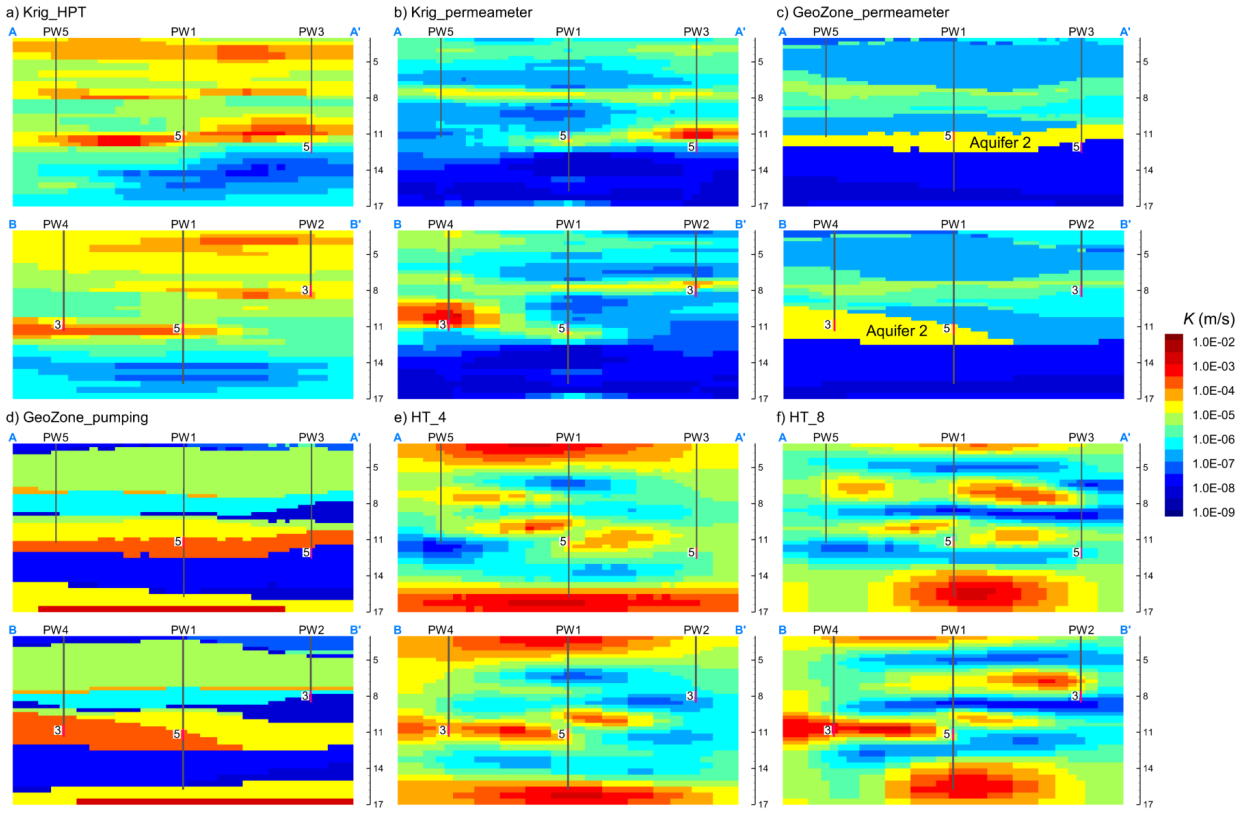


Figure 7.4: Contour plots of *K* distributions in A-A' and B-B' sections based on comparative *K* representations for heat transport simulations. These *K* fields are from models a) *Krig_HPT*, b) *Krig_permeameter*, c) *GeoZone_permeameter*, d) *GeoZone_pumping*, e) *HT_4*, and f) *HT_8*. The available observation ports for temperature monitoring are marked in the cross-sections. The location of Aquifer 2 is marked in c) *GeoZone_permeameter*.

The model-independent parameter estimation code, PEST (Doherty, 2015), was used in conjunction with the groundwater flow simulator in VSAFT3 (Variably Saturated Flow and Transport 3-D model, Yeh et al., 1993) for the *K* estimation of the *Uniform_anisotropic* and *GeoZone_pumping* models. The *Uniform_anisotropic* inverse model used an initial *K* of 1×10^{-5} m/s for K_{xx}, K_{yy}, and K_{zz}. After

calibration, K_{xx} was estimated to be 1.83×10^{-5} m/s, with upper and lower 95% confidence intervals being 2.50×10^{-5} m/s and 1.35×10^{-5} m/s, respectively. Likewise, K_{yy} after calibration was estimated to be 1.48×10^{-5} m/s, with upper and lower 95% confidence intervals of 2.02×10^{-5} m/s and 1.09×10^{-5} m/s, respectively. The K_{zz} after calibration was estimated to be 1.06×10^{-6} m/s, with upper and lower 95% confidence intervals of 1.59×10^{-6} m/s and 7.09×10^{-7} m/s, respectively.

The spatial *K* fields in models *Krig_HPT* and *Krig_permeameter* were derived using ordinary kriging estimation by the interpolation of 11 HPT-derived and nine permeameter-based *K* profiles, respectively. The *Krig_HPT* model was previously constructed by Zhao et al. (2023) and referred to as the *HPT-krig_exp* case. The exponential variogram model was used to fit experimental variograms, yielding a sill of 4.09 and correlation lengths of 4.06 m, 6.38 m, and 0.64 m for the x-, y-, and z-directions, respectively. The extrapolation of the ordinary kriging estimation was extended to the lateral boundaries of the entire simulation domain. The *Krig_permeameter* model was previously constructed by Wang et al (2025b) and referred to as the *Krig_ordinary_9 (permeameter)* case. Similarly, the exponential variogram model was used to fit the experimental variograms, yielding a sill of 7.4, a nugget of 0.4, and correlation lengths of 17.6 m, 17.6 m, and 2.2 m for the x-, y-, and z-directions, respectively. The kriged *K* fields from permeameter *K* estimates were gridded within a 20 m × 20 m data domain encompassing the borehole data. As the study space exhibits a strong layered structure, *K* values at the outer edge of the kriged fields (20 m × 20 m data domain) were horizontally extrapolated to the boundary of the numerical models (70 m × 70 m simulation domain). In sum, the *K* field of the *Krig_HPT* model was presented in Figures 7.4a and 7.5a, and the *K* field of the *Krig_permeameter* model was exhibited in Figures 7.4b and 7.5b.

The zonal *K* values for the *GeoZone_permeameter* model were determined by calculating the geometric mean of the permeameter-based *K* within each zone of the 19-layer geological zonation model (Zhao & Illman, 2017). Similarly to the *Uniform_anisotropic* case, an initial *K* of 1×10^{-5} m/s was

assigned for the zonal *K* of 19 layers in the *GeoZone_pumping* model. The optimized zonal *K* estimates from eight pumping tests, along with the upper and lower limits of the 95% confidence intervals for each layer, are listed in Table D3. Note that the full HT dataset was also used to calibrate the *Uniform_anisotropic* and *GeoZone_pumping* inverse models. The *K* fields of the *GeoZone_permeameter* and the *GeoZone_pumping* models are presented in Figures 7.4c-d and 7.5c-d. The 19 zonal *K* values from pumping tests and permeameter tests are summarized in Table 7.4.

Table 7.4: Zonal *K* and porosity values for the 19-layer geological zonation model. Zonal *K* values include permeameter-based *K* values (the *GeoZone_permeameter* model) and *K* estimates from pumping tests (the *GeoZone_pumping* model). Zonal porosity values include NMR-derived total porosity and effective porosity, as well as the ratio between effective and total porosity.

Layers	Soil type	Permeameter *K* (m/s)	Estimated *K* (m/s)	NMR-derived Porosity		
				Effective	Total	Ratio (%)
1	Clay	2.7E-07	3.1E-06	0.086	0.341	25.3
2	Silt	7.8E-07	5.7E-08	0.096	0.308	31.1
3	Sand	1.5E-07	2.0E-06	0.106	0.310	34.1
4	Clay	6.1E-08	1.1E-08	0.056	0.327	17.1
5[(1)]	Sand and Silt	2.4E-06	5.4E-04	0.073	0.296	24.7
6	Sandy silt	2.4E-06	1.0E-08	0.107	0.324	32.9
7	Silt	3.1E-07	7.2E-06	0.051	0.311	16.6
8	Clay	1.8E-06	4.9E-06	0.101	0.335	30.1
9	Sandy silt	5.0E-06	7.9E-05	0.179	0.334	53.6
10	Silt	7.5E-06	6.4E-07	0.165	0.334	49.4
11	Sand	1.3E-06	5.9E-07	0.074	0.291	25.5
12	Clay	3.7E-07	1.0E-08	0.051	0.268	18.9
13	Sandy silt	1.2E-06	5.6E-06	0.057	0.253	22.6
14	Silt	1.1E-07	2.4E-05	0.084	0.251	33.7
15	Sand and Gravel	1.2E-05	1.2E-04	0.238	0.281	84.6
16	Clay	2.0E-08	1.2E-08	0.080	0.316	25.2
17[(2)]	Silt and Clay	2.4E-08	1.0E-08	0.075	0.308	24.3
18	Clay	4.7E-09	1.2E-05	0.003	0.316	0.8
19[(2)]	Silt and Clay	2.4E-08	1.0E-03	0.075	0.308	24.3

[(1)]NMR-derived porosity for Layer 5 is estimated as the average of porosities for the Sand and the Silt layers.

[(2)]NMR-derived porosity for Layers 17 and 19 is estimated as the average of porosities for the Silt and the Clay layers.

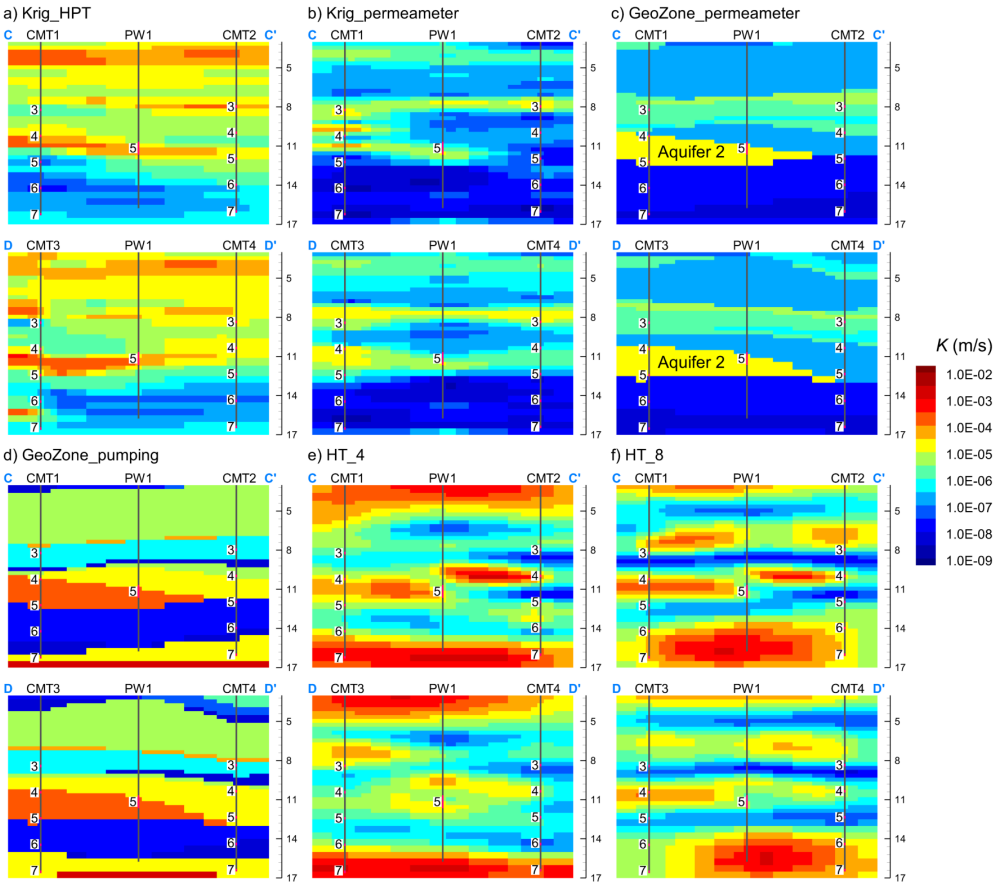


Figure 7.5: Contour plots of *K* distributions in C-C' and D-D' cross-sections based on comparative *K* representations for heat transport simulations, as shown in a) – f). The available observation ports for temperature monitoring are marked in the cross-sections. Again, the location of Aquifer 2 is marked in c) *GeoZone_permeameter*.

The *K* fields in models *HT_4* and *HT_8* were estimated by calibrating the initial homogeneous *K* fields via a geostatistical inverse algorithm with the SimSLE code (Xiang et al., 2009). An optimal iteration step was determined when both the biggest head misfit and the overall mean squared error decreased to relatively small prescribed values. Meanwhile, the range of *K* and the variance of the ln *K* field at the selected iteration step were anticipated to be close to the a priori knowledge (e.g., geology logs, permeameter measurements). Based on the comparisons between the simulated and observed head responses from the calibration datasets, both geostatistical inverse models were successfully calibrated. Figure D5 presents the calibration scatterplots of simulated versus observed drawdowns for the *HT_4* and

HT_8 models, as well as inverse models *Uniform_anisotropic* and *GeoZone_pumping*, in which the mean absolute error (MAE) and the root-mean-squared error (RMSE) for the overall drawdown fits were computed.

The geometric mean of K and the variance of ln K for the selected iteration of the *HT_4* model were 7.61×10^{-6} m/s and 5.12, with the K field ranging from 1.29×10^{-8} m/s to 3.48×10^{-3} m/s. For the optimal iteration of the *HT_8* model, the geometric mean of K and the variance of ln K are 3.69×10^{-6} m/s and 5.07, with K ranging from 8.80×10^{-9} m/s to 1.59×10^{-3} m/s. The K fields of models *HT_4* and *HT_8* are presented as Figures 7.4e and 7.5e, and Figures 7.4f and 7.5f, respectively. The statistics of all comparative K fields, including the geometric mean of K, the variance of ln K, and the maximum and minimum K values within the data domain, were compared using a line chart in Figure D6.

Zonal porosities, including total porosity and effective porosity, were determined in a similar fashion to the zonal K fields of the *GeoZone_permeameter* model. The downhole NMR-derived porosity estimates located within each zone were used to calculate the zonal porosity estimates (Figure 7.6). The effective porosity field (see Figure 7.6b) was used in all heat transport cases using various K representations. The zonal total and effective porosities from NMR logging are summarized in Table 7.4.

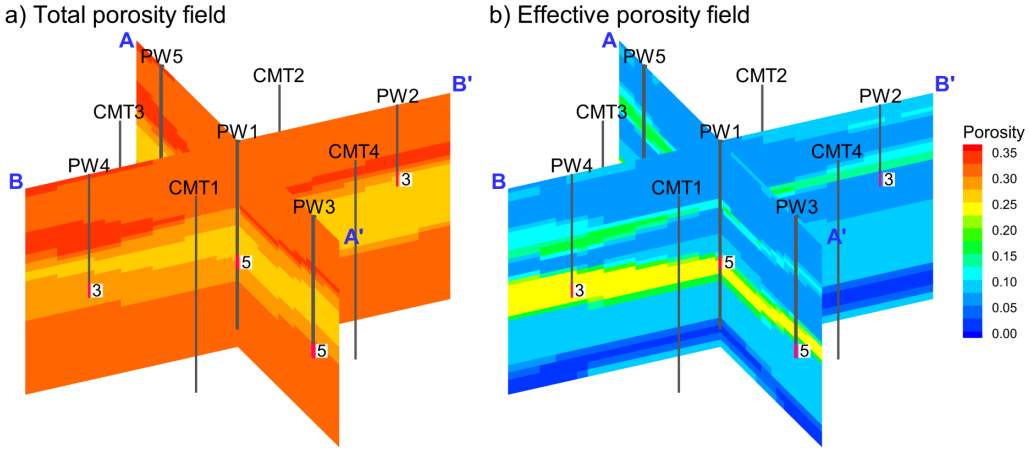


Figure 7.6: The perspective view of the porosity fields within the data domain, in which a) shows the total porosity field, and b) shows the effective porosity field. The screen locations along the two cross-

sections (A-A’ and B-B’) are highlighted in red. The chilled water was injected in PW1-5, and groundwater was simultaneously pumped in PW4-3.

7.5 Results of Heat Transport Simulations

7.5.1 Simulated Temperature Breakthrough Curves

Figure 7.7 displays the simulated versus observed temperature breakthrough curves (BTCs), which show the temperature decreases over time for all observation ports. These observed BTCs indicate that the heat tracer plume primarily migrated to the southwest of the data domain, as evidenced by the large temperature decreases at CMT1 and CMT3. In contrast, only minor temperature decreases of < 0.3 °C were observed at CMT2, CMT4, and PW3, and no temperature decreases were observed at the PW2 wells. Additionally, only a slight temperature decrease was observed at the effluent port, PW4-3. For the depth-specific temperature decreases, highly variable vertical temperature profiles were shown for CMT1 and CMT3. The most significant temperature decreases occurred at the “-4” intervals of the CMT wells, with the largest temperature decrease exceeding 2 °C at CMT1-4. Furthermore, the heat tracer plume did not travel to all “-7” intervals (i.e., the largest temperature decrease smaller than 0.1 °C).

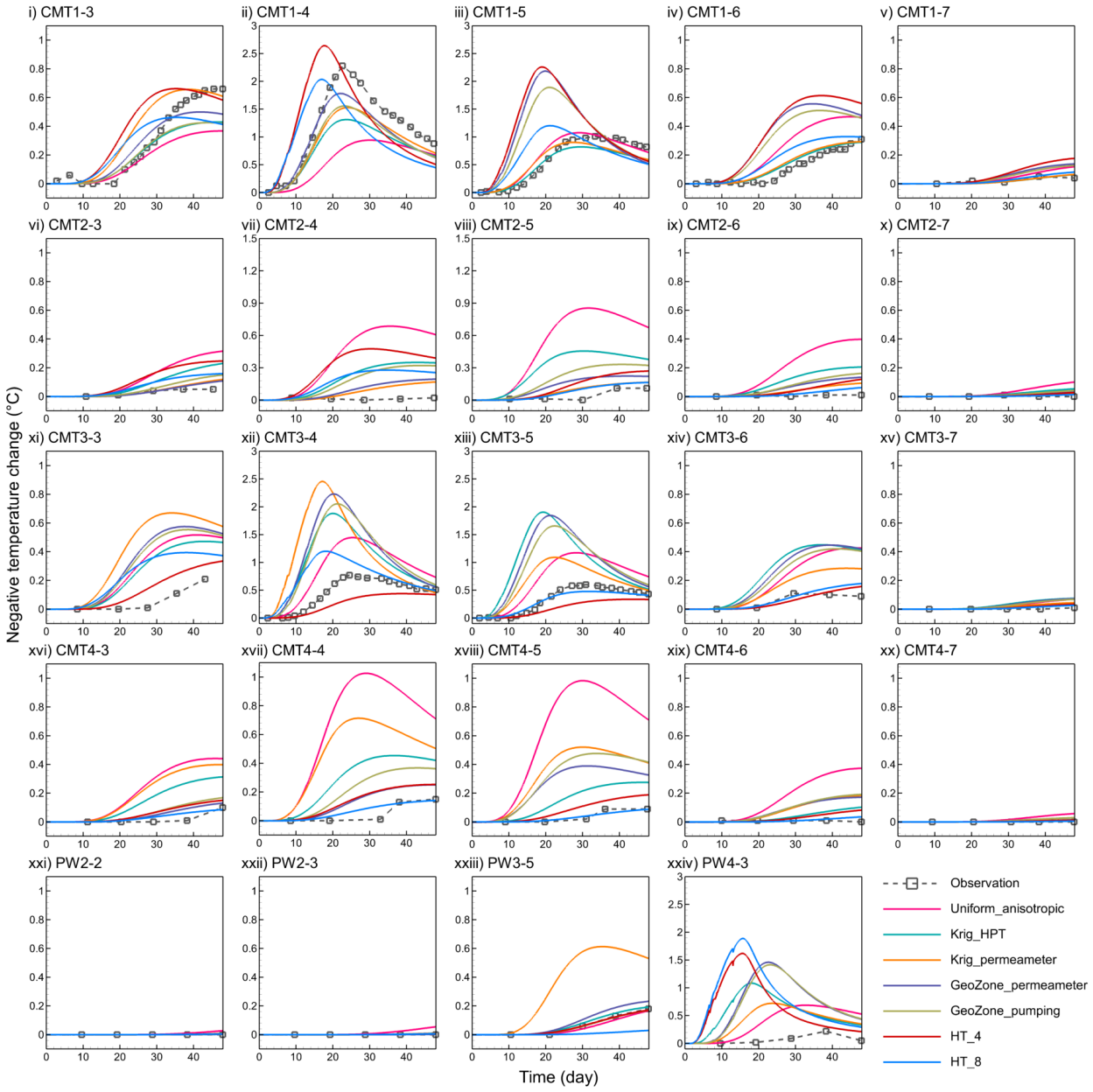


Figure 7.7: The temperature breakthrough curves from field observations and model simulations for all observation ports. Five or twenty points on the observed BTCs are used to represent their main characteristics. The numerical simulations of heat transport are based on different cases of *K* representations, including models *Uniform_anisotropic*, *Krig_HPT*, *Krig_permeameter*, *GeoZone_permeameter*, *GeoZone_pumping*, *HT_4*, and *HT_8*.

Different *K* representations led to disparate patterns of simulated BTCs in the 3D monitoring network. First, the *Uniform_anisotropic* model yielded similar shapes of BTCs for the observation ports

at the same distance from the injection port, PW1-5. For example, PW1-5 is at the mid-depth of the "-4" and "-5" channels in the CMT wells. As a result, the simulated BTCs at the "-4" and "-5" ports of CMT1 through CMT4 exhibited similar characteristics in terms of maximum temperature decreases and peak arrivals. Similar trends of simulated temperature decreases were also observed at the "-3" and "-6" channels of the CMT wells. In contrast, field observations exhibited highly non-uniform spreading and migration of the heat tracer plume. They revealed a dominant advective transport of the heat tracer towards CMT1 and CMT3, with observations of more significant temperature decreases and shorter arrival times at CMT1-4 than at CMT1-5. Moreover, the non-Fickian features of the BTCs were not captured by the *Uniform_anisotropic* model. For instance, the observed BTC at CMT1-4 displayed an early peak arrival with a significantly larger temperature decrease than the simulation result from the *Uniform_anisotropic* model.

The *Krig_HPT* and *Krig_permeameter* models yielded simulated BTCs with similar magnitudes and phases. Different from the *Uniform_anisotropic* model, reduced temperature decreases in CMT2 were simulated for the kriging cases. However, the temperature decreases in CMT4 were still overestimated. It is worth noting that both cases resulted in larger temperature decreases in CMT3 than in CMT1, which contradicts the field observations. Furthermore, the *Krig_permeameter* model was the only case that yielded significant temperature decreases in PW3-5, which was inconsistent with observed data.

Two zonation models, *GeoZone_permeameter* and *GeoZone_pumping*, also led to similar trends of variable BTCs at different locations. They reasonably predicted the dominant migration paths to CMT1 and CMT3, and lesser spreading to CMT2 and CMT4 was reproduced. Nevertheless, they still yielded larger temperature decreases in CMT3 than in CMT1. Moreover, biased predictions of depth-specific temperature decreases were derived, as the simulated temperature decreases at CMT1-5 were larger than those at CMT1-4.

Based on the comparison of BTCs, the irregular spatial distribution of the observed heat tracer plume was reasonably reproduced by models *HT_4* and *HT_8*. The two HT models successfully simulated the preferential migrating paths towards CMT1 and CMT3, with larger temperature decreases in CMT1, and the reduced spreading of the heat tracer to CMT2 and CMT4 was adequately reproduced. In the vertical profiles, the trends of simulated temperature decreases in CMT1 and CMT3 were consistent with those observed in the field. Specifically, more significant temperature decreases were simulated at the "-4" interval than at the "-5" interval, and the temperature decreases at the "-3" interval were larger than those at the "-6" interval. Compared to the *HT_4* model, *HT_8* yielded improved fits with observed BTCs, particularly in terms of more consistent peaks (e.g., at CMT1-5 and CMT1-6). However, the simulated peak arrivals at most ports were still earlier than field observations, which also occurred in the simulations using other *K* representations. Furthermore, the observed BTC at the effluent port (PW4-3) was not accurately reproduced in all simulations.

A scatterplot of the simulated versus observed temperature decreases at selected times on the BTCs is provided in Figure D7. It can be observed that more scatter points were located on the upper-left side of the 1:1 line for the simulations using the uniform, kriged, and zonation models. Based on the computed MAE and RMSE between model simulations and field observations (see Figure D7), the *HT_8* model achieved optimal fits of temperature decreases, with only a few large misfits. For a more rigorous quantitative comparison of the transport prediction performance, the temporal moment analysis (Leube et al., 2012) was employed, in which the temperature decrease at every hour of the BTCs was used to calculate the temporal moments through numerical integration of breakthrough data using the trapezoidal rule. The n^{th} temporal moments (M_{n}) of temperature decreases (ΔT) at location (x, y, z) at time (t) are given by:

$$M_n = \int_0^\infty t^n \Delta T(x, y, z, t) dt \tag{7.3}$$

For example, the zeroth (M_0), first (M_1), and second (M_2) temporal moments can be obtained using the moment generating function above by setting n as 0, 1, and 2, respectively. The absolute temperature decrease was divided by the temperature difference between the injected chilled water (1.11 °C) and the background groundwater (10.19 °C) to obtain the relative temperature decrease for the temporal moment analysis.

The tracer total response strength (M_0) at a given point, the mean arrival time (μ) of the tracer strength, and the variance (σ^2) of temperature BTCs were employed to quantitatively measure the heat transport behavior. The mean arrival time (μ) of the tracer strength and the variance (σ^2) of temperature BTCs were calculated by:

$$\mu = \frac{M_1}{M_0} \tag{7.4}$$

$$\sigma^2 = \frac{M_2}{M_0} - \left(\frac{M_1}{M_0}\right)^2 \tag{7.5}$$

Table 7.5 shows the computed M_0 for field observations and model simulations at all observation ports. The M_0 values at various observation ports were horizontally arranged in descending order based on the field observations (see Table 7.5). The MAE and RMSE between the simulated and observed M_0 were provided. The computed M_0 for field observations in Table 7.5 indicates that more tracer strength has migrated to monitoring intervals of CMT1 and CMT3, while only minor tracer strength was observed at monitoring intervals of CMT2 and CMT4. For example, CMT1-4 and CMT2-4 were located at a similar depth and were equidistant from PW1. However, the M_0 observed at CMT1-4 was more than 120 times as much as the M_0 at CMT2-4, revealing a highly irregular tracer distribution. The spatial distribution and magnitude of the M_0 from field observations can only be reasonably reproduced using the *HT_4* and *HT_8* models, in which the M_0 from the simulation by model *HT_8* had better correspondence with field observations, particularly in reproducing its spatial distribution more accurately. In contrast, the other

models failed to reproduce the dominating heat tracer migration behavior to CMT1 and CMT3 and the minor spreading to CMT2 and CMT4. For example, the M_0 at the "-4" intervals of CMT wells simulated by model *Uniform_anisotropic* was 2.90, 2.06, 4.41, and 3.15 for CMT1-4, CMT2-4, CMT3-4, and CMT4-4, respectively, showing a more uniform tracer spreading. Additionally, Table 7.5 indicates that the kriging and zonation models resulted in relatively larger M_0 (e.g., > 1.0 °C·day) at the location of CMT4 (e.g., in the interval of CMT4-4 or CMT4-5). Also, the zonation models (*GeoZone_permeameter* and *GeoZone_pumping*) yielded larger simulated M_0 at CMT1-5 than that at CMT1-4, contradicting the computed M_0 from field observations. Furthermore, the M_0 was overestimated at the effluent port (PW4-3) by all model simulations. The quantitative comparisons from the temporal moment analysis were consistent with the visual observations from the comparisons of temperature BTCs in Figure 7.7.

Table 7.5: The tracer total response strength (M_0) (°C·day) at each observation port for the observed and simulated heat transport. The observation ports are horizontally arranged in descending order of observed tracer strength from the temporal moment analysis. The MAE and RMSE between the simulated and observed M_0 for all observation ports are provided in the last two columns.

Total Strength (°C·day)	CMT1-4	CMT1-5	CMT3-4	CMT3-5	CMT1-3	CMT1-6	PW4-3	PW3-5	CMT3-6	CMT4-4	CMT3-3	CMT2-5	CMT4-5
Field observations	6.05	3.00	2.31	1.73	1.34	0.50	0.37	0.28	0.26	0.20	0.19	0.17	0.16
Uniform_anisotropic	2.90	3.32	4.41	3.63	0.85	1.15	1.97	0.23	1.00	3.15	1.13	2.62	3.03
Krig_HPT	4.02	2.57	5.16	5.10	1.05	0.61	2.72	0.34	1.30	1.38	1.00	1.48	0.72
Krig_permeameter	4.58	2.86	6.00	3.32	1.82	0.65	2.08	1.75	0.73	2.28	1.74	0.39	1.66
GeoZone_permeameter	5.00	5.66	5.85	5.05	1.29	1.65	3.65	0.43	1.21	0.63	1.35	0.65	1.26
GeoZone_pumping	4.50	5.16	5.59	4.70	1.06	1.47	3.58	0.27	1.10	0.99	1.27	0.96	1.49
HT_4	6.52	5.93	1.35	0.94	1.98	1.79	2.87	0.28	0.28	0.66	0.51	0.60	0.40
HT_8	5.22	3.61	3.76	1.57	1.41	0.86	3.62	0.04	0.35	0.32	1.01	0.36	0.16
Total Strength (°C·day)	CMT2-3	CMT1-7	CMT4-3	CMT2-4	CMT4-6	CMT2-6	CMT2-7	CMT3-7	CMT4-7	PW2-2	PW2-3	Comparison	
Field observations	0.11	0.10	0.07	0.05	0.03	0.02	0.01	0.01	0.01	0.00	0.00	MAE	RMSE
Uniform_anisotropic	0.55	0.18	1.07	2.06	0.82	0.92	0.15	0.11	0.09	0.03	0.06	1.08	1.48
Krig_HPT	0.37	0.09	0.68	0.99	0.17	0.48	0.08	0.14	0.02	0.01	0.01	0.79	1.23
Krig_permeameter	0.16	0.09	0.97	0.35	0.38	0.17	0.03	0.07	0.04	0.01	0.01	0.77	1.19
GeoZone_permeameter	0.15	0.25	0.22	0.46	0.39	0.28	0.05	0.13	0.04	0.00	0.00	0.87	1.42
GeoZone_pumping	0.22	0.22	0.30	0.85	0.41	0.33	0.06	0.12	0.04	0.00	0.01	0.89	1.35
HT_4	0.52	0.31	0.28	1.55	0.13	0.19	0.03	0.05	0.02	0.01	0.01	0.57	0.95
HT_8	0.32	0.13	0.15	0.89	0.05	0.09	0.01	0.04	0.01	0.00	0.01	0.40	0.80

The mean arrival time (μ) of the tracer strength was calculated as the first temporal moment (M_1) over the zeroth temporal moment (M_0). However, the computed M_0 for some monitoring intervals was very small and even approached zero (e.g., at PW2-2 and PW2-3), leading to biased μ values. Therefore, only the observation ports with the maximum temperature decrease > 0.3 °C (CMT1-3, CMT1-4, CMT1-5, CMT1-6, CMT3-4, and CMT3-5) and the effluent port (PW4-3), which exhibited large discrepancies between model simulations and field observations, were incorporated to compute the μ as well as the σ^2, which involved the first and the second temporal moment.

Table D4 exhibits the computed μ and σ^2 for both field observations and model simulations using various spatial K fields. The computed μ values at the seven ports were arranged in descending order based on the field observations, indicating faster to slower arrivals, while the computed σ^2 values at the seven ports were arranged in ascending order based on the field observations, indicating broader to narrower temporal distributions of the BTCs. It can be noted from Table D4 that the sequence of μ (from early to late) is consistent with, and the sequence of σ^2 (from large to small) is almost consistent with the sequence of M_0 (from large to small) for the six CMT observation ports with significant temperature decreases. Again, only the *HT_4* and *HT_8* models can generally reproduce the spatial variations of the μ and σ^2 values at these ports. However, the *HT_8* model consistently yielded earlier arrivals of the tracer center and broader temporal distributions of the temperature BTCs at all seven ports, thus resulting in unsatisfactory MAE and RMSE values for comparison of both μ and σ^2.

7.5.2 Heat Tracer Plume Migration

The comparisons of temperature BTCs (Figure 7.7) and the results of the temporal moment analysis (Tables 7.5 and D4) showed that the simulation result from model *HT_8* most closely resembled the field observations. Therefore, the simulated spatiotemporal evolution of the heat tracer plumes in model *HT_8* was used to show their migration and spreading patterns in Figure 7.8, where four

perspective views are provided, including a horizontal plane crossing the PW1-5 (influent port) and PW4-3 (effluent port), and three cross-sections of B-B', C-C', and D-D' (see Figure 7.1a). The heat tracer plumes at four time points were simulated, including 3 days, 13 days (immediately following the heat dissipation), 16 days, and 48 days (the end of the tracer test) after chilled water injection.

Highly asymmetric heat tracer plumes were obtained at these times. After three days of chilled water injection, the edge of the simulated heat tracer plume has arrived at PW4-3 and CMT1 and has almost reached CMT3. In comparison, the tracer spread slowly towards CMT2 and CMT4. After 13 days of chilled water injection, the heat tracer plume has significantly expanded, with a temperature decrease of > 4 °C near CMT1 and PW4, and a temperature decrease of ~2 °C near CMT3.

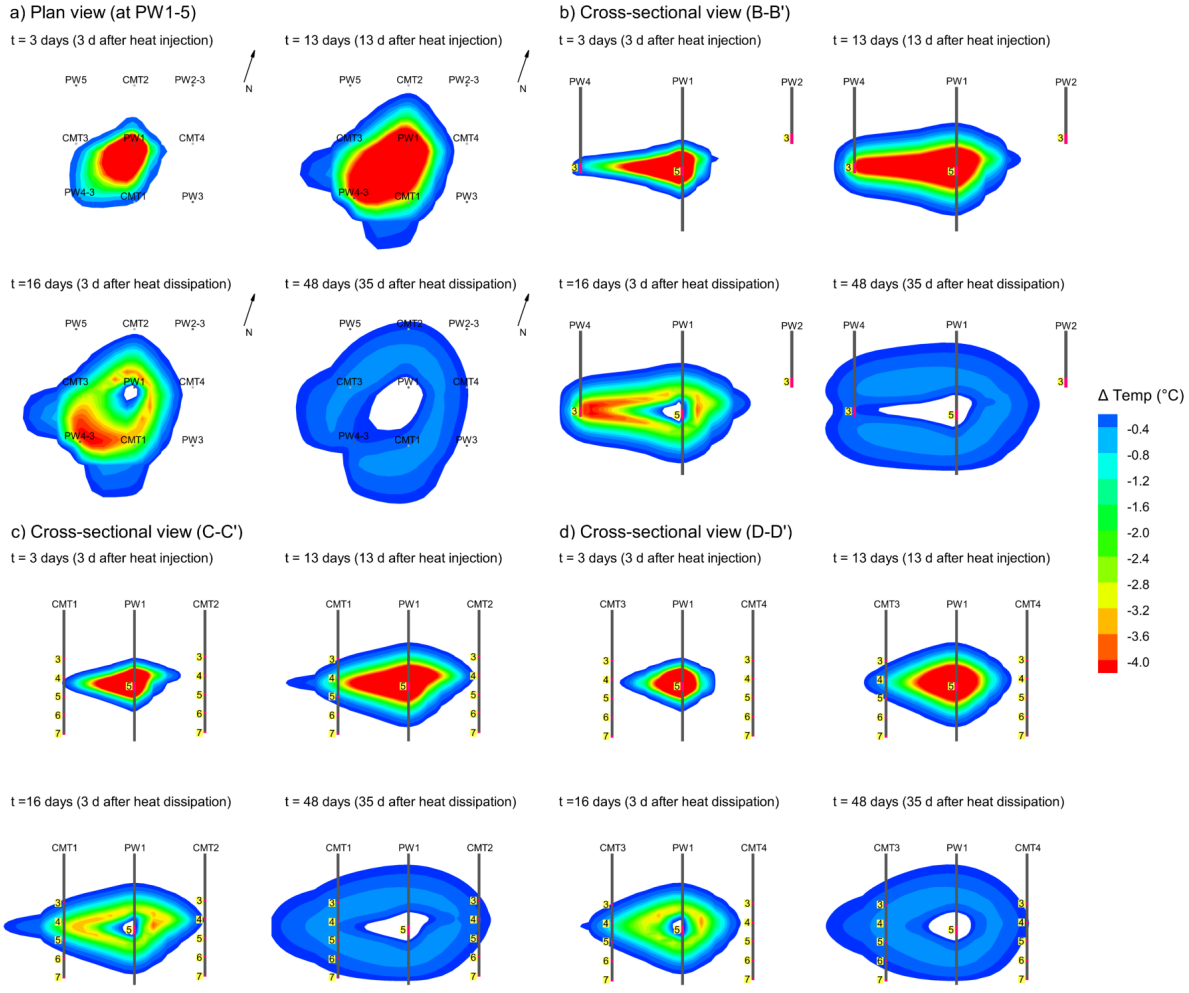


Figure 7.8: The simulated spatiotemporal evolution of heat tracer plumes in the plan view crossing PW1-5 and PW4-3, and three cross-sectional views of B-B', C-C', and D-D'. The temporal evolution of the heat tracer plume at four time points is simulated, including 3 days and 13 days after heat tracer injection, and 3 days and 35 days after heat tracer dissipation. This heat transport simulation used the *K* field of model *HT_8*.

The spatial distribution of the heat tracer plumes in those cross-sectional views exhibited that the simulated heat tracer plumes moved closer towards the "-4" interval in all four CMT wells, though the influent port (PW1-5) was at the mid-depth of the "-4" and "-5" channels of the CMT wells. After three days of tracer dissipation, a large proportion of the heat tracer plume for temperature decreases of > 3 °C has faded. In addition, the edge of the simulated heat tracer plume reached CMT2-4. After 35 days of heat dissipation, the groundwater temperature field has almost recovered to the background temperature, with a maximum temperature decrease of < 0.6 °C. At this late time, the heat tracer plume has arrived at CMT4-4.

7.5.3 Impacts of Transport and Thermal Properties

The sensitivity of heat tracer plume migration to transport and thermal parameters was examined. The studied parameters included porosity, dispersivities, and thermal conductivity. The specific heat capacity of the solid is known to vary only slightly for different soils, so this parameter was excluded from the sensitivity analysis. Here, the optimal *K* representation through HT analysis using the full calibration dataset (*HT_8*) was employed, and its numerical simulations were treated as the standard scenario for comparison. The transport and thermal parameters for the standard scenario are provided in Table 7.3.

In terms of porosity estimation, the heterogeneous total porosity field (refer to Figure 7.6a), which accounted for both immobile and mobile domains of the NMR-derived porosity, was considered as an alternative scenario ("Total porosity field"). Two more scenarios concerned the impacts of mechanical dispersion. Therefore, a "Larger dispersion" scenario assumed values of 2.5 m, 0.5 m, and 0.25 m for the longitudinal, horizontal-transverse, and vertical-transverse dispersivities, respectively. The ratio of the

dispersivities for the three directions remained the same as the standard scenario. Additionally, another scenario ("No dispersion") was considered, with all dispersivity components set to zero.

The last two scenarios were related to the uncertainty of thermal conductivity. The maximum and minimum solid thermal conductivities measured by Tarnawski et al. (2015) were used in the "Stronger thermal conduction" and "Weaker thermal conduction" scenarios. Figure 7.9 shows the temperature BTCs of the model simulations for all scenarios with different transport and thermal parameters, as well as the standard simulation scenario (*HT_8*), along with field observations at seven representative observation ports: CMT1-3, CMT1-4, CMT1-5, CMT1-6, CMT3-4, CMT3-5, and PW4-3. Additionally, the temporal moment analysis was conducted, yielding the computed M_0, μ, and σ^2 for the heat transport simulation of each scenario.

Generally, the spatiotemporal evolution of the heat tracer plume can be reasonably reproduced using the *K* representations from model *HT_8* and these transport and thermal parameters from all scenarios, as indicated by the visual comparisons of temperature BTCs at the six representative CMT observation ports (Figure 7.9) and the quantitative comparisons of computed M_0, μ, and σ^2 between model simulations and field observations (Table D5).

When the NMR-derived total porosity field was used for numerical modeling, the maximum simulated temperature decreases were generally reduced, accompanied by delayed peak arrivals. However, the discrepancies between the simulated and observed peaks increased at CMT1-3, CMT1-4, and CMT3-5, as shown in Figures 7.9a, b, and f. Based on the results of temporal moment analysis (Table D5), the misfits with the observed M_0 were comparable when using the effective and the total porosity field for simulation, while improved fits between the simulated and the observed μ and σ^2 were observed in the "Total porosity field" scenario.

When larger dispersivities were used for simulations, elevated misfits were obtained, especially for CMT3-4. Conversely, the minimum MAE and RMSE values for the comparisons of M_0, μ, and σ^2

were achieved when the dispersivities were assigned to be zero. Notably, the simulated peak arrivals were closer to the observations.

When the two boundaries of solid thermal conductivity for the Ontario soil were used in the numerical simulation, only a marginal shift of simulated BTCs was observed compared to the standard scenario. Based on the MAE and RMSE values, the "Stronger thermal conduction" scenario yielded a slightly improved M_0 fit with field observations, and the "Weaker thermal conduction" scenario achieved better μ and σ^2 fits with field observations, compared with the standard scenario. Again, no scenario was found to satisfactorily reproduce the main characteristics of the BTC at PW4-3.

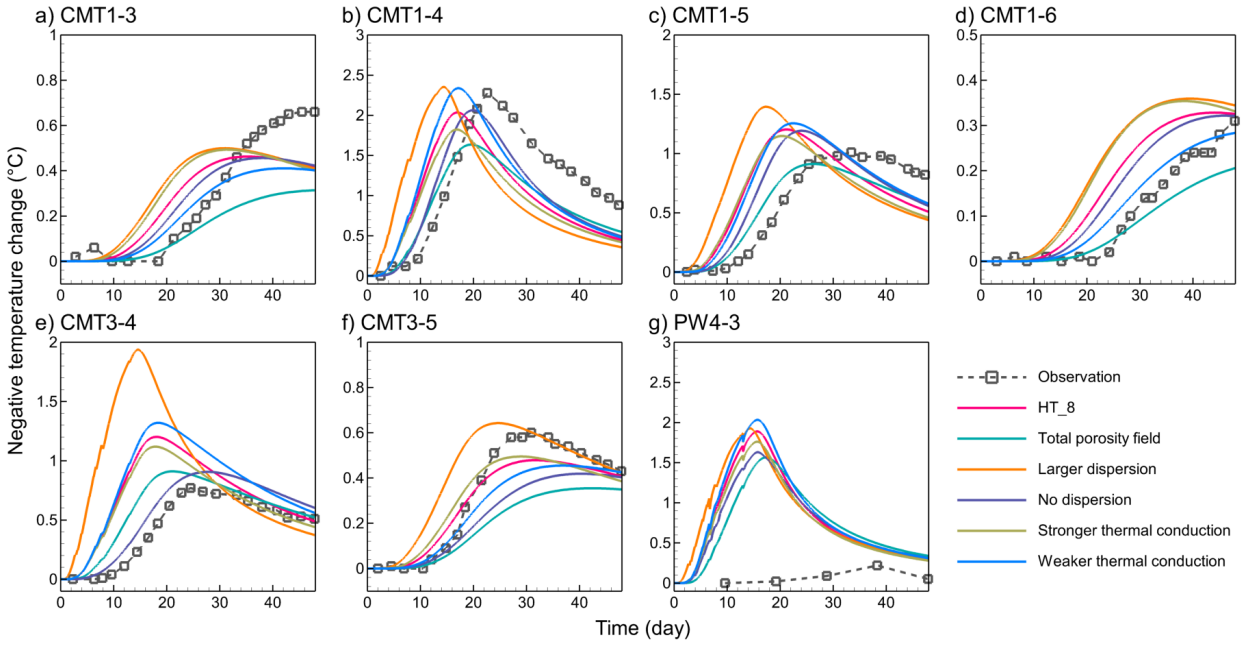


Figure 7.9: The simulated temperature BTCs at seven representative observation ports considering the impacts of transport and thermal parameters. These scenarios include "Total porosity field", "Larger dispersion", "No dispersion", "Stronger thermal conduction", and "Weaker thermal conduction". The simulated BTCs in *HT_8* are used as the standard simulation case, and field observations are provided for comparison. Figure parts a) – g) correspond to the BTCs at CMT1-3, CMT1-4, CMT1-5, CMT1-6, CMT3-4, CMT3-5, and PW4-3.

7.6 Discussion

7.6.1 Comparison of Predictions Using Different K Representations

In addition to HT analysis, various approaches have been covered in this study to represent the 3D *K* fields within the simulation domain for the heat transport predictions. An effective *K* representation was first adopted, which was popular for studies to simulate the transport behavior within a relatively homogeneous aquifer system (e.g., at the Borden site, see Sudicky, 1986). The kriging approach is a reliable spatial interpolation approach that has been extensively used in industry to derive spatial distributions of parameters based on hard data (e.g., borehole *K* measurements). The zonation approach can be used to delineate hydrostratigraphic units, particularly suitable for mapping the layered porous medium at the NCRS.

First, the numerical simulation using a homogeneous *K* representation was not able to reproduce the irregular spatial distribution of the observed heat tracer plume, as well as the highly asymmetric temperature BTCs for this highly heterogeneous aquifer system. The temporal evolution of the simulated heat tracer plume was more Gaussian-like in the *Uniform_anisotropic* model result, although a heterogeneous 3D porosity distribution was applied. The model fit with a certain observed BTC (e.g., at CMT1-4) might be improved by calibrating the transport parameters, such as porosity and dispersivities. However, the overall irregular spatial distribution of the migrating heat tracer plume cannot be faithfully reproduced in cases where the *K* heterogeneity was not accurately mapped. On the other hand, the relatively uniform spatiotemporal evolution of the simulated heat tracer plume indicates that the highly irregular shapes of the observed heat tracer plume, towards CMT1 and CMT3, were primarily caused by *K* heterogeneity rather than pumping at PW4-3.

Due to the mapping of *K* heterogeneity in the two kriging cases (*Krig_HPT* and *Krig_permeameter*), preferential flow towards CMT1 and CMT3 can be reproduced. However, the spatial

distribution of the simulated heat tracer plume did not accurately align with field observations. For example, significant temperature changes were simulated in CMT3, CMT1, and CMT4, in order of decreasing temperature decrease. This biased simulation result corroborated with our knowledge of *K* heterogeneity, in which slightly higher *K* values were observed in the region between CMT3 and PW1 compared to the region between CMT1 and PW1 (Figures 7.5a and b), and PW1 and CMT4 were also hydraulically connected with intermediate-to-high *K* materials. The inaccurate prediction of heat tracer spreading was likely due to the uncertainty in borehole *K* measurements. Based on the comparison of *K* profiles in CMT1 and CMT3 (Figures 7.3f and h), HPT logging cannot adequately reproduce the highly variable *K* values, ranging from gravel to clay, for this glaciofluvial aquifer system.

On the other hand, permeameter tests failed to precisely locate the aquifers (Wang et al., 2025a), as slight depth misalignments were found in locating Aquifer 2 among the permeameter results and the two in-situ logging surveys (see the *K* profiles in CMT1 and CMT4 in Figures 7.3f and i), which were usually encountered as soil compression and loss occurred during the core sampling process, especially for complex layered porous media, such as the NCRS (Alexander et al., 2011). Otherwise, the biased prediction may result from the uncertainty of the inter-well hydraulic connections that have not yet been captured through the kriging interpolation of existing borehole measurements.

Models *GeoZone_pumping* and *GeoZone_permeameter* can reproduce the dominant migration towards CMT1 and CMT3, as well as the secondary spreading to CMT2 and CMT4. The favorable predictions of the heat tracer plume distribution in the horizontal plane should benefit from the reasonable delineation of the layer structure for Aquifer 2 (see Figures 7.4c and 7.5c). For example, Aquifer 2 thickened towards CMT1 and CMT3, while it pinched out near CMT2 and CMT4. In terms of the vertical trends of simulated BTCs, the tracer migrating to CMT1 was biased to the “-5” interval, suggesting that the vertical layer connections between PW1 and CMT1 may still not be faithfully represented using the

19-layer geological zonation model, or the internally homogeneous zonal *K* distribution was not sufficient to accurately describe the vertical variations of tracer migration at high resolution.

Comparing the simulated BTCs for the four cases using kriging and zonation approaches (*Krig_HPT*, *Krig_permeameter*, *GeoZone_permeameter*, and *GeoZone_pumping*) for 3D *K* derivation (see Figure 7.7), the characteristics of simulated BTCs relied more heavily on the spatial conceptualization of the *K* fields than on borehole hard data. However, inter-well heterogeneity cannot be adequately mapped solely based on the interpretation of borehole data. Model *GeoZone_pumping* used head responses from multiple monitoring wells to interpret zonal *K* values; nevertheless, the zone geometry was still delineated using borehole geology information.

HT analysis is the only approach that can yield spatial distributions of the simulated heat tracer plume consistent with field observations, as indicated by comparing the simulated with the observed BTCs. The *K* fields in Figures 7.4 and 7.5 suggest that models *HT_4* and *HT_8* yielded *K* tomograms with similar patterns of spatial variability. When more pumping tests were utilized in HT analysis, the *K* representation in model *HT_8* was further refined. Based on the *K* field in model *HT_8* (see Figure 7.5f), Aquifer 2 has $K > 1 \times 10^{-4}$ m/s in between PW1 and CMT1, while lower *K* values by a factor of 5 were shown in between PW1 and CMT3. Furthermore, higher *K* values were mapped in CMT1-4 and CMT1-5 than in CMT3-4 and CMT3-5, respectively. Therefore, the high-resolution imaging of spatial *K* variability via HT analysis enabled the accurate prediction of the spatial distribution of the heat tracer plume. Moreover, it is worth noting that the HT cases characterized a high-*K* lens in between PW1 and CMT2, which intersected CMT2-4 (Figures 7.5e and f). On the other hand, the presence of low *K* materials above PW1-5 caused indirect hydraulic connections between PW1-5 and CMT2-4, leading to a minor spread of temperature decreases toward CMT2-4.

However, the observed BTC at the effluent port, PW4-3, was not satisfactorily simulated using all *K* representations. The observed BTC at PW4-3 exhibited a maximum temperature decrease of ~0.3 °C,

with a delayed peak arrival after 38 days, whereas all simulations yielded significantly larger temperature decreases accompanied by earlier peak arrivals. One possible explanation for this misfit is system measurement error. The temperature observation at PW4-3 was performed as follows: The Levellogger that measures both temperature and pressure was placed at the middle of the 1-m screen in PW4-3, and a foot valve of the water hose used for pumping was situated right above the upper boundary of the screen (50 cm above the Levellogger). Based on this measurement configuration, an increase in thermal conduction due to intraborehole convection may have occurred, such that the temperature readings from the Levellogger did not accurately represent the actual groundwater temperature before the groundwater reached the screen. Otherwise, the minor observed temperature decrease indicates that the pump in PW4-3 drew more background water than expected in the simulations during the tracer test, suggesting that a highly conductive lens probably exists in the southwest of PW4-3 (exterior to the data domain), which was not accurately mapped via HT analysis or other borehole testing methods.

7.6.2 The Advection-Dispersion Transport Model and Its Parameters

The classical advection-dispersion equation was used in this study to numerically simulate the heat transport behavior for this highly heterogeneous aquifer system. The spatially heterogeneous *K* and porosity estimates were determined based on the analyses of field measurements. Meanwhile, the heat tracer data were not used to calibrate the model parameters (e.g., porosity distribution) for optimizing the temperature fit. By comparing the simulated temperature BTCs with field observations in the dense 3D monitoring network, the optimal case of numerical simulation (*HT_8*) can reasonably reproduce the spatial distribution and temporal evolution of the heat tracer plume.

In addition to the reliable *K* estimation via HT analysis, accurately representing the porosity in 3D is also beneficial to transport predictions. It has been rare to obtain in situ measurements of effective porosity. In previous studies, porosity was usually determined on the basis of expert knowledge of the soil or calibration of tracer data. Hence, the most favorable porosity value was assigned to the medium,

ensuring the optimal model fit with observed BTCs. For example, in the transport studies at the MADE site, the value of total porosity was usually assigned between 0.30 and 0.35 (e.g., Bianchi et al., 2011a; Bianchi & Zheng, 2016; Feehley et al., 2000; Liu et al., 2010; Ronayne et al., 2010). Furthermore, effective porosity varies significantly with total porosity for finer-grained materials. Ronayne et al. (2010) assigned 75% of the total porosity to the mobile domain, yielding a mobile-domain porosity of 0.225. For the same tracer test ("MADE-4"), the mobile-domain porosity was assigned to be 0.044 by Liu et al. (2010). Feehley et al. (2000) suggested that a mobile-domain porosity should be 1/8 of the total porosity.

In this study, NMR logging yielded high-resolution total and effective porosities along eight boreholes. Meanwhile, the soft information from the borehole geology logs was used to delineate layer boundaries (19-layer geological zonation model), enabling the construction of the 3D effective porosity distribution. The effective porosity field (see Figure 7.6b) exhibited highly contrasting patterns between aquifer and aquitard layers. The ratio between effective and total porosity (refer to Table D4) can vary from 0.8% in the Clay layer to 84.6% in the Sand and gravel layer (Aquifer 2). Based on the observations and results of our study, the spatial heterogeneity of effective porosity should be accounted for in the numerical simulation of tracer transport, especially for highly heterogeneous porous media with a wide range of soil grain sizes.

Compared to the total porosity field (the "Total porosity field" scenario), the simulations using the effective porosity field can better reproduce the magnitude of the peak of observed temperature decreases at more observation ports (see Figure 7.9). However, the simulation using the effective porosity field resulted in earlier peak arrivals, as indicated by the computed mean arrival time (μ).

Uncertainty existed in the 3D distribution of effective porosity, which can be attributed to two key aspects. First, the NMR-derived effective porosity was determined by distinguishing the mobile fluid volume of the relaxation time distribution from the immobile fluid volume. A cutoff time was used to partition the relaxation time distribution, yielding the proportion of mobile-domain porosity to total

porosity. A smaller cutoff time can lead to a larger estimate of effective porosity. A value of 33 ms was employed in this study, which is a typical cutoff time for saturated sandstones and widely used for unconsolidated materials (e.g., Dlubac et al., 2013; Kendrick et al., 2023; Walsh et al., 2013). Behroozmand et al. (2015) noted that no appropriate cutoff time was defined for unconsolidated sediments.

Another source of uncertainty came from the spatial correlation of borehole porosity estimates using the zonation model. In this study, the layering of NMR-derived effective porosity was not very consistent with the zone geometry from the geological zonation model. This also resulted in an overall underestimated effective porosity for Aquifer 2, as lower effective porosity values were also incorporated in the zonal effective porosity estimate for Aquifer 2. The lower the estimated effective porosity, the higher the linear velocity of groundwater is, resulting in an earlier arrival of the simulated BTCs. In summary, a refined effective porosity field is required to improve the temperature fits with field observations, which entails both optimized borehole effective porosity estimates and more accurate mapping of its spatial variability.

The impacts of the mechanical dispersion and thermal conduction processes on transport behavior were also investigated. Based on the simulation results, the M_0, μ, and σ^2 were more consistent with field observations when the dispersivity components were removed from the simulation. On the contrary, larger misfits were obtained when a more significant mechanical dispersion process was included in the simulation. The results in this study suggest that the role of dispersion can be reduced after K heterogeneity is accurately mapped via HT analysis. Additionally, the thermal conduction process can be favorably represented in numerical simulations when an estimate of thermal conductivity within a reasonable range is provided. In summary, the heat transport behavior for this porous medium was dominated by the advective transport process. Therefore, accurately mapping the spatial heterogeneity of

K and effective porosity is key to successfully predicting heat transport in highly heterogeneous aquifer systems such as the NCRS.

8 Conclusions and Recommendations

8.1 Summary and Conclusions

Study I: *Hydraulic Tomography Estimates Improved by Zonal Information from the Clustering of Geophysical Survey Data*

In this study, we investigated how to integrate geophysical survey data into hydraulic tomography (HT) surveys by constructing clustering-based zonation models as the initial model of the geostatistical inversion of HT analysis. In addition, a new type of spatial constraints was proposed to remove outliers and to include geological knowledge for creating improved zonation models. Disparate borehole geophysical survey logs were incorporated into clustering, including electrical conductivity (EC) and gamma ray (GR) logs, to investigate the progressive improvements to HT analysis performance through the integration of non-redundant geophysical information. Meanwhile, the usefulness of K measurements from local-scale hydraulic tests in HT analyses integrated with geophysical survey was also investigated through comparative case studies. Initially, the general framework of using clustering-based zonation models with spatial constraints to improve HT estimates was introduced. After generating the zone geometry from clustering of ln K profiles along six columns, clustering validity with spatial constraints was checked by comparing external Clustering Validity Indices (CVIs) of different scenarios. A numerical sandbox was generated based on a real laboratory sandbox, in which groundwater flow and solute transport simulations were conducted to highlight the performance of HT integrated with geophysical survey logs. Four pumping tests were performed with 31 observation ports utilized to observe constant drawdowns under steady-state flow conditions, and observation data were provided fairly to all cases to calibrate the K field through geostatistical inverse modeling. Comparative cases were designed in the sandbox experiment, in which different initial models of K field were provided in the inversion,

including: (1) a homogeneous *K* field; (2) zonation models built by the clustering of disparate geophysical surveys with/without spatial constraints; and (3) zonation models improved by incorporating local *K* measurements. The performance of HT analyses for all cases were assessed by comparison of simulated and true ln *K* fields, model validation by analyzing drawdown responses from independent pumping tests not included in the model calibration process, and predictions of tracer migration patterns through two forced-gradient tracer injection tests. Our study resulted in the following major findings and conclusions:

(1) Clustering with the proposed spatial constraints is a feasible approach to improve the zonation performance in hydrostratigraphic characterization. When spatial constraints are incorporated into clustering, outliers are significantly removed, and only adjacent logging data with similar geophysical properties are assigned to the same cluster due to the existence of increasing spatial penalty with farther distance. It indicates that the extraction of zone geometry by clustering with spatial constraints can better identify hydrostratigraphy in layered porous medium.

(2) The resolution of zone geometry and the estimated zonal *K* values in zonation models are progressively improved when spatial constraints, disparate geophysical survey logs and local *K* measurements are successively involved in cluster analysis. Hydrostratigraphic zones may be erroneously interpreted if only EC logs are involved in clustering, and this situation can be substantially alleviated when spatial constraints are involved. Non-redundant geophysical information from disparate surveys can significantly improve the delineation of layer structures. In addition, the inclusion of local *K* measurements can yield better zonal *K* estimates.

(3) All cases with/without initial models can be calibrated well via the SimSLE code. Dominant high-*K* and low-*K* regions of the aquifer-aquitard system are generally mapped in the inversion initialized with a homogeneous *K* field although their locations are slightly deviated from true hydrostratigraphy. The resolution of the *K* tomogram is significantly improved with the use of zonal information as an initial guess in the inverse model, and reliable zonation models can

further increase the mapping ability of aquifer characteristics. These results suggest that the fusion of hydraulic and geophysical characterization techniques can yield more accurate *K* tomograms.

(4) In terms of the quantitative comparison of spatial *K* distribution, the accuracy of HT estimates can be effectively improved by integrating geophysical survey data using the clustering approach. Besides, significant improvements to subsurface heterogeneity characterization are obtained in HT analysis initialized with spatially-constrained zonation models when local *K* measurements are involved, but only marginally improvements are found when no spatial constraints are included.

(5) Different from the results of ln *K* comparison, HT initialized with zonation models from the clustering without spatial constraints cannot improve drawdown prediction performance. Inaccurate zonal information incorporated into geostatistical inversion leads to adverse HT performance. This finding corroborates a conclusion from Zhao et al. (2016). Nonetheless, HT analyses involved with spatial constraints in clustering can better predict drawdown responses.

(6) The significance of integrating geophysical information into HT analysis is further highlighted through tracer tests to predict tracer migration patterns. In terms of results from temporal moment analysis, integration of geophysical survey logs leads to better performance of HT estimates on solute transport predictions, and the involvement of local *K* measurements further improve prediction performance.

(7) In an overall consideration of ln *K* field comparison, assessment of drawdown prediction performance, and prediction of tracer migration, the effectiveness of integrating geophysical surveys to improve HT analysis by clustering with spatial constraints is sufficiently demonstrated. Furthermore, incorporating local *K* measurements can yield better interpretation on zonation models, which further leads to improved HT estimates. It again illustrates the importance of

fusing different non-redundant data sets into HT analysis for improved subsurface heterogeneity characterization.

Study II: *Comparison of NMR-Derived Hydraulic Conductivity with Various Hydraulic Testing Methods*

In this study, we evaluated the performance of NMR petrophysical models to estimate hydraulic conductivity (*K*) in a highly heterogeneous glaciofluvial aquifer system. Hence, we first calibrated NMR-derived *K* profiles at four well locations with permeameter test data. The optimized model constants suitable for glaciofluvial deposits were listed. To highlight the performance of NMR logging, we compared the NMR-derived *K* with various hydraulic testing methods, including a) permeameter tests that provided accurate *K* measurements in the laboratory, b) multi-level slug tests that sampled a larger extent local to the screen locations, and c) direct-push hydraulic profiling tool (HPT) logging tests, which is an in-situ physical method to measure *K* at a dense sampling interval. We evaluated the four popular NMR models, including Schlumberger-Doll Research (SDR), Seevers, Sum-of-echoes (SOE), and Kozeny-Godefroy (KGM), to estimate *K* by quantifying the model fits with *K* measurements from permeameter and slug tests. Our study resulted in the following findings and conclusions:

(1) NMR logging can yield *K* estimates over five orders of magnitude for the glaciofluvial deposits at the NCRS, which consist of gravel, sand, silt, and clay. The main aquifer characteristics for this complex layered porous medium can be reproduced through the cross-hole comparison of four NMR-derived *K* profiles, including the shape and volume of alternating aquifers and aquitards with their geometric mean.

(2) Site characterization performance of NMR logging is highlighted after comparing it with various hydraulic testing methods featured on their respective advantages. Meanwhile, NMR logging can avoid the shortcomings of these testing methods found in this study. Specifically, NMR logging

yields comparable *K* estimates to permeameter-based *K*, and its high-resolution delineation of hydrostratigraphy is similar to HPT logging. After upscaling, the NMR-derived *K* estimates can generally account for the slug-based *K* estimates at a larger scale.

(3) Through quantitative comparison with permeameter and slug tests, all four petrophysical models can produce reliable *K* estimates. The SDR model always yields slightly better model fits than the Seevers model, and including porosity does not improve the predictive performance of the SDR model. The SDR constant is demonstrated to be stable across different sites for glacial deposits. After adopting NMR signals with better noise control, the SOE model yields slightly improved fits than the SDR model. The KGM model leads the predictive performance, especially on the model fits with slug tests, which may contribute to the calibration of surface relaxivity in this physically derived NMR estimator.

Based on these findings, we provide the following recommendations for selecting suitable NMR models to estimate *K* in glacial deposits: a) when there is limited information about lithology, the SDR model without porosity can be sufficient for basic site characterization; b) when the background noise is high, the SOE model is recommended given the improved noise control of the SOE signal; c) the KGM model can provide more optimal *K* estimates if hydraulic calibration points are available at the site.

Study III: *Characterizing Spatial Heterogeneity of Hydraulic Conductivity using Borehole NMR in a Complex Groundwater Flow System*

Borehole NMR logging was limited to estimating *K* along 1D vertical profiles. Here, we extended its use to determine 3D *K* distributions in a highly heterogeneous, glaciofluvial deposit. We employed geostatistical interpolation methods (ordinary kriging and indicator kriging), zonal representation methods (visually-interpreted geological zonation and spatially associated zonation with clustering), and a multi-level heterogeneity characterization method to model inter- and intralayer heterogeneous *K* fields.

Through numerically simulating the drawdown responses in ten pumping tests, a multi-level heterogeneity model provided the most optimal representation of interlayer and intralayer heterogeneity as well as their connectivity. Our results demonstrate the effectiveness of using borehole NMR data to reproduce high-resolution aquifer characteristics suitable for simulations of hydraulic responses to pumping tests in a complex glaciofluvial environment. Our study led to the following conclusions:

(1) The spatial *K* models derived using kriging estimation methods can recognize alternating aquifer and aquitard units. However, our study reveals that *K* connectivity between boreholes cannot be faithfully represented for this glaciofluvial deposit. These biased *K* representations lead to overpredicted drawdowns in groundwater flow simulations of multiple pumping tests. Additionally, the predictive capability of the groundwater flow models based on kriged *K* fields relies heavily on the number of NMR logs.

(2) Compared with the kriging interpolation models, the spatially associated zonation model improves the correspondence of model-derived *K* to permeameter measurements and drawdown predictions of multiple pumping tests. Compared with the visually interpreted aquifer zones from borehole geologic logs, the zonation based on clustering of NMR logs can correlate aquifer zones more efficiently and delineate layer boundaries more consistently with realistic geologic layering.

(3) Incorporating zonally kriged *K* into the spatially associated zonation model can further improve the spatial heterogeneity characterization. The *K* connectivity is refined, especially in mapping the gradual or crisp boundaries between aquifer and aquitard layers. Therefore, the multi-level heterogeneity model yields the optimal drawdown prediction performance for this layered aquifer system. Moreover, its drawdown predictive capability remains stable with a limited NMR dataset.

(4) The drawdown predictions based on NMR logging outperform those on permeameter measurements, though less data is used. This may contribute to in-situ logging procedures and

more precise *K* estimation using borehole NMR, thereby mapping the spatial *K* variability more representative of the layered glaciofluvial deposits at the NCRS.

Overall, NMR logging shows great potential in characterizing the spatial heterogeneity of *K*. On the one hand, NMR logging yields *K* profiles along boreholes that provide a wide range of *K* estimates, accounting for soil types from gravel/sand to silt and clay for the highly heterogeneous glaciofluvial deposits. On the other hand, NMR logging characterizes the subsurface at a high resolution, facilitating the spatial correlation of borehole *K* estimates. These factors contribute to the successful prediction of hydraulic responses to multiple pumping tests. However, additional studies are needed to investigate the model utility in other hydrogeologic environments.

Study IV: *Integrating NMR Logging with Three-Dimensional Steady-State Hydraulic Tomography for Improved Mapping of Hydraulic Conductivity Distribution*

In this study, we investigated whether integrating nuclear magnetic resonance (NMR) logging with hydraulic tomography (HT) analysis can enhance the mapping of hydraulic conductivity (*K*) heterogeneity, thereby facilitating more accurate predictions of drawdown responses to pumping tests in a complex groundwater flow system. The effectiveness of integrated models was evaluated by comparing them with NMR models, which were based on the spatial correlation of borehole NMR *K* estimates, and the HT models, which were solely based on the inverse modeling of pumping/injection test data. In addition to evaluating the drawdown simulations of the groundwater flow models based on these *K* fields, this study also compared the *K* profiles extracted from these spatial *K* models with co-located permeameter measurements. Furthermore, previous studies suggested that smooth tomograms may be derived when insufficient pumping data were used for HT analysis, resulting in aquifer characteristics not being faithfully reproduced. Thus, this study examined whether the integration of NMR logging can

improve the resolution and accuracy of HT results using a limited calibration dataset, where both the number of pumping tests and observation wells were reduced. The results and observations of this study led to the following conclusions:

1) When sufficient head observations are incorporated in HT analysis, the main characteristics of this aquifer system, including the alternating layers of aquifers and aquitards and their *K* magnitudes, can be reasonably mapped. Meanwhile, the hydraulic connectivities can be adequately described, yielding reliable drawdown prediction performance for different pumping tests. However, when an insufficient number of pumping tests and a lower observation interval density are used in HT analysis, a smooth tomogram may be derived, which is incapable of delineating the hydrostratigraphy for this highly heterogeneous glaciofluvial aquifer system. Still, reasonable drawdown predictions can be obtained.

2) The NMR models using different spatial interpolation methods, including ordinary kriging, clustering-based zonation, and the hybrid multi-level characterization, can be successfully calibrated via HT analysis. The patterns of spatial *K* variability in these integrated models resemble those in their initial NMR models with refined hydraulic connections, leading to significantly improved drawdown predictions, particularly for the pumping tests in regions where NMR logging encounters higher measurement noise. Additionally, the joint use of HT and NMR logging data can yield enhanced drawdown prediction performance and characterize *K* heterogeneity at a higher resolution, compared to using HT analysis relying on pumping/injection test data alone.

3) After calibration via HT analysis with sufficient observation data, the calibrated *K* fields of different NMR models show high degrees of consistency in the *K* magnitudes and spatial variability. Furthermore, the integrated models initialized with different NMR *K* models can yield comparable drawdown predictions, though the drawdown predictive capability varies for these

initial *K* models. Despite similar overall drawdown predictions, the multi-level heterogeneity model can produce more stable drawdown predictions for each pumping test, and vertical *K* variations in *K* tomograms are more consistent with the layered structure of this glaciofluvial aquifer system.

4) The initial NMR models can still be refined when a limited HT dataset is used for model calibration. The heterogeneity patterns of these integrated *K* models resemble those obtained by calibrating a full HT dataset. In addition, they can provide substantially improved drawdown prediction performance compared to their initial NMR models, and the drawdown predictions of the optimal case even slightly exceed those using the HT model with a full calibration dataset. Compared to HT analysis with more sufficient head observations, the integration of NMR logging with the HT using a limited calibration dataset can substantially improve the mapping of downhole *K* variations, as observed in comparisons with permeameter measurements.

HT analysis through geostatistical inverse modeling is a powerful means of depicting spatial *K* heterogeneity, provided that a dedicated monitoring network can be installed; however, this may not always be feasible for use in the environmental and industrial sectors. By incorporating reliable hydrostratigraphic information and borehole *K* estimates from NMR logging, *K* tomograms with limited head observations can be more representative of highly heterogeneous aquifer systems. Moreover, the accuracy and resolution of *K* tomograms can be further refined when more pumping tests with additional head observations are available for model calibration, making the joint use of NMR logging and HT analysis suitable for a wide variety of subsurface characterization tasks.

Study V: *Predictions of Heat Transport Behaviors in a Complex Groundwater Flow System*

In this study, the ability of various approaches (e.g., Hydraulic Tomography (HT) analysis) to accurately characterize hydraulic conductivity (*K*) heterogeneity was investigated by predicting heat transport behavior at high resolution. Accordingly, a heat tracer test was conducted in a highly heterogeneous glaciofluvial deposit, in which chilled water was injected into the subsurface to yield a temperature difference with a reduced density effect. The highly non-uniform spreading of the heat tracer plume was captured using a dense temperature monitoring network, which recorded depth-specific temperature decreases at different observation locations. Transport predictions were performed through numerical simulations using the classical advection-dispersion equation. Different spatial conceptualization approaches and data sources were used to yield *K* representations for transport predictions. Meanwhile, NMR-derived effective porosity was used to derive a 3D porosity field. The temporal moment analysis of the temperature breakthrough curves (BTCs) was employed to quantitatively compare various model simulations with field observations. Lastly, the sensitivities of heat tracer plume migration to transport and thermal parameters were examined. This study led to the following findings and conclusions:

1) A homogeneous *K* representation is not capable of representing the irregular distribution of the migrating plume and the highly asymmetric shapes of the temperature BTCs for this complex glaciofluvial aquifer system. When heterogeneous *K* fields are described using kriging and zonation models, the preferential migration of the heat tracer plume can be reproduced. However, these kriging and zonation models fail to faithfully represent the spatial distribution of the observed temperature decreases. This may be attributed to the borehole hydraulic testing methods, which cannot provide accurate *K* variations at high resolution, or to the fact that inter-well *K* heterogeneity is not effectively mapped based on the spatial correlation of borehole data.

2) The transport simulations using HT analysis for K estimation can adequately represent the spatial distribution of temperature responses, even in cases with a limited HT calibration dataset. Specifically, the improved mapping of connected high-K networks via HT analysis enables the accurate tracing of heat tracer plume migration and spreading in both horizontal and vertical directions. When more sufficient head data are used in HT analysis, its simulations can reasonably reproduce the primary characteristics of the observed BTCs at different well locations and depths.

3) The NMR-derived effective porosity estimates vary significantly for aquifer and aquitard layers. Therefore, it is imperative to estimate effective porosity in 3D for predictions of tracer migration in highly heterogeneous aquifer systems. The simulations using the effective porosity field yield more accurate predictions of the maximum temperature decreases than using the total porosity field. However, the effective porosity field using a geologically based zonal representation of uncalibrated effective porosity from NMR logging resulted in a higher migration rate compared to field observations. Therefore, the calibration of effective porosity estimates with tracer data can next be considered to further improve heat transport predictions.

4) Through additional numerical simulations with varying dispersivities and thermal conductivities, it is found that dispersivity is not a very sensitive parameter that controls heat transport behavior in this groundwater flow system, provided that K estimates and their spatial variability are accurately described. Additionally, the determination of thermal conductivity within a reasonable range yields comparable transport predictions. In summary, tracer migration is most sensitive to K heterogeneity, second most sensitive to effective porosity, and only mildly sensitive to dispersivities and thermal conductivity.

Based on these findings, the highly non-uniform spatiotemporal evolution of temperature BTCs can be reasonably reproduced using the classical advection-dispersion equation, provided that reliable

estimates of heterogeneous *K* (e.g., from HT analysis) and porosity fields (e.g., from NMR logging) are available. Here, transport models using HT analysis for *K* estimation yield the most optimal prediction performance, demonstrating that HT can successfully map the spatial *K* variability for predicting transport behavior in highly heterogeneous groundwater flow systems. Therefore, a paradigm shift is required for improved transport predictions in complex unconsolidated aquifer systems. Instead of collecting a larger number of borehole *K* profiles that are typically used in the environmental and engineering sectors to characterize *K* heterogeneity, HT analysis is a more powerful and cost-effective technique for predicting contaminant transport.

8.2 Recommendations for Future Studies

Through four independent studies, this thesis demonstrated that integrating geophysical logging data with HT analysis can yield improved representations of *K* heterogeneity, which can refine drawdown predictions in groundwater flow models through both synthetic sandbox and field HT studies. In addition, integrating geophysical logging with HT can also enhance predictions of solute transport in the synthetic sandbox study. Through simulating the field observations of a heat tracer test, this thesis demonstrates that advective transport models using HT analysis and NMR logging for *K* and porosity estimation can reasonably predict the transport behavior of a heat tracer test conducted in a complex groundwater system. Based on these completed studies, workable recommendations for future studies are provided:

(1) Future characterizations of layered porous media based on borehole testing procedures can incorporate multiple non-co-located datasets to improve the accuracy of layer delineation. Here, numerous datasets of geophysical parameters or borehole *K* measurements have been collected at the NCRS, including electrical conductivity (EC), gamma ray (GR) radiation, permeameter measurements, HPT- and NMR-derived *K* data. They were only partially co-located with different locations and varying depth ranges. Different petrophysical relationships exist between geophysical parameters and *K*. Meanwhile, the reliability of various testing techniques is affected

by local mineralogy or hydrogeological conditions. For example, the standard parameters of NMR petrophysical models may not be applicable to regions with a high content of iron oxides. Therefore, the collective use of disparate geophysical logging or borehole hydraulic testing methods can facilitate a more reliable characterization of layer structures for determining spatial *K* models at high resolution.

(2) Future HT studies on transport simulations can investigate the efficacy of HT analysis with an optimal monitoring network to predict transport behaviors in highly heterogeneous aquifer systems. Yeh and Liu (2000) provided design criteria for the HT survey, including the optimal monitoring network and pumping intervals, in which the performances of comparative designs were evaluated by comparing their drawdown responses. In Study IV, two HT datasets with varying numbers of observed heads were used to interpret *K* heterogeneity for transport simulations. However, none of them followed an optimal design of the monitoring network. In addition, the spacing of monitoring intervals in the CMT wells (i.e., 2 m) remained constant in this field HT survey. Therefore, a future synthetic study is suggested to investigate the optimal transport prediction performance using HT analysis, as well as the transport predictions with different designs of monitoring and pumping intervals when gradually altering the available data for model calibration. This future study may guide the practical designs of the HT survey for transport predictions in the industry.

(3) Future studies can focus on determining optimal porosity fields for improved transport predictions. In Study IV, discrepancies were found between the pore-scale effective porosity estimated from NMR logging and the Darcy-scale effective porosity for transport simulations. Therefore, tracer test data can be used in a future study to calibrate the effective porosity field. The NMR-derived effective porosity is determined by distinguishing the mobile fluid volume of the relaxation time distribution from the immobile fluid volume, in which a cutoff time is used (typically 33 ms for unconsolidated materials). The optimized effective porosity can contribute to

determining a site-specific cutoff time for effective porosity estimation, which can be suggested for use in transport studies within similar sedimentations (e.g., glacial deposits).

(4) A key future research topic is the joint inversion of tracer and head data to enhance the mapping of K heterogeneity. The geostatistical inversion of head data is typically underdetermined, as multiple K realizations can yield simulated heads that are consistent with the observed data. The joint inversion of head and tracer (e.g., concentration, temperature, tracer arrival times) can effectively reduce the nonuniqueness of calibrated K fields. However, new algorithms are required for the joint inversion of tracer and head data.

References

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

Anderson, M. P. (2005). Heat as a ground water tracer. *Ground Water*, *43*(6), 951–968. https://doi.org/10.1111/j.1745-6584.2005.00052.x

Aquanty Inc. (2025). HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. Waterloo. ON.

Archie, G. E. (1942). The electrical resistivity log as an aid in determining some reservoir characteristics. *Petroleum Technology*, 54–62. https://doi.org/10.2118/942054-g

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. https://doi.org/10.1145/1283383.1283494

Behroozmand, A. A., Keating, K., & Auken, E. (2015). A review of the principles and applications of the NMR technique for near-surface characterization. *Surveys in Geophysics*, *36*(1), 27–85. https://doi.org/10.1007/s10712-014-9304-0

Berg, S. J., & Illman, W. A. (2011a). Capturing aquifer heterogeneity: Comparison of approaches through controlled sandbox experiments. *Water Resources Research*, *47*, W09514. 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*, W10507. 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. *Groundwater*, *53*(1), 71–89. https://doi.org/10.1111/gwat.12159

Bianchi, M., & Zheng, C. (2016). A lithofacies approach for modeling non-Fickian solute transport in a heterogeneous alluvial aquifer. *Water Resources Research*, *52*, 552–565. https://doi.org/10.1002/2015WR018186

Bianchi, M., Zheng, C., Tick, G. R., & Gorelick, S. M. (2011a). Investigation of small-scale preferential flow with a forced-gradient tracer test. *Groundwater*, *49*(4), 503–514. https://doi.org/10.1111/j.1745-6584.2010.00746.x

Bianchi, M., Zheng, C., Wilson, C., Tick, G. R., Liu, G., & Gorelick, S. M. (2011b). Spatial connectivity in a highly heterogeneous aquifer: From cores to preferential flow paths. *Water Resources Research*, *47*(5), 1–18. https://doi.org/10.1029/2009WR008966

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., Butler, J. J., Zhan, X., & Knoll, M. D. (2007). A field assessment of the value of steady shape hydraulic tomography for characterization of aquifer heterogeneities. *Water Resources Research*, *43*(5), 1–23. https://doi.org/10.1029/2006WR004932

Borden, R. C., Cha, K. Y., & Liu, G. (2021). A Physically Based Approach for Estimating Hydraulic Conductivity from HPT Pressure and Flowrate. *Groundwater*, *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

Brownstein, K. R., & Tarr, C. E. (1979). Importance of classical diffusion in NMR studies of water in biological cells. *Physical Review A*, *19*(6), 2446–2453. https://doi.org/10.1103/PhysRevA.19.2446

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

Cardiff, M., & Barrash, W. (2011). 3-D transient hydraulic tomography in unconfined aquifers with fast drainage response. *Water Resources Research*, *47*(12). https://doi.org/10.1029/2010WR010367

Cardiff, M., Barrash, W., & Kitanidis, P. K. (2013). Hydraulic conductivity imaging from 3-D transient hydraulic tomography at several pumping/observation densities. *Water Resources Research*, *49*(11), 7311–7326. https://doi.org/10.1002/wrcr.20519

Carle, S. F., Labolle, E. M., Weissmann, G. S., Brocklin, D. Van, & Fogg, G. E. (1998). Conditional Simulation of Hydrofacies Architecture. In *Hydrogeologic Models of Sedimentary Aquifers, Concepts in Hydrogeoly and Environmental Geology* (Issue 1, pp. 147–170). https://doi.org/10.2110/sepmcheg.01.147

Carr, J. C., Beatson, R. K., Mitchell, T. J., & Evans, T. R. (2001). Reconstruction and Representation of 3D Objects with Radial Basis Functions. *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, *August*, 67–76. https://doi.org/10.1145/383259.383266

Coates, G. R., & Dumanoir, J. L. (1973). A new approach to improved log-derived permeability. *SPWLA 14th Annual Logging Symposium*.

Coates, G. R., Peveraro, R. C. A., Hardwick, A., & Roberts, D. (1991). The Magnetic Resonance Imaging Log Characterized by Comparison with Petrophysical Properties and Laboratory Core Data. *SPE 66th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers*, *9*.

Crow, H., Paradis, D., Grunewald, E., Liang, X. X., & Russell, H. A. J. (2022). Hydraulic Conductivity from Nuclear Magnetic Resonance Logs in Sediments with Elevated Magnetic Susceptibilities. *Groundwater*, *60*(3), 377–392. https://doi.org/10.1111/gwat.13158

Dagan, G. (1982). Stochastic modeling of groundwater flow by unconditional and conditional probabilities 1. Conditional simulation and the direct problem. *Water Resources Research*, *18*(4), 813–833. https://doi.org/10.1029/WR018i004p00813

Dagan, G. (1984). Solute transport in heterogeneous porous formations. *Journal of Fluid Mechanics*, *145*(2), 151–177. https://doi.org/10.1017/S0022112084002858

Daranond, K., Yeh, T. C. J., Hao, Y., Wen, J. C., & Wang, W. (2020). Identification of groundwater basin shape and boundary using hydraulic tomography. *Journal of Hydrology*, *588*, 125099. https://doi.org/10.1016/j.jhydrol.2020.125099

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

Deutsch, C. V., & Journel, A. G. (1998). GSLIB: geostatistical software library and user's guide. Second edition. In *Oxford Univeristy Press*.

Díaz-Curiel, J., Miguel, M. J., Biosca, B., & Arévalo-Lomas, L. (2021). Gamma ray log to estimate clay content in the layers of water boreholes. *Journal of Applied Geophysics*, *195*. https://doi.org/10.1016/j.jappgeo.2021.104481

Dietrich, P., & Tronicke, J. (2009). Integrated analysis and interpretation of cross-hole P- And S-wave tomograms: A case study. *Near Surface Geophysics*, *7*(2), 101–109. https://doi.org/10.3997/1873-0604.2008041

Dlubac, K., Knight, R., Song, Y. Q., Bachman, N., Grau, B., Cannia, J., & Williams, J. (2013). Use of NMR logging to obtain estimates of hydraulic conductivity in the High Plains aquifer, Nebraska,

USA. *Water Resources Research*, *49*(4), 1871–1886. https://doi.org/10.1002/wrcr.20151

Dlugosch, R., Günther, T., Müller-Petke, M., & Yaramanci, U. (2013). Improved prediction of hydraulic conductivity for coarse-grained, unconsolidated material from nuclear magnetic resonance. *Geophysics*, *78*(4), EN55–EN64. https://doi.org/10.1190/GEO2012-0187.1

Doetsch, J., Linde, N., Coscia, I., Greenhalgh, S. A., & Green, A. G. (2010). Zonation for 3D aquifer characterization based on joint inversions of multimethod crosshole geophysical data. *Geophysics*, *75*(6). https://doi.org/10.1190/1.3496476

Dogan, M., Van Dam, R. L., Liu, G., Meerschaert, M. M., Butler, J. J., Bohling, G. C., Benson, D. A., & Hyndman, D. W. (2014). Predicting flow and transport in highly heterogeneous alluvial aquifers. *Geophysical Research Letters*, *41*(21), 7560–7565. https://doi.org/10.1002/2014GL061800

Doherty, J. (2015). *Calibration and uncertainty analysis for complex environmental models*. Watermark Numerical Computing.

Doro, K. O., Cirpka, O. A., & Leven, C. (2015). Tracer Tomography: Design Concepts and Field Experiments Using Heat as a Tracer. *Groundwater*, *53*(S1), 139–148. https://doi.org/10.1111/gwat.12299

Doveton, J. H. (1986). Log analysis of subsurface geology: concepts and computer methods. In *Wiley*.

Dumont, M., Reninger, P. A., Pryet, A., Martelet, G., Aunay, B., & Join, J. L. (2018). Agglomerative hierarchical clustering of airborne electromagnetic data for multi-scale geological studies. *Journal of Applied Geophysics*, *157*, 1–9. https://doi.org/10.1016/j.jappgeo.2018.06.020

Estivill-Castro, V. (2002). Why so many clustering algorithms. *ACM SIGKDD Explorations Newsletter*, *4*(1), 65–75. https://doi.org/10.1145/568574.568575

ETB (2025). *The engineering toolbox. Solids: specific heat capacities.*

https://www.engineeringtoolbox.com/specific-heat-solids-d_154.html (accessed 13 July 2025)

Feehley, C. E., Zheng, C., & Molz, F. J. (2000). A dual-domain mass transfer approach for modeling solute transport in heterogeneous aquifers: Application to the macrodispersion experiment (MADE) site. *Water Resources Research*, *36*(9), 2501–2515. https://doi.org/10.1029/2000WR900148

Fowlkes, E. B., & Mallows, C. L. (1983). A Method for Comparing Two Hierarchical Clusterings. *Journal of the American Statistical Association*, *78*(383), 553–569.

Garabedian, S. P., Leblanc, D. R., Gelhar, L. W., & Celia, M. A. (1991). Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 2. Analysis of spatial moments for a nonreactive tracer. *Water Resources Research*, *27*(5), 911–924. https://doi.org/10.1029/91WR00242

Gelhar, L. W., & Axness, C. L. (1983). Three-dimensional stochastic analysis of macrodispersion in aquifers. *Water Resources Research*, *19*(1), 161–180. https://doi.org/10.1029/WR019i001p00161

Giambastiani, B. M. S., Colombani, N., & Mastrocicco, M. (2013). Limitation of using heat as a groundwater tracer to define aquifer properties: Experiment in a large tank model. *Environmental Earth Sciences*, *70*(2), 719–728. https://doi.org/10.1007/s12665-012-2157-2

Gottlieb, J., & Dietrich, P. (1995). Identification of the permeability distribution in soil by hydraulic tomography. *Inverse Problems*, *11*(2), 353–360. https://doi.org/10.1088/0266-5611/11/2/005

Gueting, N., Klotzsche, A., van der Kruk, J., Vanderborght, J., Vereecken, H., & Englert, A. (2015). Imaging and characterization of facies heterogeneity in an alluvial aquifer using GPR full-waveform inversion and cone penetration tests. *Journal of Hydrology*, *524*, 680–695. https://doi.org/10.1016/j.jhydrol.2015.03.030

Gutjahr, A., & Wilson, J. (1989). Co-kriging for stochastic flow models. *Transport in Porous Media*, *4*(6), 585–598. https://doi.org/10.1007/BF00223629

Hantush, M. S. (1960). Modification of the theory of leaky aquifers. *Journal of Geophysical Research*, *65*(11), 3713–3725. https://doi.org/10.1029/jz065i011p03713

Harvey, C. F., & Gorelick, S. M. (1995). Mapping hydraulic conductivity: Sequential conditioning with measurements of solute arrival Time, hydraulic Head, and local conductivity. *Water Resources Research*, *31*(7), 1615–1626. https://doi.org/10.1029/95WR00547

Harvey, C., & Gorelick, S. M. (2000). Rate-limited mass transfer or macrodispersion: Which dominates plume evolution at the Macrodispersion Experiment (MADE) site? *Water Resources Research*, *36*(3), 637–650.

Hazen, A. (1892). Some physical properties of sands and gravels, with special reference to their use in filtration. In *Massachusetts State Board of Health, 24th Annual Report*. https://doi.org/10.4159/harvard.9780674600485.c25

Hochstetler, D. L., Barrash, W., Leven, C., Cardiff, M., Chidichimo, F., & Kitanidis, P. K. (2016). Hydraulic tomography: Continuity and discontinuity of high-K and low-K zones. *Groundwater*, *54*(2), 171–185. https://doi.org/10.1111/gwat.12344

Hoeksema, R. J., & Kitanidis, P. K. (1984). An Application of the Geostatistical Approach to the Inverse Problem in Two-Dimensional Groundwater Modeling. *Water Resources Research*, *20*(7), 1003–1020. https://doi.org/10.1029/WR020i007p01003

Hoeksema, R. J., & Kitanidis, P. K. (1989). Prediction of transmissivities, heads, and seepage velocities using mathematical modeling and geostatistics. *Advances in Water Resources*, *12*(2), 90–102. https://doi.org/10.1016/0309-1708(89)90007-9

Hou, X., Hu, R., Yeh, T. C. J., Li, Y., Qi, J., Song, Y., & Qiu, H. (2023). A Short-Term Pumping Strategy for Hydraulic Tomography Based on the Successive Linear Estimator. *Water Resources Research*, *59*(4). https://doi.org/10.1029/2022WR033831

Huang, S. Y., Wen, J. C., Yeh, T. C. J., Lu, W., Juan, H. L., Tseng, C. M., Lee, J. H., & Chang, K. C. (2011). Robustness of joint interpretation of sequential pumping tests: Numerical and field experiments. *Water Resources Research*, *47*(10), 1–18. https://doi.org/10.1029/2011WR010698

Hvorslev, M. J. (1951). Time lag and soil permeability in ground-water observations. In *Corps of Engineers, U.S.Army* (Issue 36).

Illman, W. A., Berg, S. J., & Yeh, T. C. J. (2012). Comparison of approaches for predicting solute transport: Sandbox experiments. *Groundwater*, *50*(3), 421–431. https://doi.org/10.1111/j.1745-6584.2011.00859.x

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

Illman, W. A., Liu, X., & Craig, A. (2007). Steady-state hydraulic tomography in a laboratory aquifer with deterministic heterogeneity: Multi-method and multiscale validation of hydraulic conductivity tomograms. *Journal of Hydrology*, *341*(3–4), 222–234. https://doi.org/10.1016/j.jhydrol.2007.05.011

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

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

Jaccard, P. (1912). The distribution of the Flora in the Alpine Zone. *New Phytologist*, *11*(2), 37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

Jia, R., Lv, Y., Wang, G., Carranza, E. J. M., Chen, Y., Wei, C., & Zhang, Z. (2021). A stacking methodology of machine learning for 3D geological modeling with geological-geophysical datasets, Laochang Sn camp, Gejiu (China). *Computers and Geosciences*, *151*(November 2020), 104754. https://doi.org/10.1016/j.cageo.2021.104754

Jimenez, S., Brauchler, R., Hu, R., S. Schmidt, T. Ptak, & Bayer, P. (2015). Predition of solute transport in a heterogeneous aquifer utilizing hydraulic conductivity and specific storage tomograms. *Water Resources Research*, *51*, 5504–5520. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Kawo, N. S., Korus, J., Kishawi, Y., Haacker, E. M. K., & Mittelstet, A. R. (2024). Three-dimensional probabilistic hydrofacies modeling using machine learning. *Water Resources Research*, *60*(7). https://doi.org/10.1029/2023WR035910

Kendrick, A. K., Knight, R., Johnson, C. D., Liu, G., Hart, D. J., Butler, J. J., & Hunt, R. J. (2023). Evaluation of models for estimating hydraulic conductivity in glacial aquifers with NMR logging. *Groundwater*, *61*(6), 778–792. https://doi.org/10.1111/gwat.13318

Kendrick, A. K., Knight, R., Johnson, C. D., Liu, G., Knobbe, S., Hunt, R. J., & Butler, J. J. (2021). Assessment of NMR Logging for Estimating Hydraulic Conductivity in Glacial Aquifers. *Groundwater*, *59*(1), 31–48. https://doi.org/10.1111/gwat.13014

Kenyon, W. E., Day, P. I., Straley, C., & Willemsen, J. F. (1988). Three-part study of NMR longitudinal relaxation properties of water-saturated sandstones. *SPE Formation Evaluation*, *3*(3), 622–636. https://doi.org/10.2118/15643-pa

Kitanidis, P. K. (1995). Quasi-linear geostatistical theory for inversing. *Water Resources Research*, *31*(10), 2411–2419. https://doi.org/10.1029/95WR01945

Kitanidis, P. K., & Vomvoris, E. G. (1983). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resources Research*,

19(3), 677–690. https://doi.org/10.1029/WR019i003p00677

Kleinberg, R. L., & Horsfield, M. A. (1990). Transverse relaxation processes in porous sedimentary rock. *Journal of Magnetic Resonance*, *88*(1), 9–19. https://doi.org/10.1016/0022-2364(90)90104-H

Klepikova, M., Brixel, B., & Jalali, M. (2020). Transient hydraulic tomography approach to characterize main flowpaths and their connectivity in fractured media. *Advances in Water Resources*, *136*, 103500. https://doi.org/10.1016/j.advwatres.2019.103500

Klepikova, M., Wildemeersch, S., Hermans, T., Jamin, P., Orban, P., Nguyen, F., Brouyère, S., & Dassargues, A. (2016). Heat tracer test in an alluvial aquifer: Field experiment and inverse modelling. *Journal of Hydrology*, *540*, 812–823. https://doi.org/10.1016/j.jhydrol.2016.06.066

Knight, R., Walsh, D. O., Butler, J. J., Grunewald, E., Liu, G., Andrew, D., Reboulet, E. C., Knobbe, S., & Barrows, M. (2016). NMR Logging to Estimate Hydraulic Conductivity in Unconsolidated Aquifers. *Groundwater*, *54*(1), 104–114. https://doi.org/10.1111/gwat.12324

Kodinariya, T., & Makwana, P. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, *1*(6), 90–95.

Leblanc, D. R., Garabedian, S. P., Hess, K. M., Gelhar, L. W., Quadri, R. D., Stollenwerk, K. G., & Wood, W. W. (1991). Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts 1. Experimental design and observed tracer movement. *Water Resources Research*, *27*(5), 895–910. https://doi.org/10.1029/91WR00241

Leube, P. C., Nowak, W., & Schneider, G. (2012). Temporal moments revisited: Why there is no better way for physically based model reduction in time. *Water Resources Research*, *48*(11), 1–14. https://doi.org/10.1029/2012WR011973

Liu, F., Yeh, T. C. J., Wang, Y. L., Hao, Y., Wen, J. C., & Wang, W. (2020a). Characterization of basin-

scale aquifer heterogeneity using transient hydraulic tomography with aquifer responses induced by groundwater exploitation reduction. *Journal of Hydrology*, *588*, 125137. https://doi.org/10.1016/j.jhydrol.2020.125137

Liu, F., Yeh, T. C. J., Wang, Y. L., Song, X., Lei, X., Wen, J. C., Wang, W., & Hao, Y. (2020b). Potential of hydraulic tomography in identifying boundary conditions of groundwater basins. *Water Resources Research*, *56*(12), 1–18. https://doi.org/10.1029/2020WR028331

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

Liu, G., Zheng, C., & Gorelick, S. M. (2007b). Evaluation of the applicability of the dual-domain mass transfer model in porous media containing connected high-conductivity channels. *Water Resources Research*, *43*(12), 1–12. https://doi.org/10.1029/2007WR005965

Liu, G., Zheng, C., Tick, G. R., Butler, J. J., & Gorelick, S. M. (2010). Relative importance of dispersion and rate-limited mass transfer in highly heterogeneous porous media: Analysis of a new tracer test at the Macrodispersion Experiment (MADE) site. *Water Resources Research*, *46*(3), 1–8. https://doi.org/10.1029/2009WR008430

Liu, S., Yeh, T. C. J., & Gardiner, R. (2002). Effectiveness of hydraulic tomography: Sandbox experiments. *Water Resources Research*, *38*(4), 5-1-5–9. https://doi.org/10.1029/2001wr000338

Liu, X., Illman, W. A., Craig, A. J., Zhu, J., & Yeh, T. C. J. (2007a). Laboratory sandbox validation of transient hydraulic tomography. *Water Resources Research*, *43*(5), 1–13. https://doi.org/10.1029/2006WR005144

Lu, G. Y., & Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers and Geosciences*, *34*(9), 1044–1055. https://doi.org/10.1016/j.cageo.2007.07.010

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*, 127911. https://doi.org/10.1016/j.jhydrol.2022.127911

Luo, N., Illman, W. A., Zha, Y., Park, Y. J., & Berg, S. J. (2020). Three-dimensional hydraulic tomography analysis of long-term municipal wellfield operations: Validation with synthetic flow and solute transport data. *Journal of Hydrology*, *590*, 125438. https://doi.org/10.1016/j.jhydrol.2020.125438

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

Luo, N., Zhao, Z., Illman, W. A., Zha, Y., & Mok, C. M. W. (2023). Three-dimensional steady-state hydraulic tomography analysis with integration of cross-hole flowmeter data at a highly heterogeneous site. *Water Resources Research*, *59*, e2022WR034034. https://doi.org/10.1029/2022WR034034

Ma, R., Zheng, C., Zachara, J. M., & Tonkin, M. (2012). Utility of bromide and heat tracers for aquifer characterization affected by highly transient flow conditions. *Water Resources Research*, *48*(8), 1–18. https://doi.org/10.1029/2011WR011281

Mackay, D. M., Freyberg, D. L., Roberts, P. V., & Cherry, J. A. (1986). A natural gradient experiment on solute transport in a sand aquifer: 1. Approach and overview of plume movement. *Water Resources Research*, *22*(13), 2017–2029. https://doi.org/10.1029/WR022i013p02017

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *5th Berkeley Symposium on Math*, 281–297. https://doi.org/10.1007/s11665-016-2173-6

Maurer, J., & Knight, R. (2016). Models and methods for predicting hydraulic conductivity in near-

surface unconsolidated sediments using nuclear magnetic resonance. *Geophysics*, *81*(5), D503–D518. https://doi.org/10.1190/GEO2015-0515.1

McCall, W. (2010). *Tech Guide for Calculation of Estimated Hydraulic Conductivity (Est . K) Log from HPT Data*. Geoprobe Systems.

McCall, W., & Christy, T. M. (2020). The Hydraulic Profiling Tool for Hydrogeologic Investigation of Unconsolidated Formations. *Groundwater Monitoring and Remediation*, *40*(3), 89–103. https://doi.org/10.1111/gwmr.12399

Molz, F. J., & Young, S. C. (1993). Development and Application of Borehole Flowmeters for Environmental Assessment. *Log Analyst*, *34*(1), 13–23.

Morozov, D., Walsh, D., McLaughlin, C., Nguyen, A., Grunewald, E., Caldwell, W., Pipp, D., Basore, N., Christy, T., & Macedo, J. (2024). High-Resolution Direct Push NMR Tools for Groundwater Investigations. *Groundwater Monitoring and Remediation*. https://doi.org/10.1111/gwmr.12645

Neuman, S. P. (1972). Theory of flow in unconfined aquifers considering delayed response of the water table. *Water Resources Research*, *8*(4), 1031–1045. https://doi.org/10.1029/WR008i004p01031

Ni, C. F., Yeh, T. C. J., & Chen, J. S. (2009). Cost-effective hydraulic tomography surveys for predicting flow and transport in heterogeneous aquifers. *Environmental Science and Technology*, *43*(10), 3720–3727. https://doi.org/10.1021/es8024098

Ning, Z., Luo, N., Inaba, K., Nakashima, T., Shimizu, T., & Illman, W. A. (2023). Three-dimensional hydraulic tomography analyses to investigate commingling issues of reproducibility, data density, and geological prior models. *Journal of Hydrology*, *616*(September 2022), 128785. https://doi.org/10.1016/j.jhydrol.2022.128785

Paasche, H., Tronicke, J., & Dietrich, P. (2010). Automated integration of partially colocated models: Subsurface zonation using a modified fuzzy c-means cluster analysis algorithm. *Geophysics*, *75*(3).

https://doi.org/10.1190/1.3374411

Paradis, D., Lefebvre, R., Gloaguen, E., & Rivera, A. (2015). Predicting hydrofacies and hydraulic conductivity from direct-push data using a data-driven relavance vector machine approach: Motivations, algorithms, and application. *Water Resources Research*, *51*, 481–505. https://doi.org/10.1002/2013WR014979.Reply

Parsekian, A. D., Dlubac, K., Grunewald, E., Butler, J. J., Knight, R., & Walsh, D. O. (2015). Bootstrap Calibration and Uncertainty Estimation of Downhole NMR Hydraulic Conductivity Estimates in an Unconsolidated Aquifer. *Groundwater*, *53*(1), 111–121. https://doi.org/10.1111/gwat.12165

Perlo, Josefina, Danieli, E., Perlo, Juan, Blümich, B., & Casanova, F. (2013). Optimized slim-line logging NMR tool to measure soil moisture in situ. *Journal of Magnetic Resonance*, *233*, 74–79. https://doi.org/10.1016/j.jmr.2013.05.004

Pham, D. L. (2001). Spatial models for fuzzy clustering. *Computer Vision and Image Understanding*, *84*(2), 285–297. https://doi.org/10.1006/cviu.2001.0951

Pollock, D., & Cirpka, O. A. (2008). Temporal moments in geoelectrical monitoring of salt tracer experiments. *Water Resources Research*, *44*(12), 1–13. https://doi.org/10.1029/2008WR007014

Pollock, D., & Cirpka, O. A. (2010). Fully coupled hydrogeophysical inversion of synthetic salt tracer experiments. *Water Resources Research*, *46*(7), 1–12. https://doi.org/10.1029/2009WR008575

Pollock, D., & Cirpka, O. A. (2012). Fully coupled hydrogeophysical inversion of a laboratory salt tracer experiment monitored by electrical resistivity tomography. *Water Resources Research*, *48*(1), 1–13. https://doi.org/10.1029/2011WR010779

Poupon, A., & Gaymard, R. (1970). The evaluation of clay content from logs. *SPWLA Eleventh Annual Logging Symposium*.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*(336), 846–850. https://doi.org/10.1080/01621459.1971.10482356

Raymer, L., Hunt, E., & Gardner, J. (1980). An improved sonic transit time-to-porosity transform. *SPWLA 21st Annual Logging Symposium*.

Rehfeldt, K. R., Boggs, J. M., & Gelhar, L. W. (1992). Field study of dispersion in a heterogeneous aquifer: 4. Investigation of adsorption and sampling bias. *Water Resources Research*, *28*(12), 3325–3336. https://doi.org/10.1029/92WR01759

Remy, N., Boucher, A., & Wu, J. (2007). *Applied Geostatistics with SGeMS: A user's guide*. https://doi.org/10.1017/CBO9781139150019

Robin, M. J. L., Gutjahr, A. L., Sudicky, E. A., & Wilson, J. L. (1993). Cross-correlated random field generation with the direct Fourier Transform Method. *Water Resources Research*, *29*(7), 2385–2397. https://doi.org/10.1029/93WR00386

Ronayne, M. J., Gorelick, S. M., & Zheng, C. (2010). Geological modeling of submeter scale heterogeneity and its influence on tracer transport in a fluvial aquifer. *Water Resources Research*, *46*(10), 1–9. https://doi.org/10.1029/2010WR009348

Rubin, Y., & Dagan, G. (1987). Stochastic identification of transmissivity and effective recharge in steady groundwater flow: 1. Theory. *Water Resources Research*, *23*(7), 1185–1192. https://doi.org/10.1029/WR023i007p01185

Salamon, P., Fernàndez-Garcia, D., & Gómez-Hernández, J. J. (2007). Modeling tracer transport at the MADE site: The importance of heterogeneity. *Water Resources Research*, *43*(8), 1–15. https://doi.org/10.1029/2006WR005522

Sanchez-León, E., Leven, C., Haslauer, C. P., & Cirpka, O. A. (2016). Combining 3D Hydraulic Tomography with Tracer Tests for Improved Transport Characterization. *Groundwater*, *54*(4), 498–

507. https://doi.org/10.1111/gwat.12381

Sarris, T. S., Close, M., & Abraham, P. (2018). Using solute and heat tracers for aquifer characterization in a strongly heterogeneous alluvial aquifer. *Journal of Hydrology*, *558*, 55–71. https://doi.org/10.1016/j.jhydrol.2018.01.032

Schwede, R. L., Li, W., Leven, C., & Cirpka, O. A. (2014). Three-dimensional geostatistical inversion of synthetic tomographic pumping and heat-tracer tests in a nested-cell setup. *Advances in Water Resources*, *63*, 77–90. https://doi.org/10.1016/j.advwatres.2013.11.004

Seequent Ltd. (2021). Leapfrog Geo 2021. 3D Geological Modeling Software.

Seevers, D. O. (1966). A nuclear magnetic method for determining the permeability of sandstones. *SPWLA 7th Annual Logging Symposium 1966*.

Sezginer, A., Minh, C. C., Heaton, N., Herron, M., & Freedman, R. (1999). An NMR High-Resolution Permeability Indicator. *SPWLA 40th Annual Logging Symposium*.

Slater, L. (2007). Near surface electrical characterization of hydraulic conductivity: From petrophysical properties to aquifer geometries - A review. *Surveys in Geophysics*, *28*(2–3), 169–197. https://doi.org/10.1007/s10712-007-9022-y

Slater, L., & Lesmes, D. P. (2002). Electrical-hydraulic relationships observed for unconsolidated sediments. *Water Resources Research*, *38*(10), 31-1-31–13. https://doi.org/10.1029/2001wr001075

Somogyvari, M., & Bayer, P. (2017). Field validation of thermal tracer tomography for reconstruction of aqufier heterogeneity. *Water Resources Research*, *53*, 5070–5084. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Sucre, O., Pohlmeier, A., Minière, A., & Blümich, B. (2011). Low-field NMR logging sensor for measuring hydraulic parameters of model soils. *Journal of Hydrology*, *406*(1–2), 30–38.

https://doi.org/10.1016/j.jhydrol.2011.05.045

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/10.1029/92WR01758

Sudicky, E. A., Cherry, J. A., & Frind, E. O. (1983). MIGRATION OF CONTAMINANTS IN GROUNDWATER AT A LANDFILL : A CASE STUDY 4 . A Natural-Gradient Dispersion Test. *Journal of Hydrology*, *63*, 81–108. https://doi.org/10.1016/0022-1694(83)90224-X

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), 1–16. 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. Technical report submitted to Geoprobe Inc.

Sun, D., Luo, N., Vandenhoff, A., McCall, W., Zhao, Z., Wang, C., Rudolph, D. L., & Illman, W. A. (2024). Evaluation of Hydraulic Conductivity Estimates from Various Approaches with Groundwater Flow Models. *Groundwater*, *62*(3), 384–404. https://doi.org/10.1111/gwat.13348

Tarnawski, V. R., Momose, T., McCombie, M. L., & Leong, W. H. (2015). Canadian Field Soils III. Thermal-Conductivity Data and Modeling. *International Journal of Thermophysics*, *36*(1), 119–156. https://doi.org/10.1007/s10765-014-1793-z

Theis, C. V. (1935). The relation between the lowering of the Piezometric surface and the rate and duration of discharge of a well using ground-water storage. In *Eos, Transactions American Geophysical Union* (Vol. 16, Issue 2). https://doi.org/10.1029/TR016i002p00519

Tiedeman, C. R., & Barrash, W. (2020). Hydraulic Tomography: 3D Hydraulic Conductivity, Fracture Network, and Connectivity in Mudstone. *Groundwater*, *58*(2), 238–257. https://doi.org/10.1111/gwat.12915

Timur, A. (1968). An investigation of permeability, porosity, and residual water saturation relationships. *SPWLA 9th Annual Logging Symposium*.

Topp, G. C., Davis, J. L., & Annan, A. P. (1980). Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resources Research*, *16*(3), 574–582. https://doi.org/10.1029/WR016i003p00574

Tso, C. M., Zha, Y., Yeh, T.-C. J., & Wen, J.-C. (2015). The relative importance of head, flux, and prior information in hydraulic tomography analysis. *Water Resources Research*, *51*, 3–20. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Urumović, K. (2016). The referential grain size and effective porosity in the Kozeny-Carman model. *Hydrology and Earth System Sciences*, *20*(5), 1669–1680. https://doi.org/10.5194/hess-20-1669-2016

Vandenbohede, A., Louwyck, A., & Lebbe, L. (2009). Conservative solute versus heat transport in porous media during push-pull tests. *Transport in Porous Media*, *76*(2), 265–287. https://doi.org/10.1007/s11242-008-9246-4

Wagner, V., Li, T., Bayer, P., Leven, C., Dietrich, P., & Blum, P. (2014). Thermal tracer testing in a sedimentary aquifer:Field experiment (Lauswiesen, Germany) and numerical simulation. *Hydrogeology Journal*, *22*(1), 175–187. https://doi.org/10.1007/s10040-013-1059-z

Walsh, D., Turner, P., Grunewald, E., Zhang, H., Jr, J. J. B., Reboulet, E., Knobbe, S., Christy, T., Jr, J. W. L., Johnson, C. D., Munday, T., & Fitzpatrick, A. (2013). A Small-Diameter NMR Logging Tool for Groundwater Investigations. *Groundwater*, *51*(6), 914–926. https://doi.org/10.1111/gwat.12024

Wang, C., & Illman, W. A. (2023). Hydraulic Tomography Estimates Improved by Zonal Information From the Clustering of Geophysical Survey Data. *Water Resources Research*, *59*, e2023WR035191. https://doi.org/10.1029/2023WR035191

Wang, C., Steelman, C. M., & Illman, W. A. (2025b). Characterizing Spatial Heterogeneity of Hydraulic Conductivity using Borehole NMR in a Complex Groundwater Flow System. *Water Resources Research*, *61*, e2024WR039717. https://doi.org/10.1029/2024WR039717

Wang, C., Steelman, C. M., Ning, Z., Walsh, D. O., & Illman, W. A. (2025a). Comparison of NMR-Derived Hydraulic Conductivity with Various Hydraulic Testing Methods. *Groundwater*, *63*(5), 713–724. https://doi.org/10.1111/gwat.70016

Weller, A., Slater, L., & Nordsiek, S. (2013). On the relationship between induced polarization and surface conductivity: Implications for petrophysical interpretation of electrical measurements. *Geophysics*, *78*(5). https://doi.org/10.1190/GEO2013-0076.1

Williamson, P. (2016). *Examination of the Electrical-Hydraulic Conductivity Relationship at a Highly Heterogeneous Site*. University of Waterloo.

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

Xia, R., Liang, Y., Jim Yeh, T. C., Dang, X., Gu, X., & Xu, B. (2023). Characterizing hydraulic heterogeneity of Bayin River Basin using river stage tomography. *Journal of Hydrology*, *621*, 129547. https://doi.org/10.1016/j.jhydrol.2023.129547

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), 1–14. https://doi.org/10.1029/2008WR007180

Yeh, T.-C. J. C. J., Jin, M., & Hanna, S. (1996). An Iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields. *Water Resources Research*, *32*(1), 85–92. https://doi.org/10.1029/95WR02869

Yeh, T.-C. J., Gutjahr, A. L., & Jin, M. (1995). An Interative Cokriging-like Technique for Ground-Water Flow Modeling. *Ground Water*, *33*(1), 33–41.

Yeh, T.-C. J., Srivastava, R., Guzman, A., & Harter, T. (1993). A Numerical Model for Water Flow and Chemical Transport in Variably Saturated Porous Media. *Ground Water*, *31*(4), 634–644.

Yeh, T. C. J., Jin, M., & Hanna, S. (1996). An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields. *Water Resources Research*, *32*(1), 85–92. https://doi.org/10.1029/95WR02869

Yeh, T. C. J., Lee, C. H., Hsu, K. C., Illman, W. A., Barrash, W., Cai, X., Daniels, J., Sudicky, E., Wan, L., Li, G., & Winter, C. L. (2008b). A view toward the future of subsurface characterization: CAT scanning groundwater basins. *Water Resources Research*, *44*(3), 1–9. https://doi.org/10.1029/2007WR006375

Yeh, T. C. J., Lee, C. H., Hsu, K. C., & Wen, J. C. (2008a). Fusion of hydrologic and geophysical tomographic surveys. *Geosciences Journal*, *12*(2), 159–167. https://doi.org/10.1007/s12303-008-0017-6

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

Yeh, T. C. J., Liu, S., Glass, R. J., Baker, K., Brainard, J. R., Alumbaugh, D., & LaBrecque, D. (2002). A geostatistically based inverse model for electrical resistivity surveys and its applications to vadose zone hydrology. *Water Resources Research*, *38*(12), 14-1-14–13. https://doi.org/10.1029/2001wr001204

Yeh, T. C. J., Xiang, J., Suribhatla, R. M., Hsu, K. C., Lee, C. H., & Wen, J. C. (2009). River stage tomography: A new approach for characterizing groundwater basins. *Water Resources Research*, *45*(5), 1–14. https://doi.org/10.1029/2008WR007233

Zha, Y., Yeh, T. C. J., Illman, W. A., Onoe, H., Mok, C. M. W., Wen, J. C., Huang, S. Y., & Wang, W. (2017). Incorporating geologic information into hydraulic tomography: A general framework based on geostatistical approach. *Water Resources Research*, *53*(4), 2850–2876. https://doi.org/10.1002/2016WR019185

Zha, Y., Yeh, T. C. J., Illman, W. A., Tanaka, T., Bruines, P., Onoe, H., & Saegusa, H. (2015). What does hydraulic tomography tell us about fractured geological media? A field study and synthetic experiments. *Journal of Hydrology*, *531*, 17–30. https://doi.org/10.1016/j.jhydrol.2015.06.013

Zha, Y., Yeh, T. C. J., Illman, W. A., Tanaka, T., Bruines, P., Onoe, H., Saegusa, H., Mao, D., Takeuchi, S., & Wen, J. C. (2016). An Application of Hydraulic Tomography to a Large-Scale Fractured Granite Site, Mizunami, Japan. *Groundwater*, *54*(6), 793–804. https://doi.org/10.1111/gwat.12421

Zhang, Z., Wang, G., Carranza, E. J. M., Yang, S., Zhao, K., Yang, W., & Sha, D. (2022). Three-Dimensional Pseudo-Lithologic Modeling Via Adaptive Feature Weighted k-Means Algorithm from Multi-Source Geophysical Datasets, Qingchengzi Pb–Zn–Ag–Au District, China. *Natural Resources Research*, *31*(4), 2163–2179. https://doi.org/10.1007/s11053-021-09927-0

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*, *615*, 128673. 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. (2021). On the importance of considering specific storage heterogeneity in hydraulic tomography: Laboratory sandbox and synthetic studies. *Journal of Hydrology*, *593*(September 2020), 125874. https://doi.org/10.1016/j.jhydrol.2020.125874

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*, 127971. 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*, 128124. 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

Zhao, Z., Illman, W. A., Zha, Y., Yeh, T. C. J., Mok, C. M. B., Berg, S. J., & Han, D. (2019). Transient hydraulic tomography analysis of fourteen pumping tests at a highly heterogeneous multiple aquifer-aquitard system. *Water (Switzerland)*, *11*(9). https://doi.org/10.3390/w11091864

Zhao, Z., Luo, N., & Illman, W. A. (2023). Geostatistical analysis of high-resolution hydraulic conductivity estimates from the hydraulic profiling tool and integration with hydraulic tomography at a highly heterogeneous field site. *Journal of Hydrology*, *617*, 120960. https://doi.org/10.1016/j.jhydrol.2023.129060

Zheng, C., Bianchi, M., & Gorelick, S. M. (2011). Lessons learned from 25 years of research at the

MADE site. *Groundwater*, *49*(5), 649–662. https://doi.org/10.1111/j.1745-6584.2010.00753.x

Zhu, J., Wang, H., Zhao, H., & Lin, T. (2021). High depth resolution design of small diameter nuclear magnetic resonance logging probe based on compensating magnets for soil moisture measurement in situ. *Journal of Hydrology*, *603*, 127031. https://doi.org/10.1016/j.jhydrol.2021.127031

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

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

Appendix A: Supplementary Information for Study I: Hydraulic Tomography Estimates Improved by Zonal Information from the Clustering of Geophysical Survey Data

Text A1: Detailed descriptions of EC data generation.

To restore the complex hydrogeophysical environment, a non-monotonic relationship between EC and *K* is built for the generation of the true EC field. As there is no direct relationship between EC and *K* (Slater, 2007), grain size is introduced to associate EC with *K*. The Hazen method (Hazen, 1892) is applied to link the reference grain size with *K* in this experiment:

$$K = 1000 \times {d_{10}}^2 \tag{A.1}$$

where d_{10} is the reference grain size (mm) at which 10 percent of the weight is finer, and K has a unit of m/day.

The bulk EC (σ) has two components: the pore-surface conductance and the pore-fluid conductance, which exhibit disparate correlation patterns to the reference grain size. In terms of the pore-surface part, two main controls of electrical conductance are considered, including clay content (i.e., chargeability of clay mineral) and specific surface area (i.e., contact area of the electrical double layer). Referring to a direct-current resistivity survey at the NCRS site, a constant clay conductivity of 40 mS/m is assumed. Effective surface conductivity (σ_{11}) from clay conductance at each layer is calculated as the product of clay content times clay electrical conductivity. In addition, imaginary conductivity is introduced to derive the surface conductivity in clay-free minerals. A grain size model from Slater and

Lesmes (2002) is coupled with a linear relationship (Weller et al., 2013) between surface and imaginary conductivity to link surface conductivity in clay-free minerals to the grain size.

$$\sigma^{"} = 0.0005 \times {d_{10}}^{-1} \tag{A.2}$$

$$\sigma^{"} = 0.042\sigma_{12} \tag{A.3}$$

where $\sigma^{"}$ is the imaginary conductivity (mS/m), and σ_{12} is the surface conductivity of clay-free minerals. In sum, the pore-surface part of the reference EC field is shown in Figure A1a. In terms of pore-fluid part, the following form of Archie's law (Archie, 1942) is used for the estimation of electrolyte conductivity σ_2 (mS/m):

$$\sigma_2 = {n_e}^{c} \cdot \sigma_f \tag{A.4}$$

where σ_f is fluid conductivity (mS/m), n_e is effective porosity, and c is the Archie cementation factor, an empirical value which is suggested to be 1.3 for unconsolidated soil (Doveton, 1986). However, fluid conductivity varies significantly in the field. In this study, a constant value of 100 mS/m is assigned as the fluid conductivity, which is a common value for low-salinity groundwater. Afterall, Weller et al. (2013) indicates that surface conductivity values parallel with the variation of fluid conductivity over a wide range of salinity. In addition, Urumović (2016) provides a model in which effective porosity is a function of reference grain size. Thus, a relationship can be built between the electrolyte conductivity and the reference grain size, and the bulk conductivity is a linear superposition of all components:

$$\sigma = \sigma_{11} + \sigma_{12} + \sigma_2 \tag{A.5}$$

Text A2: Detailed descriptions of the three external CVIs

External Cluster Validity Indices (CVIs) (Fowlkes & Mallows, 1983; Jaccard, 1912; Rand, 1971) are employed to compare the similarity of zone geometry between clustering results and the reference model. For dataset $\boldsymbol{D}^* = \{(\boldsymbol{x_1}, \boldsymbol{y_1}), (\boldsymbol{x_2}, \boldsymbol{y_2}), \dots, (\boldsymbol{x_m}, \boldsymbol{y_m})\}$, assuming that a clustering result is $\boldsymbol{C} = \{\boldsymbol{C_1}, \boldsymbol{C_2}, \dots, \boldsymbol{C_k}\}$, the partitioning in the reference model is $\boldsymbol{C}^* = \{\boldsymbol{C_1^*}, \boldsymbol{C_2^*}, \dots, \boldsymbol{C_s^*}\}$. Accordingly, $\boldsymbol{\varphi} = (\varphi_1; \varphi_2; \dots; \varphi_k)$ and $\boldsymbol{\varphi}^* = (\varphi_1^*; \varphi_2^*; \dots; \varphi_s^*)$ are the label vectors for $\boldsymbol{C}$ and $\boldsymbol{C}^*$, respectively. For a pair of samples $(\boldsymbol{x_i}, \boldsymbol{y_i})$ and $(\boldsymbol{x_j}, \boldsymbol{y_j})$, one can define:

$$a = |SS|,\ SS = \left\{(\boldsymbol{x_i}, \boldsymbol{y_i}), (\boldsymbol{x_j}, \boldsymbol{y_j}) \middle| \varphi_i = \varphi_j, \varphi_i^* = \varphi_j^*, i < j\right\} \tag{A.6}$$

$$b = |SD|,\ SD = \left\{(\boldsymbol{x_i}, \boldsymbol{y_i}), (\boldsymbol{x_j}, \boldsymbol{y_j}) \middle| \varphi_i = \varphi_j, \varphi_i^* \neq \varphi_j^*, i < j\right\} \tag{A.7}$$

$$c = |DS|,\ DS = \left\{(\boldsymbol{x_i}, \boldsymbol{y_i}), (\boldsymbol{x_j}, \boldsymbol{y_j}) \middle| \varphi_i \neq \varphi_j, \varphi_i^* = \varphi_j^*, i < j\right\} \tag{A.8}$$

$$c = |DS|,\ DS = \left\{(\boldsymbol{x_i}, \boldsymbol{y_i}), (\boldsymbol{x_j}, \boldsymbol{y_j}) \middle| \varphi_i \neq \varphi_j, \varphi_i^* = \varphi_j^*, i < j\right\} \tag{A.9}$$

where *a*, *b*, *c* and *d* are the quantity of sample pairs assigned to: 1) the same clusters in both clustering results and the reference model; 2) the same clusters in clustering results while to the different clusters in the reference model; 3) assigned to different clusters in clustering results while to the same clusters in the reference model and 4) assigned to different clusters in both clustering results and the reference model, respectively.

Based on Equations (A.6)–(A.9), three common CVIs are considered, including Jaccard Coefficient (JC) (Jaccard, 1912), Fowlkes and Mallows Index (FMI) (Fowlkes & Mallows, 1983) and Rand Index (RI) (Rand, 1971):

$$JC = \frac{a}{a + b + c} \tag{A.10}$$

$$FMI = \sqrt{\frac{a}{a+b} \cdot \frac{a}{a+c}} \tag{A.11}$$

$$RI = \frac{2(a+d)}{m(m-1)} \tag{A.12}$$

CVIs range from 0 to 1. The larger the CVIs are, the better clustering results can reproduce the reference model.

Text A3: Detailed descriptions of the additional tracer test.

A conservative tracer with relative concentration (C/C_0) of 1.0 was injected at Port 6-1 at a rate of 0.1 mL/s, while water was extracted at Port 1-5 at a rate of 0.1 mL/s. Tracer injection and dissipation each lasted over a period of 48 h. The same effective transport parameters (see Table 3.1 in the main text) were used in the simulation of this additional tracer test.

C/C_0 distributions at timesteps of 4 h, 12 h, 48 h, and 96 h for all cases and the true K field are provided in Figure A11. The tracer plume based on the true K field (see Figure A11k) initially evolved to an elliptical shape with a long axis developing in the x-direction after 12 h. However, the tracer plumes migrated towards the top boundary during this stage from the simulations in Cases I, II-b, and IV-b, which may result from the inaccurate characterization of Layer 17. Next, tracer plumes mainly migrated along the high-K Layers 14, 15, and 17 to the effluent port after 48 h. At 96 h (i.e., 48 h after the stop of tracer injection), most of the tracer mass has been extracted.

Likewise, concentration data at sampling ports were collected every two hours. In terms of distances from the source, nine representative sampling ports are utilized to yield BTCs during 96 hours of tracer migration, and three classes are designated: (1) near the source: Ports 6-3, 5-1 and 5-2; (2) intermediate distances: Ports 5-3, 4-2 and 3-4; and (3) large distances: Ports 2-3, 2-5 and 1-5. BTCs of the nine sampling ports during the additional tracer test are plotted in Figure A12. For clarity, BTCs of six representative cases are provided in this figure, including Cases I, II-b, III-b, IV-b, V-b and VI. Scatterplots of simulated versus C/C_0 at 12 h and 48 h, and the scatterplots at all timesteps at an interval of 2 h, are provided in Figures A13 and A14.

The temporal evolution of tracer plumes is further studied through temporal moment analyses. Scatterplots of simulated versus observed M_0, μ, and σ^2 are provided in Figures A15-17, respectively. Figure A18 summarizes the computed MAE and RMSE norms for M_0, μ, and σ^2. A similar ranking of all cases based on the solute transport prediction performances of this additional tracer test was obtained. The

main difference lies in the performance of Cases II-b and IV-b, which rank behind Case I in the simulation of this additional tracer test. It reveals that zonation models based on the clustering of indirect geophysical information might not be so reliable when spatial constraints are not incorporated. The other findings from the analyses of this additional tracer test corroborate the conclusions from the primary test discussed in the main text.

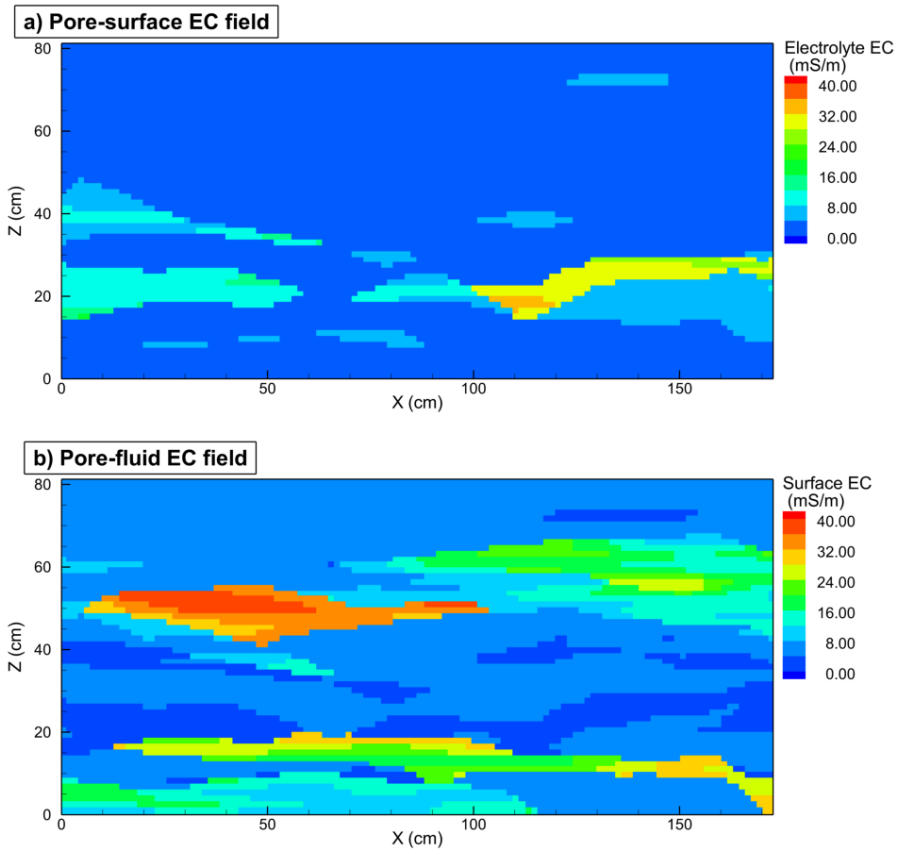


Figure A1: Generation of two components of the bulk EC field. a) spatial distribution of the pore-surface EC field; b) spatial distribution of the pore-fluid EC field.

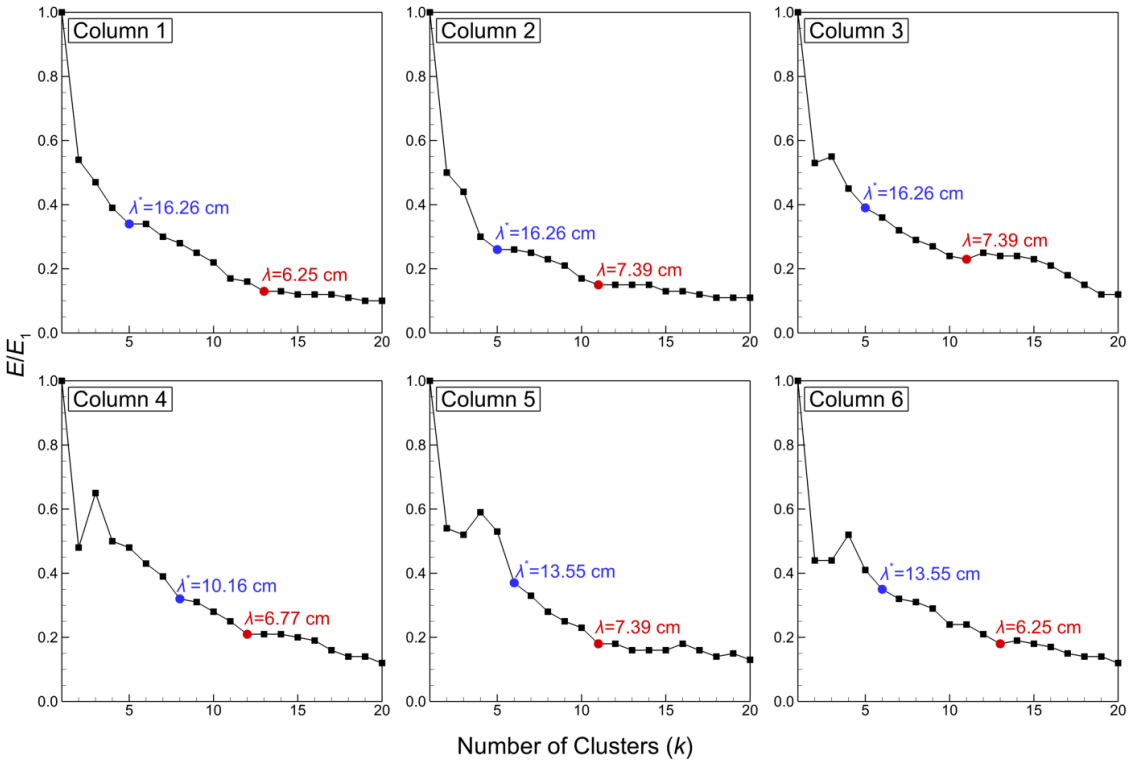


Figure A2: Estimation of average unit thickness using borehole ln *K* profiles in which Relative sum of squares (E/E_1) are plotted as a function of Number of Clusters (k) for six columns, respectively. Red and blue circles denote the preferential and the alternative elbows selected from the curves. Corresponding average unit thicknesses (λ) for each elbow point are also indicated.

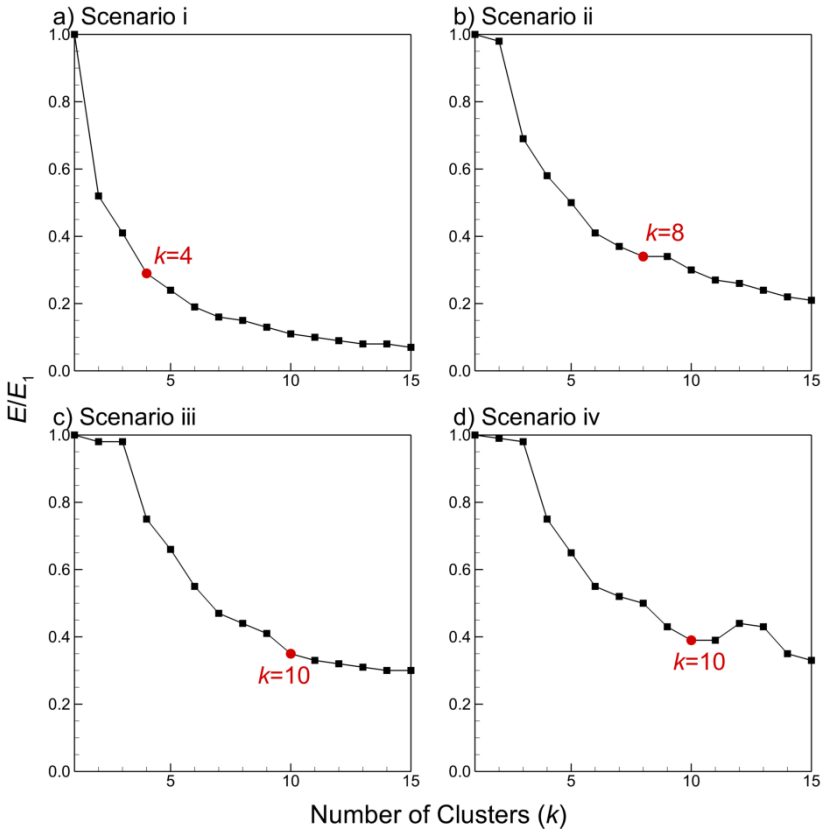


Figure A3: Number of Clusters (k) determined by the "elbow" curve method for all scenarios. a) Scenario i: k-means clustering without spatial constraints; b) Scenario ii: k-means clustering with a preferential vertical constraint and a horizontal constraint generated by multiplying a factor of 10; c) Scenario iii: k-means clustering with an alternative vertical constraint and a horizontal constraint generated by multiplying a factor of 10; d) Scenario iv: k-means clustering with the preferential vertical constraint and a horizontal constraint generated by multiplying a factor of 6. Red circles in the plot represent the inflection points selected as "elbows" with corresponding k values indicated.

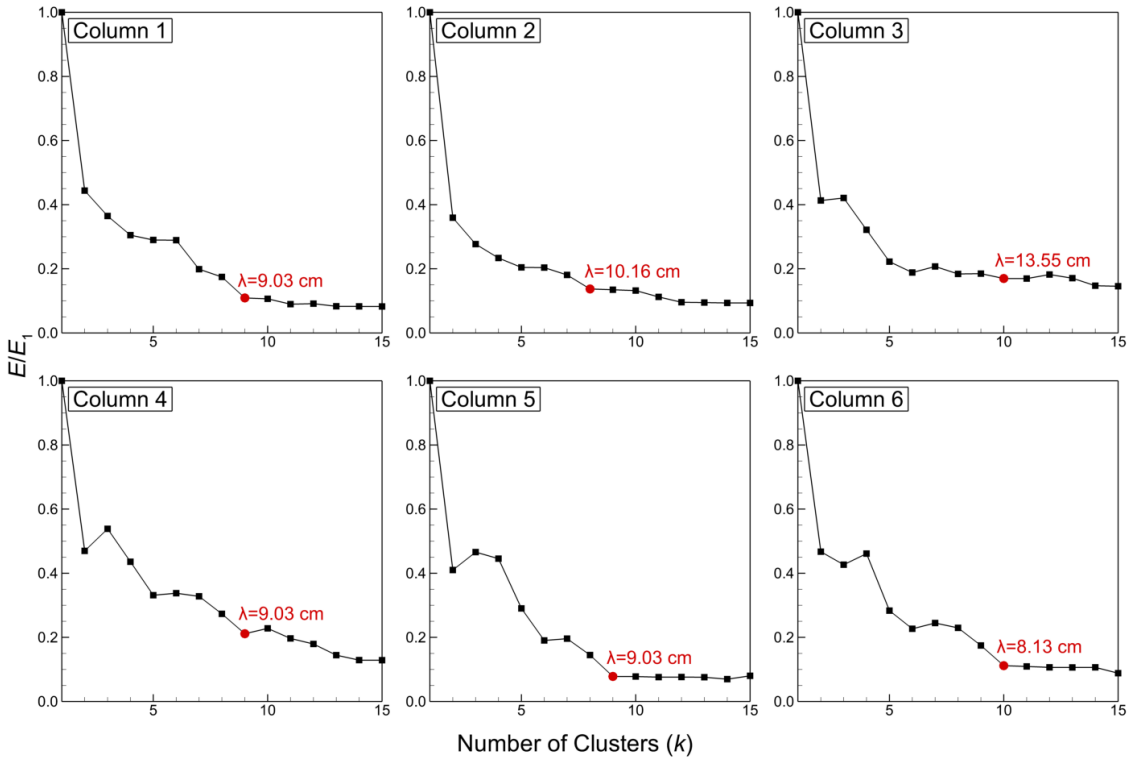


Figure A4: Estimation of average unit thicknesses (λ) using borehole EC profiles in which E/E_1 is plotted as a function of Number of Clusters (k) for six columns, respectively. Red circles denote elbows selected from the curves, while the λ values are also indicated.

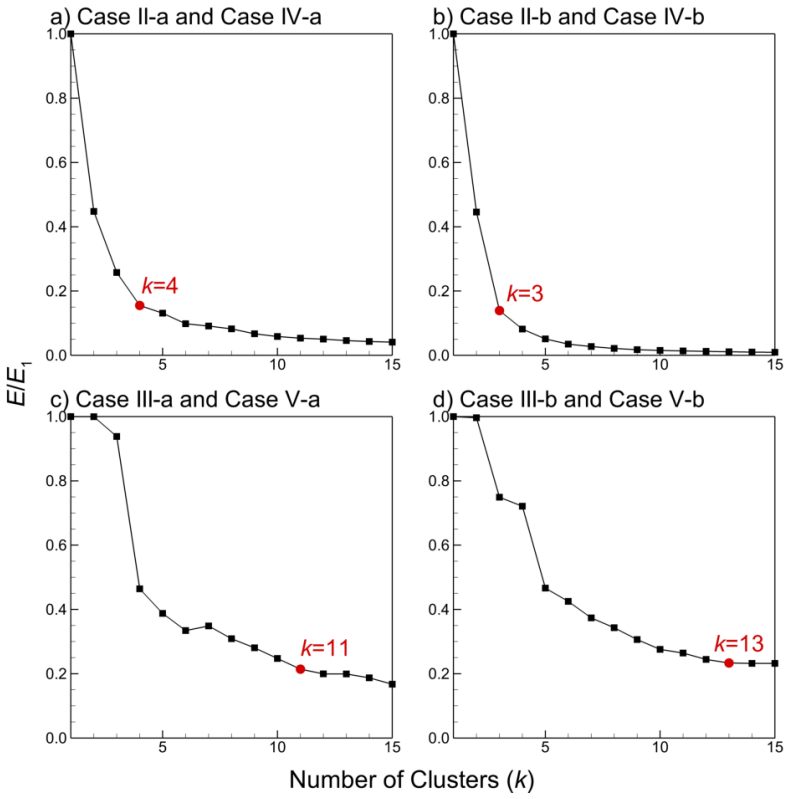


Figure A5: Number of Clusters (k) determined by the "elbow" curve method for all the cases. a) Case II-a and Case IV-a: the clustering of EC logs only; b) Case II-b and Case IV-b: the clustering of EC and GR logs; c) Case III-a and Case V-a: the clustering of EC logs with spatial constraints; d) Case III-b and Case V-b: the clustering of EC and GR logs with spatial constraints. Red circles in the plot represent the inflection points selected as "elbows" with corresponding k values indicated.

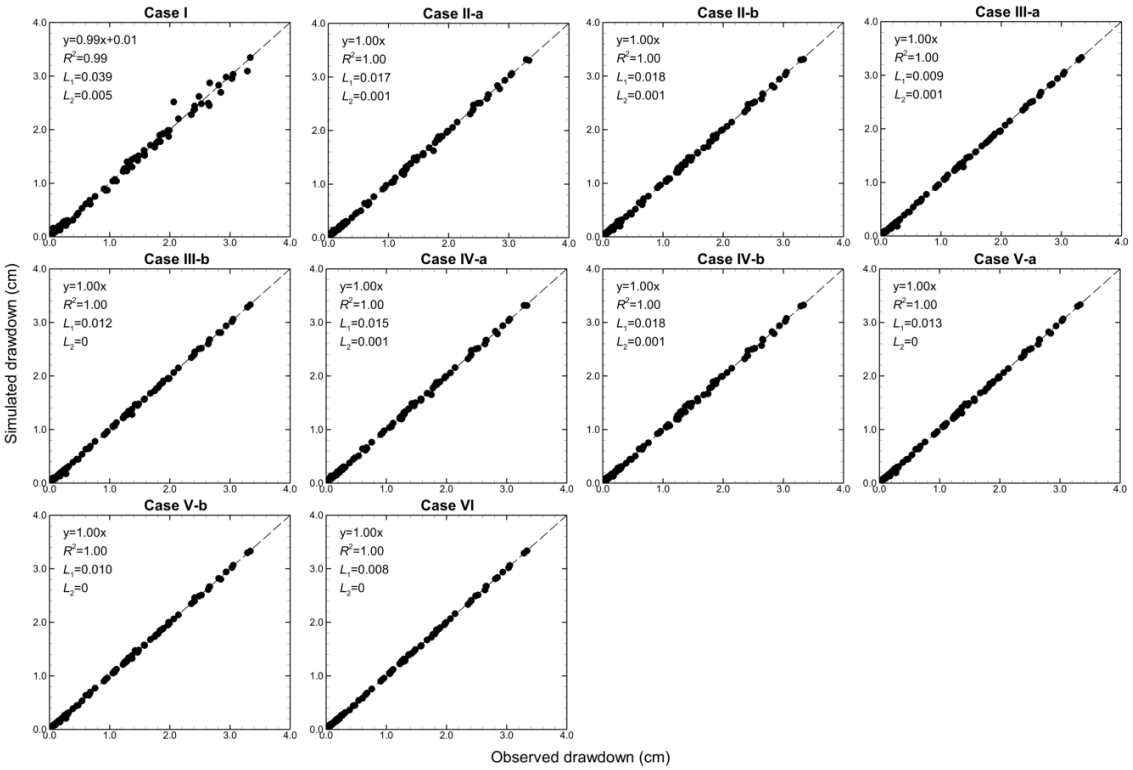


Figure A6: Calibration scatterplots of simulated versus observed drawdowns for all cases. R^2, L_1 and L_2 norms are indicated on each plot.

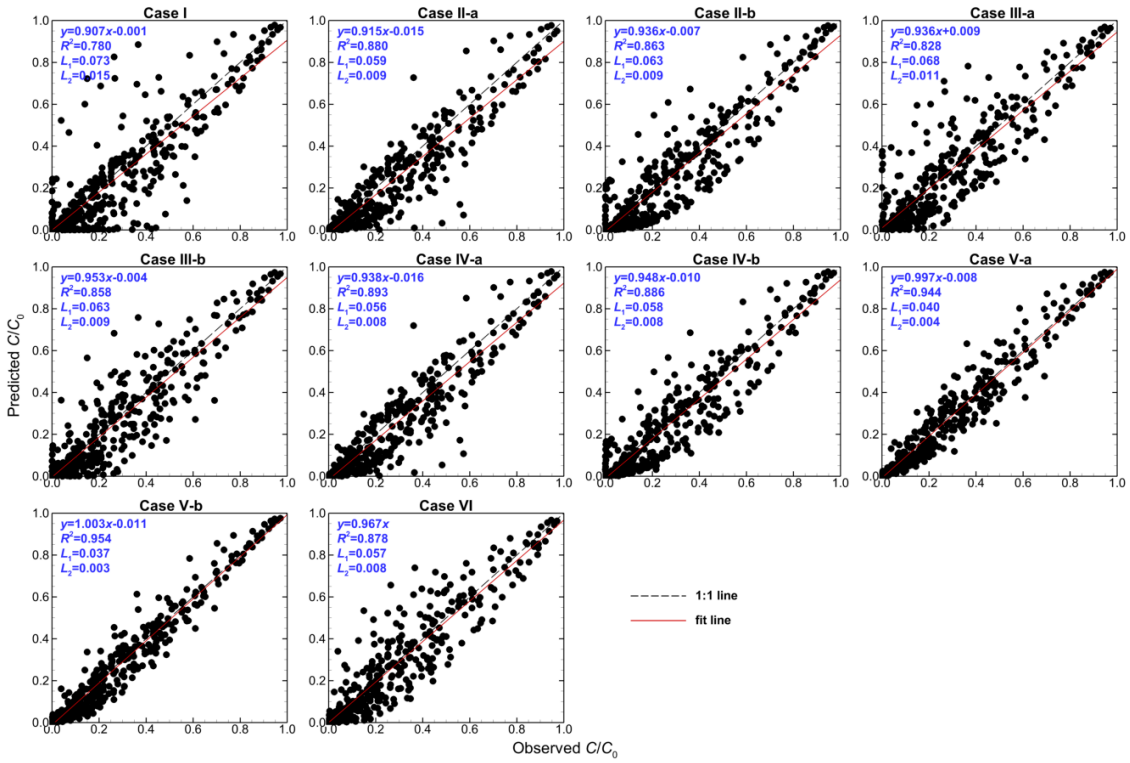


Figure A7: Scatterplots of predicted versus observed C/C_0 at all 47 sampling ports (exclusive of the influent port) for all cases. R^2, L_1 and L_2 norms are indicated on each plot.

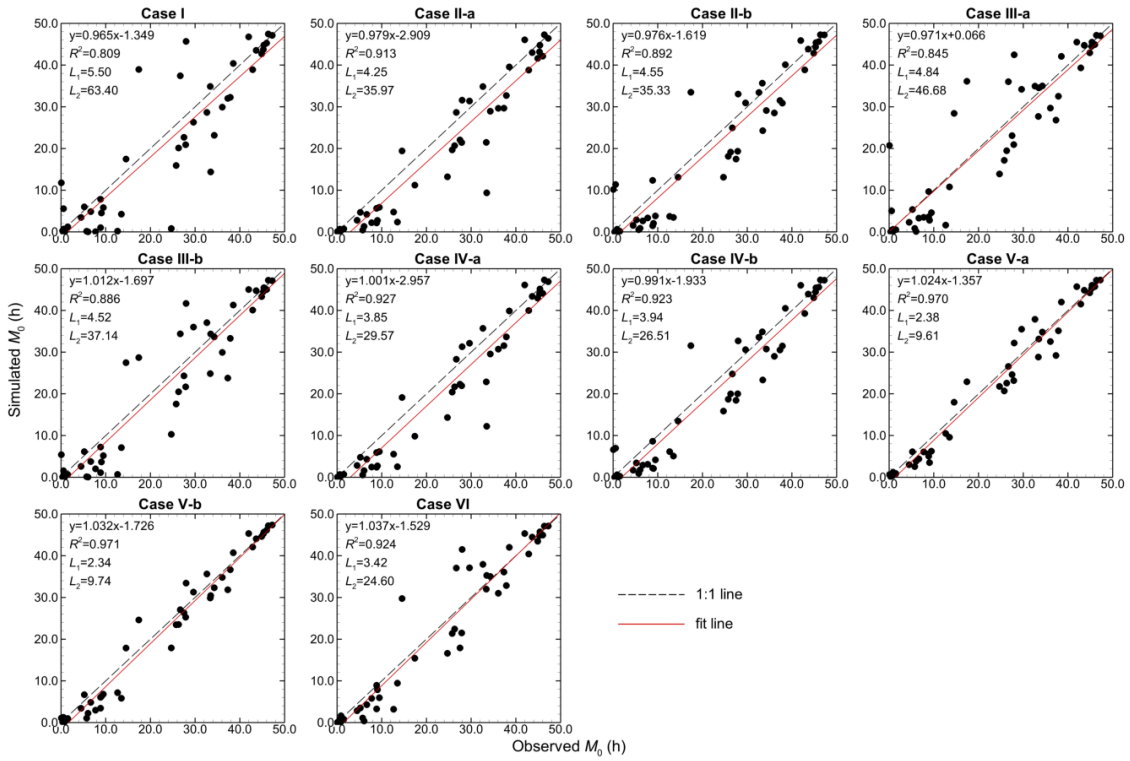


Figure A8: Scatterplots of simulated versus observed M_0 at 47 sampling ports for all cases. R^2, L_1 and L_2 norms are indicated on each plot.

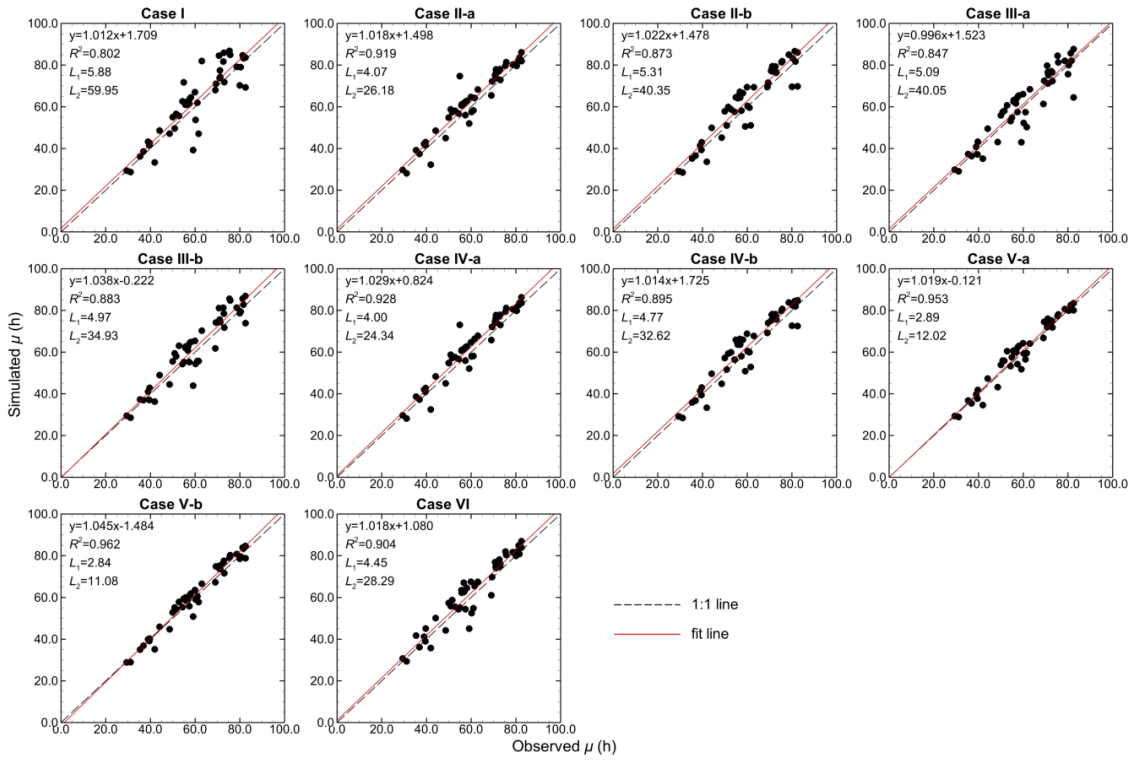


Figure A9: Scatterplots of simulated versus observed μ at 47 sampling ports for all cases. R^2, L_1 and L_2 norms are indicated on each plot.

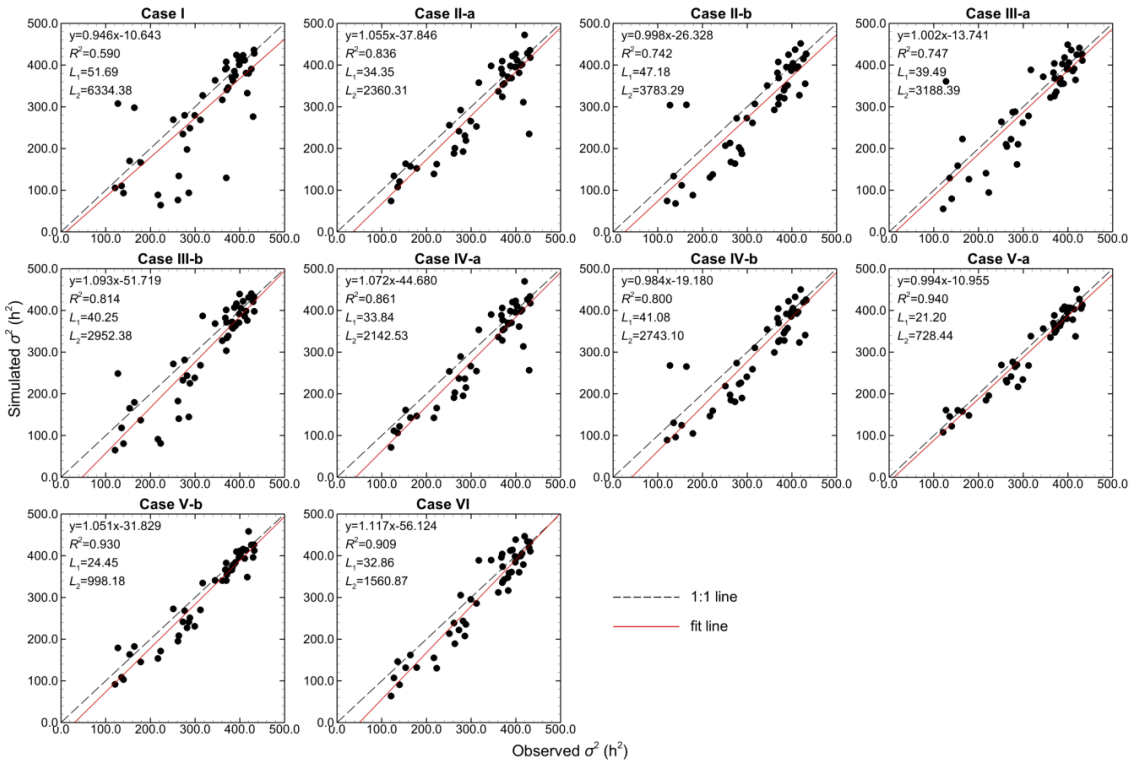


Figure A10: Scatterplots of simulated versus observed σ^2 of BTCs at 47 sampling ports for all cases. R^2, L_1 and L_2 norms are indicated on each plot.

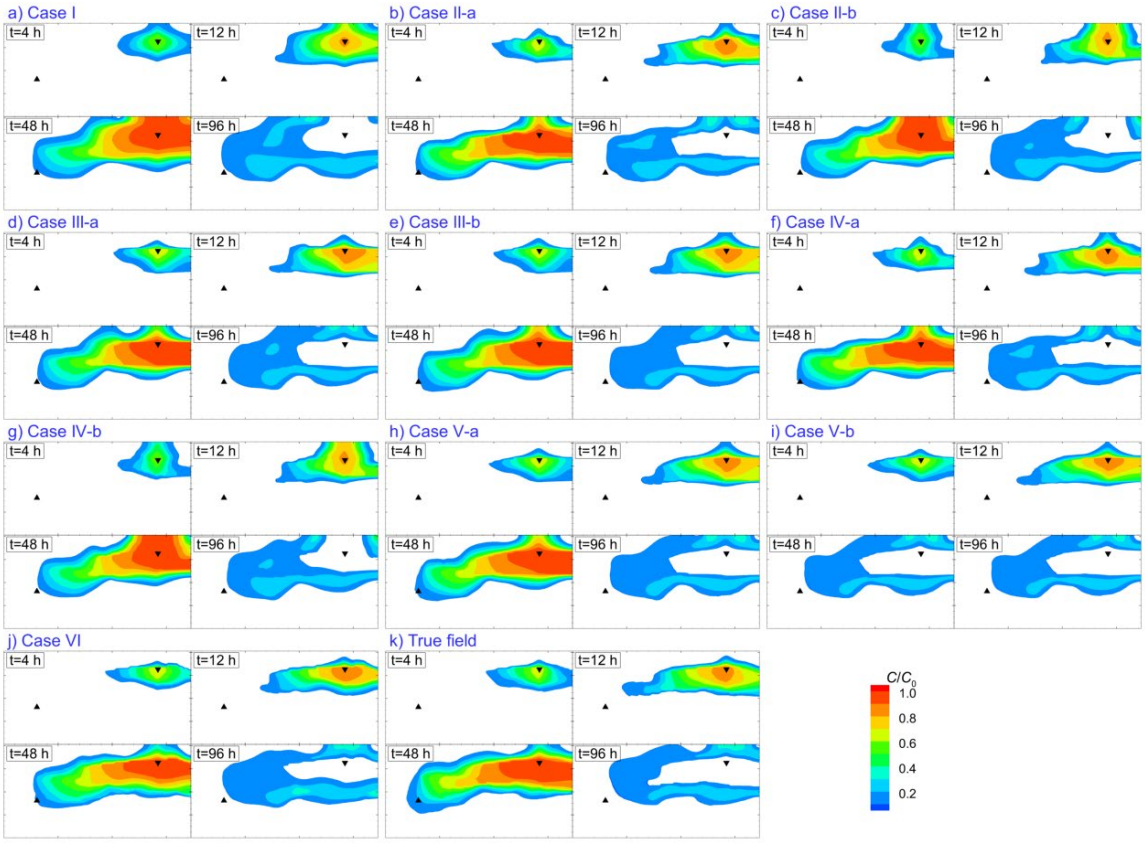


Figure A11: Relative concentration distributions (C/C_0) at 4 h, 12 h, 48 h, and 96 h during the additional tracer test. a) through j) represent C/C_0 profiles for Cases I, II-a, II-b, III-a, III-b, IV-a, IV-b, V-a, V-b, and VI. k) shows the C/C_0 distributions based on the spatial distribution of the true K field. Influent and effluent ports for each subplot are denoted with downward and upward-pointing triangles, respectively. Contours with C/C_0 under 0.1 are removed for clarity.

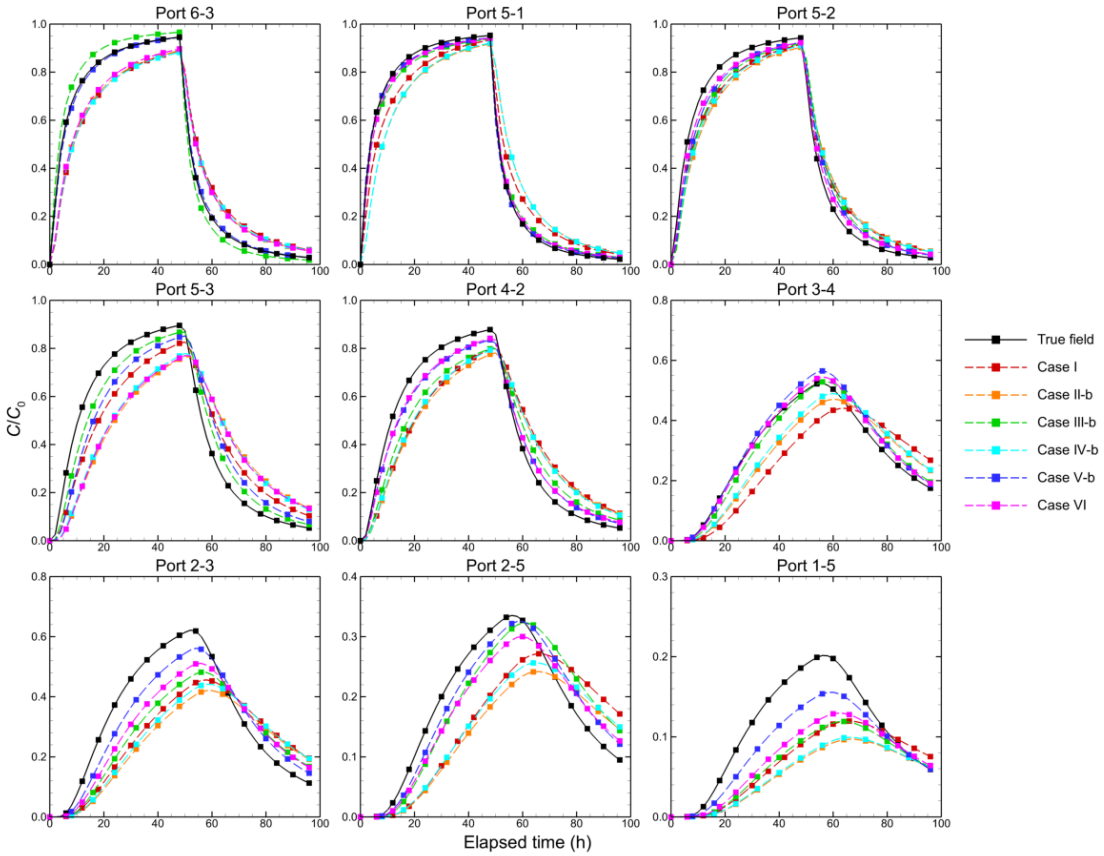


Figure A12: Breakthrough curves of nine sampling ports during the additional tracer test. The upper three ports (Ports 6-3, 5-1 and 5-2) are near the source, the intermediate three ports (Ports 5-3, 4-2 and 3-4) are at intermediate distances, and the lower three ports (Ports 2-3, 2-5 and 1-5) are at large distances from the influent port.

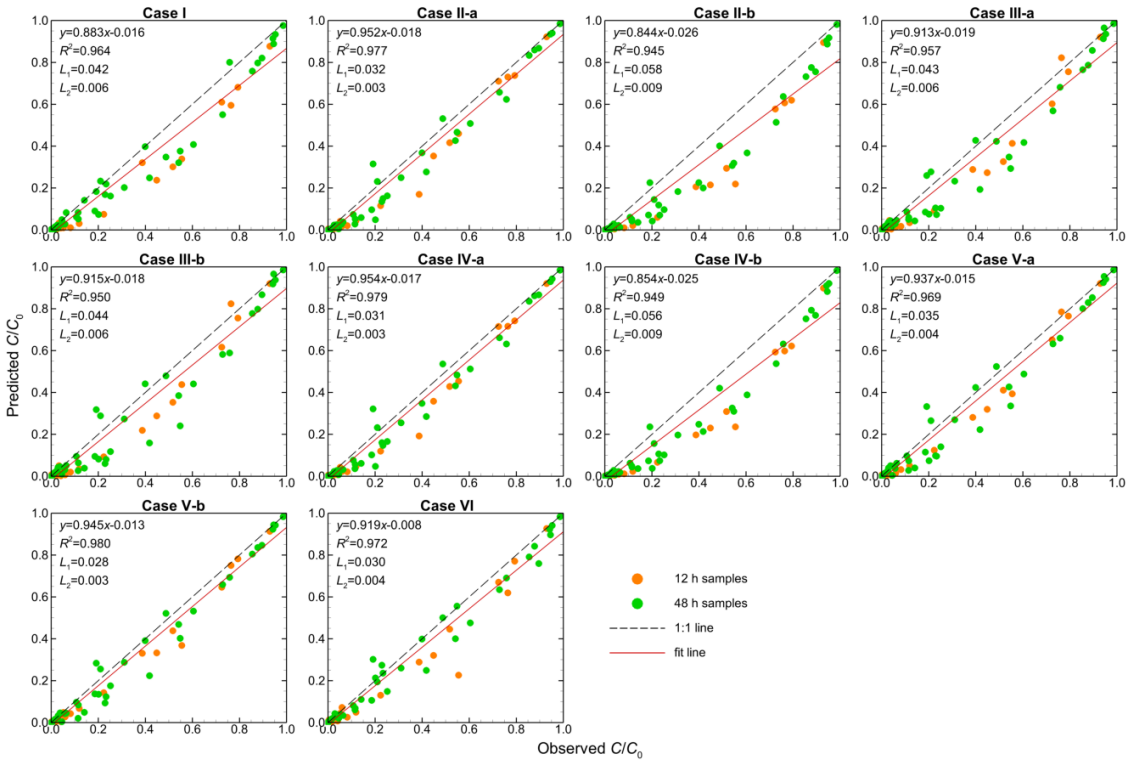


Figure A13: Scatterplots of predicted versus observed C/C_0 at 12 h and 48 h at 47 sampling ports for all cases during the additional tracer test. R^2, L_1 and L_2 norms are indicated on each plot.

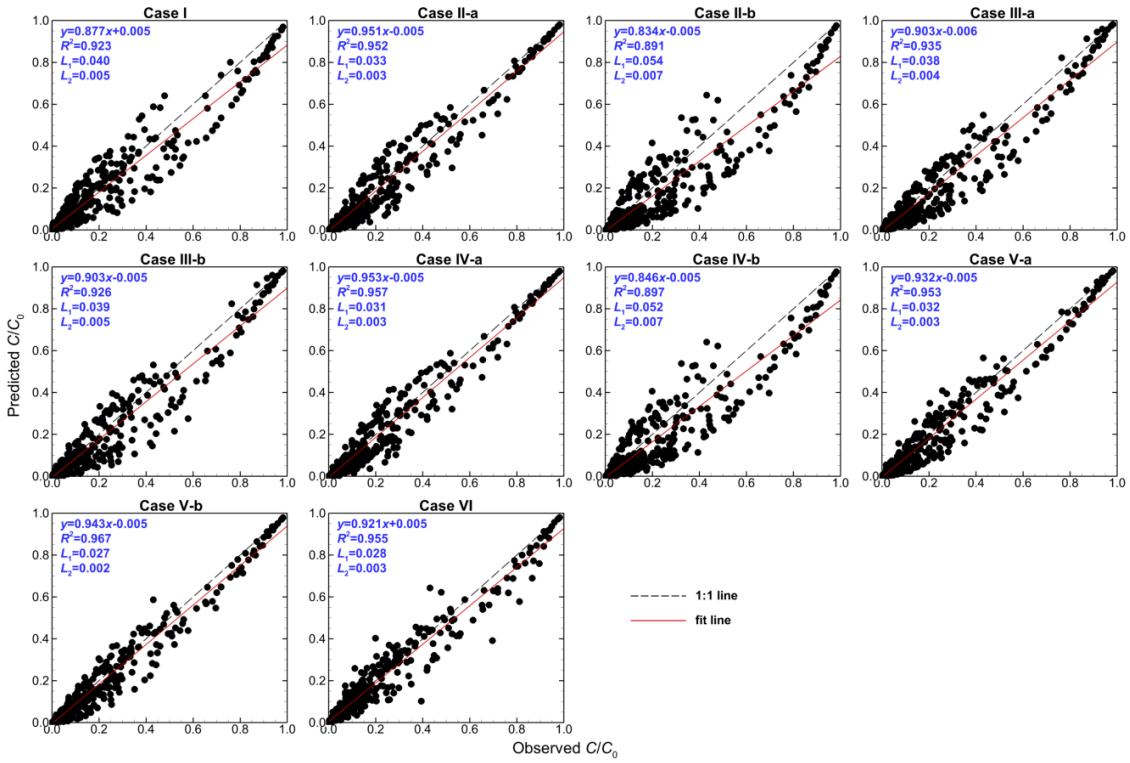


Figure A14: Scatterplots of predicted versus observed C/C_0 at all 47 sampling ports (exclusive of the influent port) for all cases during the additional tracer test. R^2, L_1 and L_2 norms are indicated on each plot.

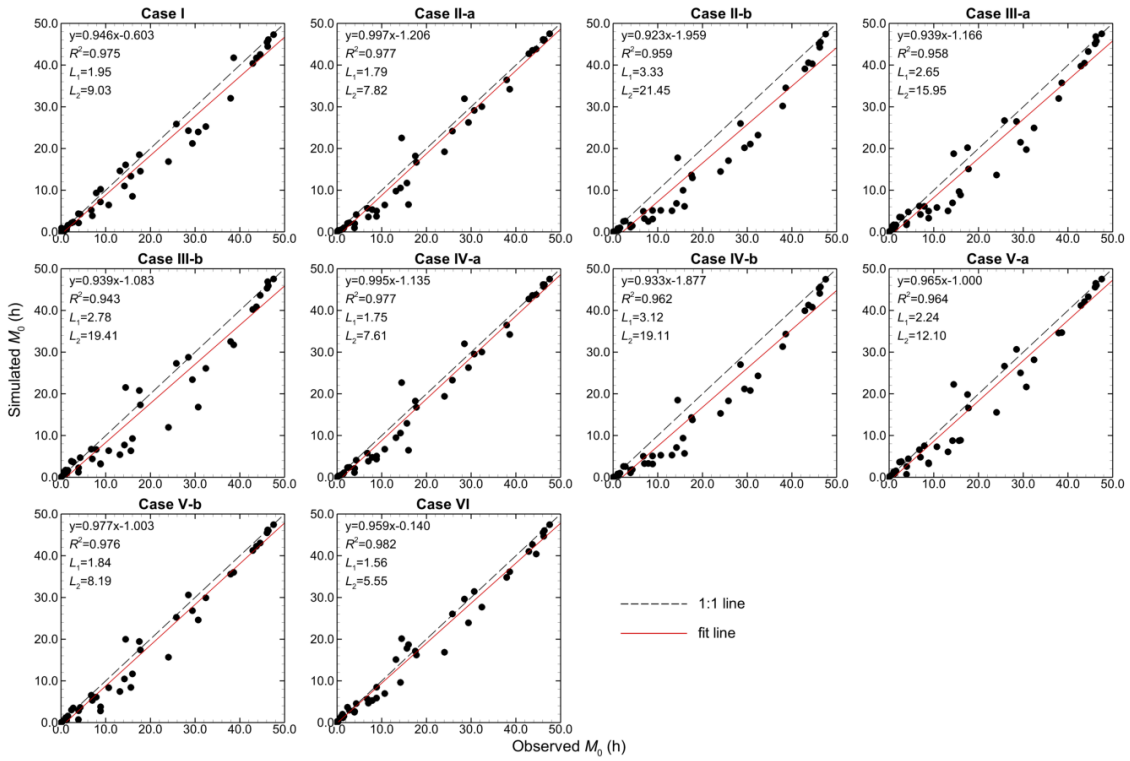


Figure A15: Scatterplots of simulated versus observed M_0 at 47 sampling ports during the additional tracer test. R^2, L_1 and L_2 norms are indicated on each plot.

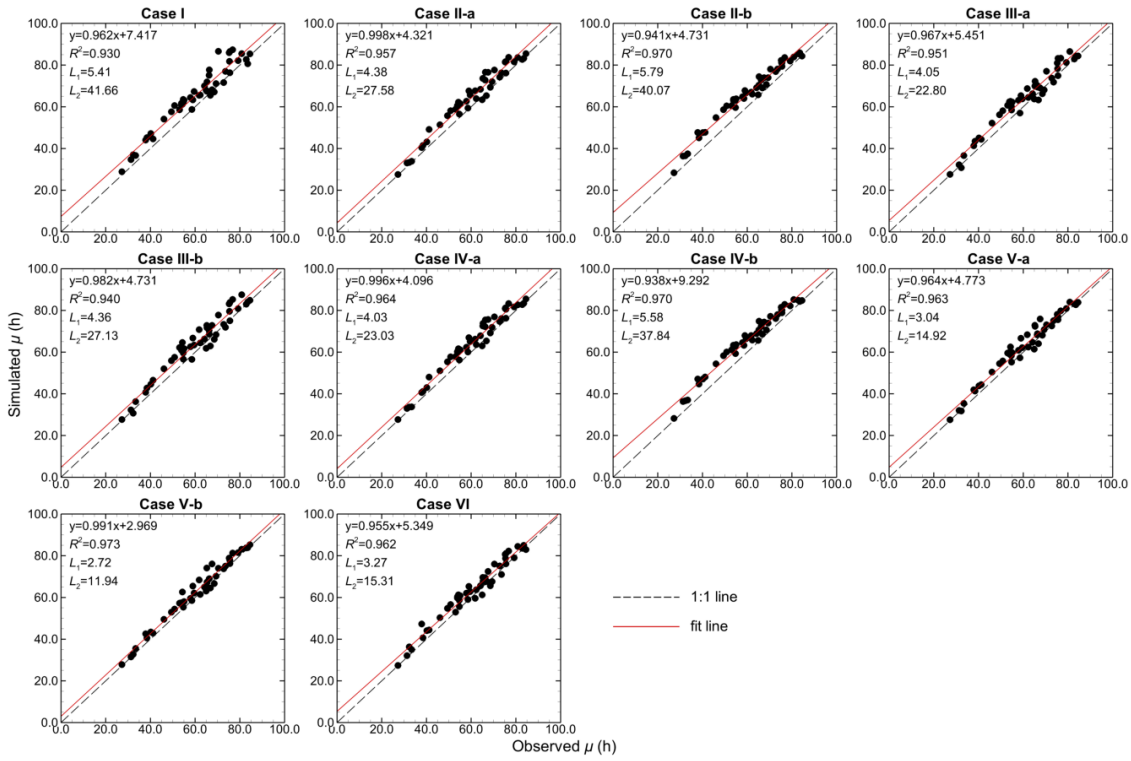


Figure A16: Scatterplots of simulated versus observed μ at 47 sampling ports for all cases during the additional tracer test. R^2, L_1 and L_2 norms are indicated on each plot.

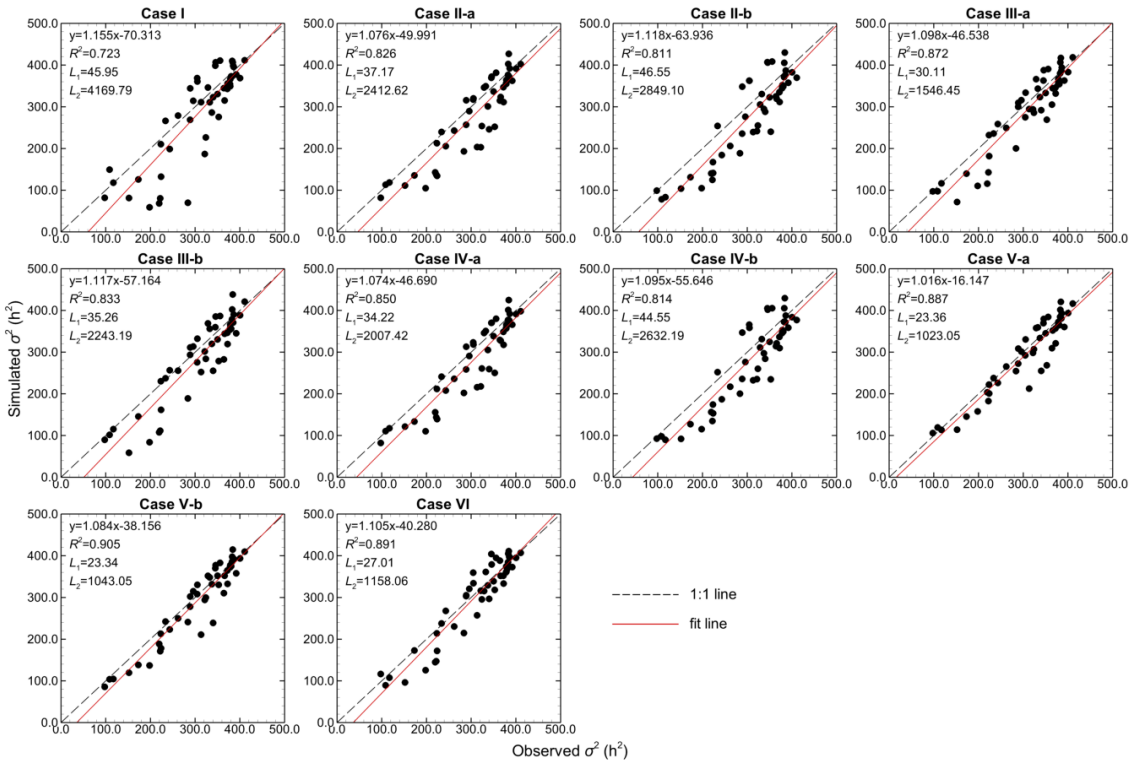


Figure A17: Scatterplots of simulated versus observed σ^2 of BTCs at 47 sampling ports for all cases. R^2, L_1 and L_2 norms are indicated on each plot.

Case	M_0 (h)		μ (h)		σ^2 (h^2)		Rank
	L_1	L_2	L_1	L_2	L_1	L_2	
I	1.95	9.03	5.41	41.66	45.95	4169.79	8
II-a	1.79	7.82	4.38	27.58	37.17	2412.62	6
II-b	3.33	21.45	5.79	40.07	46.55	2849.10	10
III-a	2.65	15.95	4.05	22.80	30.11	1546.45	5
III-b	2.78	19.41	4.36	27.13	35.26	2243.19	7
IV-a	1.75	7.61	4.03	23.03	34.22	2007.42	4
IV-b	3.12	19.11	5.58	37.84	44.55	2632.19	9
V-a	2.24	12.10	3.04	14.92	23.36	1023.05	2
V-b	1.84	8.19	2.72	11.94	23.34	1043.05	1
VI	1.56	5.55	3.27	15.31	27.01	1158.06	3

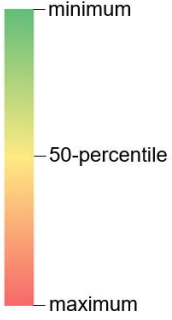


Figure A18: A summary of the color-coded L_1 and L_2 norms for M_0, μ and σ^2 for all cases of the additional tracer test. A ranking of all cases is provided on the right-most column.

Table A1: The effective values of *K* for the 18 layers. The effective values are obtained by calculating the geometric mean of all elemental *K* values for each layer.

Layer number	Effective value of *K* (m/s)	Layer number	Effective value of *K* (m/s)
1	1.71E-05	10	1.79E-07
2	1.07E-06	11	1.26E-07
3	2.49E-07	12	4.60E-06
4	9.14E-05	13	8.25E-07
5	1.43E-07	14	9.03E-04
6	6.15E-08	15	1.48E-05
7	1.11E-06	16	2.63E-06
8	4.29E-08	17	4.10E-05
9	6.48E-07	18	5.41E-07

Table A2: Pumping rates (Q) for all pumping tests, including those for model calibration and those for model validation.

Port ID	Q (mL/s)	Use	Port ID	Q (mL/s)	Use
2-3	5.0E-02	Calibration	1-3	5.0E-02	Validation
2-7	3.0E-02	Calibration	3-3	5.0E-02	Validation
5-2	5.0E-02	Calibration	3-7	3.0E-02	Validation
5-5	5.0E-03	Calibration	4-3	5.0E-02	Validation
			4-7	3.0E-02	Validation
			6-5	5.0E-03	Validation

Appendix B: Supplementary Information for Study III: Characterizing Spatial Heterogeneity of Hydraulic Conductivity using Borehole NMR in a Complex Groundwater Flow System

Text B1: Detailed descriptions of the site-specific optimization of the SDR model.

The C_{SDR} constant was calibrated using the K measurements from permeameter tests. For permeameter tests, a 10-cm split spoon sampler was used to collect continuous vertical core samples at an interval of around 10 cm during the drilling of CMT wells. The core samples were delivered to the laboratory for falling-head permeameter tests by Alexander et al. (2011). In total, 248 measurements of permeameter-based K obtained at the locations of CMT1, CMT2, CMT3, and CMT4 within similar depth ranges of NMR logging were included for this site-specific optimization. The sampling interval of permeameter tests is 10 cm. In comparison, the sampling interval of NMR logging is 15 cm. For statistical analysis, we paired NMR-derived K estimates with permeameter-based K measurements within a 10 cm distance in the vertical direction, resulting in 171 co-located K measurements associated with CMT1, CMT2, CMT3, and CMT4.

A linear regression after log transform was performed to calibrate the C_{SDR} constant. Here, we calibrated two C_{SDR} constants in different cases: when $m = 0$ and $n = 2$, the optimized C_{SDR} for the data domain is 3.86×10^{-3} m/s, with a range from 2.70×10^{-3} m/s to 7.21×10^{-3} m/s for the scale of each CMT well; when $m = 1$ and $n = 2$, the overall optimized $C_{SDR} = 0.0130$ m/s, with a range from 0.0093 m/s to 0.0244 m/s for the well scale. Note that the former case excluded porosity from the SDR model.

Two metrics, mean absolute error (MAE) and root mean squared error (RMSE), were computed to quantify the SDR model fits with permeameter measurements. For the SDR model excluding porosity (i.e., $m = 0$), the MAE and RMSE are 1.950 and 2.787, respectively, whereas the SDR model including porosity (i.e., $m = 1$) yielded the MAE and RMSE of 1.957 and 2.811, which are slightly larger than the former case. Therefore, porosity does not improve the K estimation using the SDR model for the glaciofluvial deposits at the NCRS, and we adopted the SDR model excluding porosity.

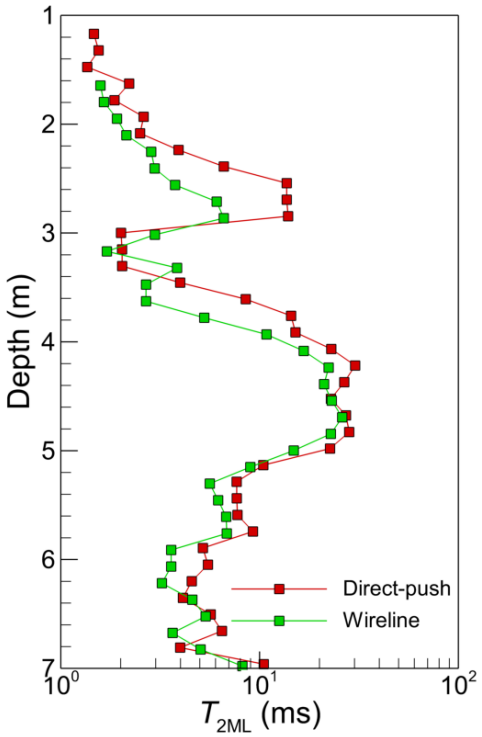


Figure B1: Two co-located NMR logs in PW4s. The line in red shows the NMR logging lowered into an open hole using the direct-push technique, while the line in green indicates the logging inserted into a PVC-lined monitoring well (i.e., PW4-3). These two logs are within 1 m distance.

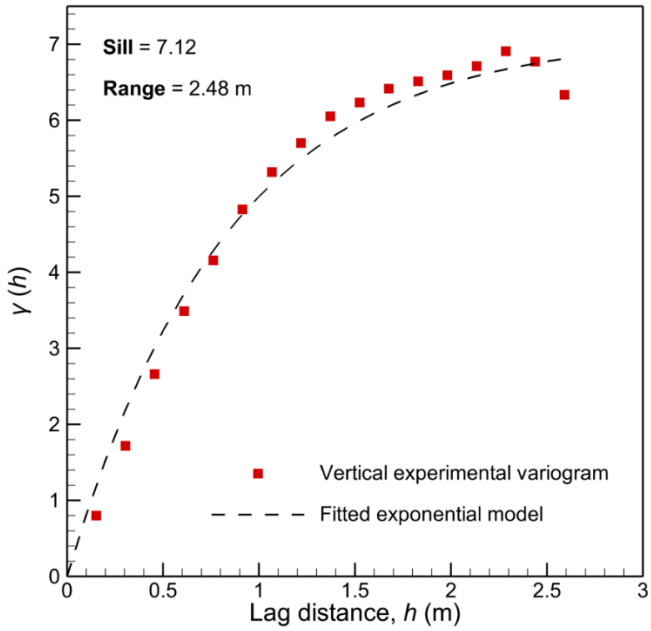


Figure B2: The vertical experimental variogram and the fitted exponential model. The ln *K* data used for variogram analysis are from eight NMR logs.

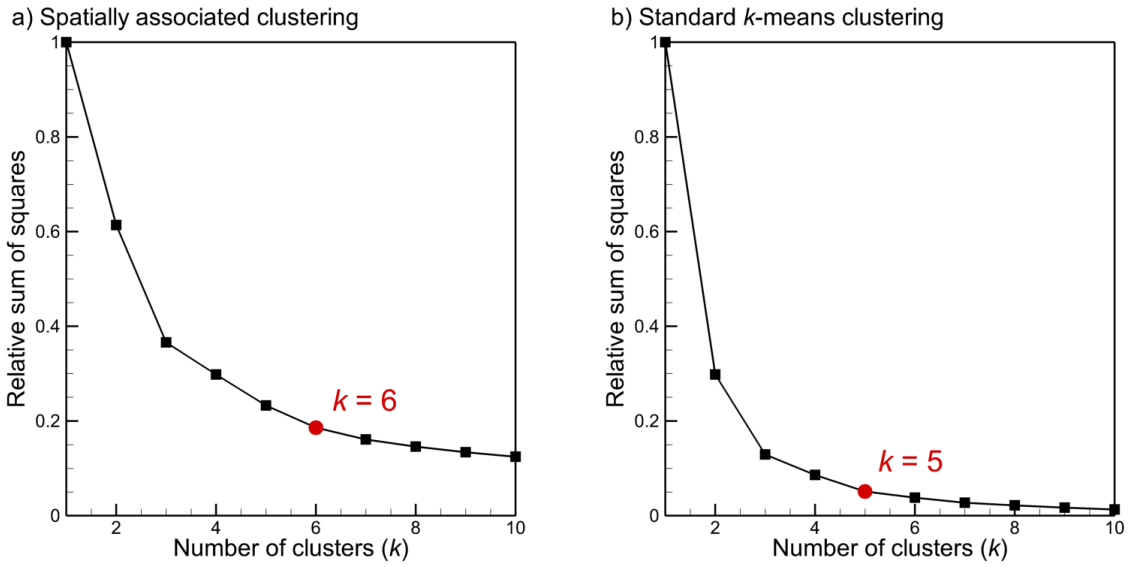


Figure B3: Number of clusters (k) determined by the "elbow" curve method for clustering of ln K data from NMR logs a) with and b) without spatial constraints. The red circle represents the inflection point selected as the "elbow" with the corresponding k value indicated in the plots. The relative sum of squares is calculated as the sum of squares for the current k over the sum of squares at $k = 1$.

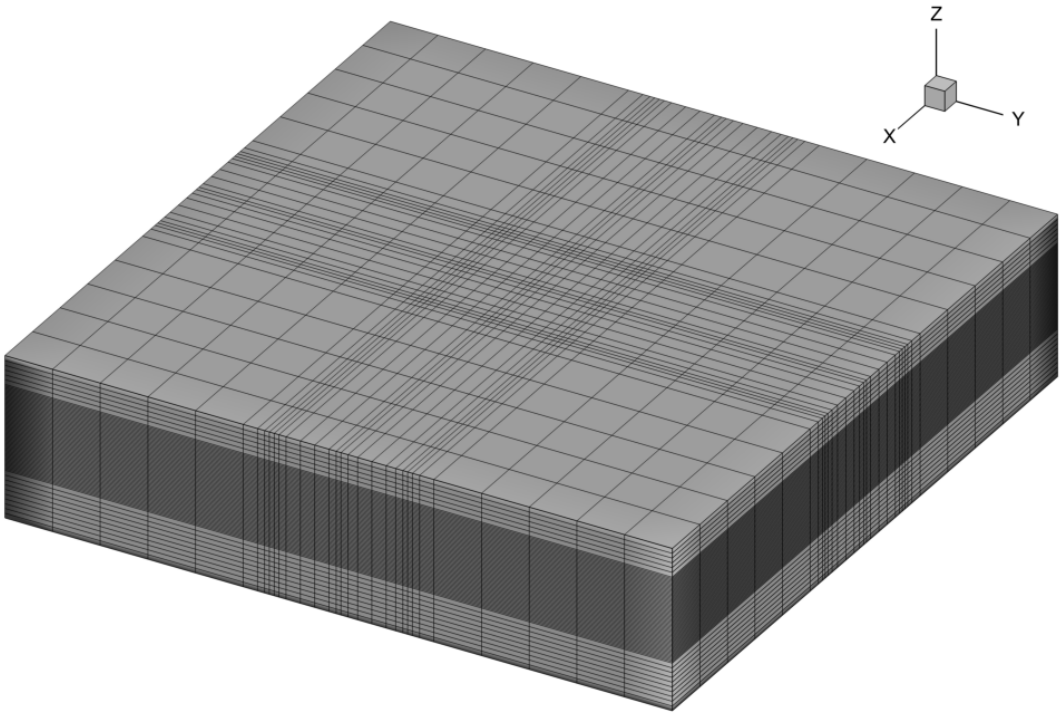


Figure B4: Grid discretization for the groundwater flow modeling in HydroGeoSphere. The top boundary is assigned to be 340.4 masl, with a thick aquitard layer treated as the bottom boundary. In total, 73036 elements were generated in model discretization, and the block size decreases gradually from 5 m × 5 m × 0.5 m near the boundaries to 0.5 m × 0.5 m × 0.15 m around the pumping and observation wells.

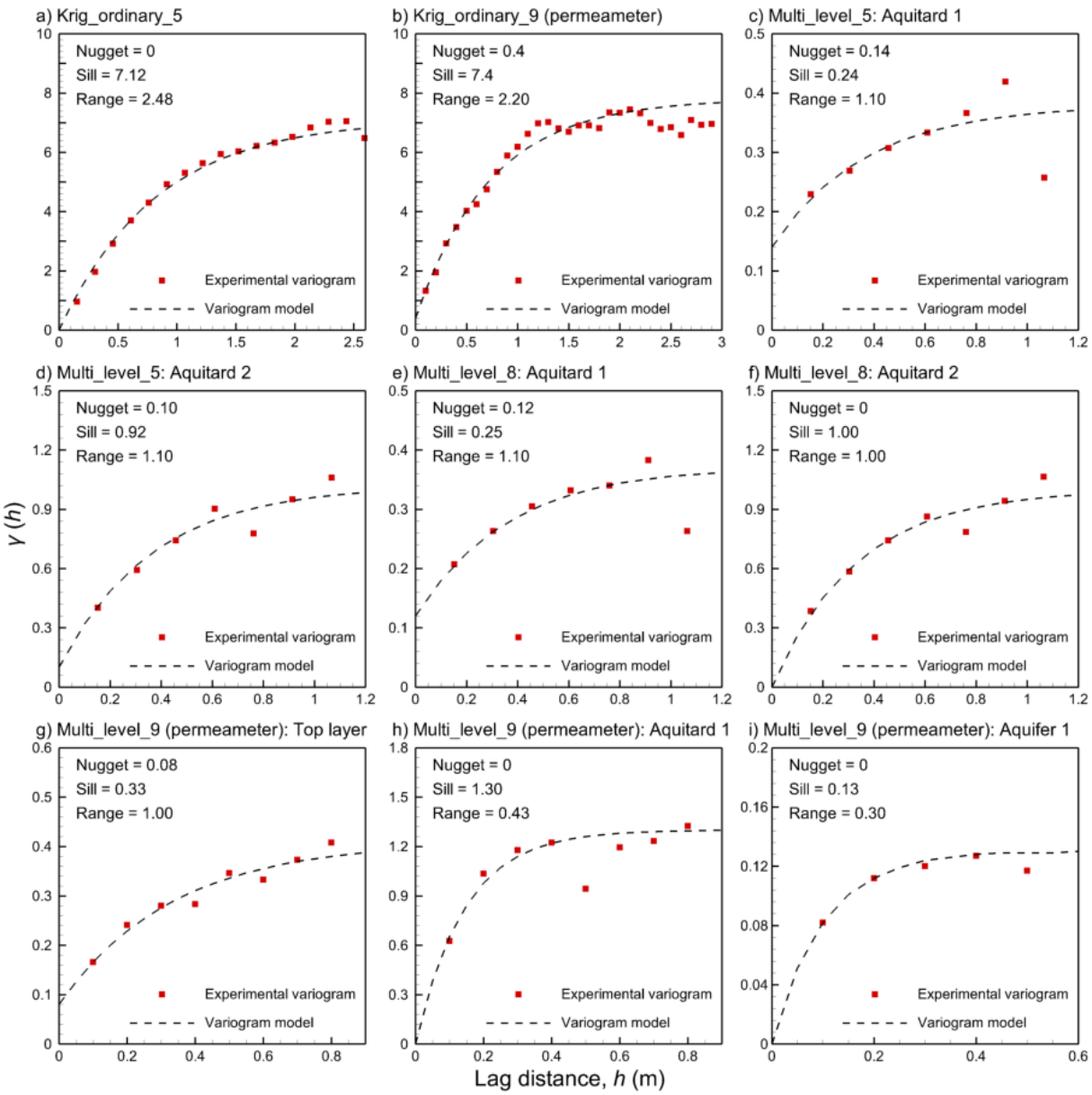


Figure B5: The vertical experimental variograms of ln *K* observations and the fitted variogram models for different cases. Parts a) – b) show the variogram fitting for two ordinary kriged models (i.e., *Krig_ordinary_5* and *Krig_ordinary_9 (permeameter)*). Parts c) – i) illustrate the variogram fitting for different layers of multi-level heterogeneity models, including *Multi-level_5*, *Multi-level_8*, and *Multi-level_9 (permeameter)*. Variogram models were fitted for all layers with a sufficient number of observations. The fitted models in d), f), and h) were adopted for the other layers of the same *K* model, which have insufficient observations for variogram analysis. In terms of multi-level heterogeneity representations, their omnidirectional horizontal variogram models took the structure of vertical models, while their horizontal correlation lengths remain the same as the ensemble variogram models.

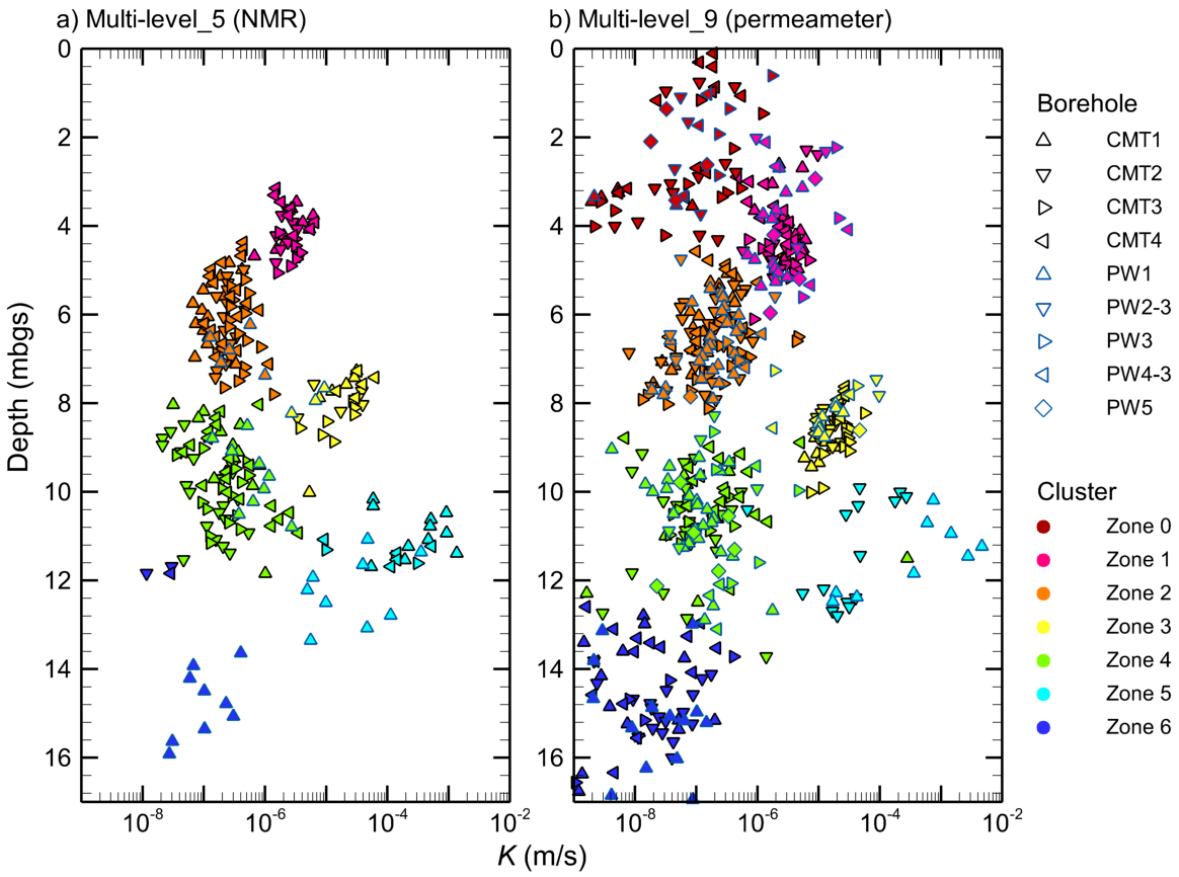


Figure B6: Clustered *K* profiles for the zonation of Cases *Multi-level_5* and *Multi-level_9 (permeameter)*. *K* observations in different clusters are filled in contrasting colors for visualization. The spatially associated clustering of five NMR logs and nine boreholes of permeameter measurements derived six and seven clusters, respectively. However, the distributions of clustered *K* are consistent between these two datasets. The uppermost cluster of *K* observations for *Multi-level_9 (permeameter)* (i.e., Zone 0) is located above the water table, while only the NMR data below the water table are included for this study.

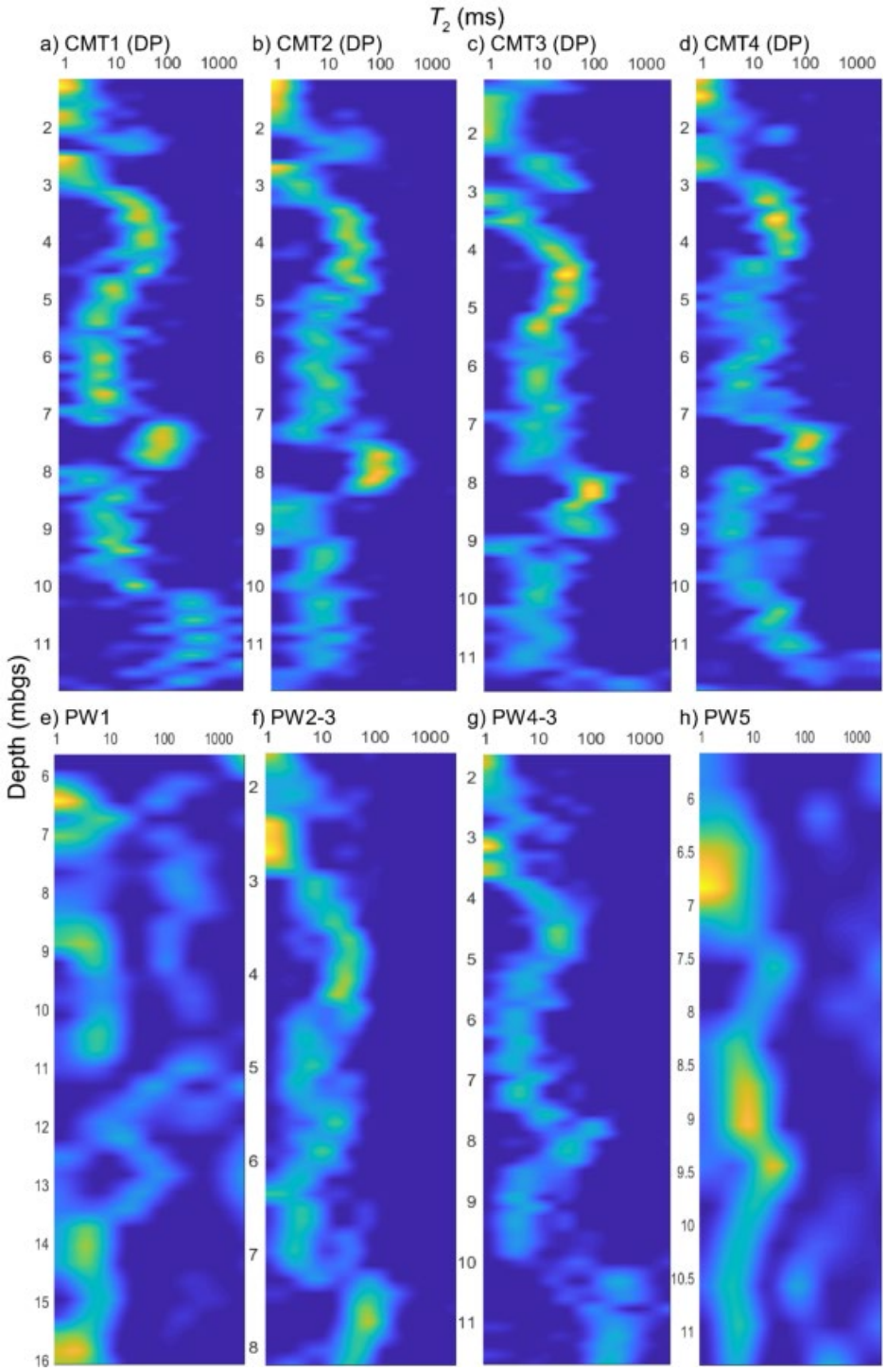


Figure B7: Relaxation time distributions of eight NMR logs. Parts a) – h) correspond to the NMR logging in CMT1, CMT2, CMT3, CMT4, PW1, PW2-3, PW4-3, and PW5. The PW1 and PW5 logs were conducted in the 4" multi-screened wells, showing bi-modal distributions due to different local sorting, e.g., the locations of Aquifers 1 and 2 at ~8 and 11 mbgs.

Table B1: Geostatistical model parameters for the vertical indicator variogram.

Threshold K (m/s)	Num of obs.	F(z)	Nugget	Sill	Range (m)	Variogram model
5.0×10^{-8}	20	0.055	0.008	0.052	1.3	Exponential
1.0×10^{-7}	46	0.127	0.038	0.092	1.7	Exponential
5.0×10^{-7}	191	0.529	0.020	0.274	1.4	Spherical
1.0×10^{-6}	214	0.593	0.000	0.310	1.7	Spherical
5.0×10^{-6}	282	0.781	0.000	0.200	2.0	Exponential
1.0×10^{-5}	302	0.837	0.010	0.150	2.2	Exponential
5.0×10^{-5}	331	0.917	0.006	0.077	4.4	Exponential
1.0×10^{-4}	336	0.931	0.005	0.055	2.8	Exponential
5.0×10^{-4}	355	0.983	0.011	0.006	0.7	Spherical

* F(z) provides the probability of observations not exceeding this threshold value.

Appendix C: Supplementary Information for Study IV: Integrating NMR logging with Hydraulic Tomography for Improved Mapping of Hydraulic Conductivity

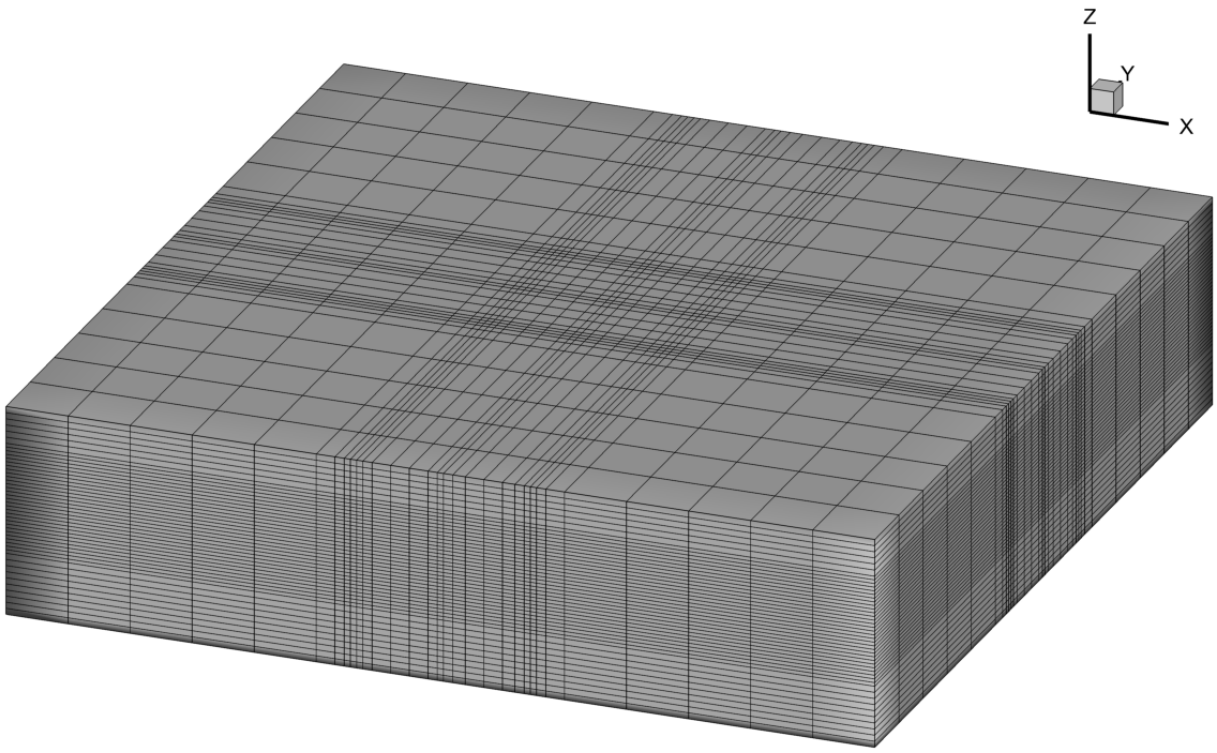


Figure C1: Grid discretization for the groundwater flow modeling in HydroGeoSphere. The top boundary is assigned to be 340.4 masl, with a thick aquitard layer treated as the bottom boundary. In total, 44206 elements were generated in model discretization, and the block size decreases gradually from 5 m × 5 m × 0.5 m near the boundaries to 0.5 m × 0.5 m × 0.3 m around the pumping and observation wells.

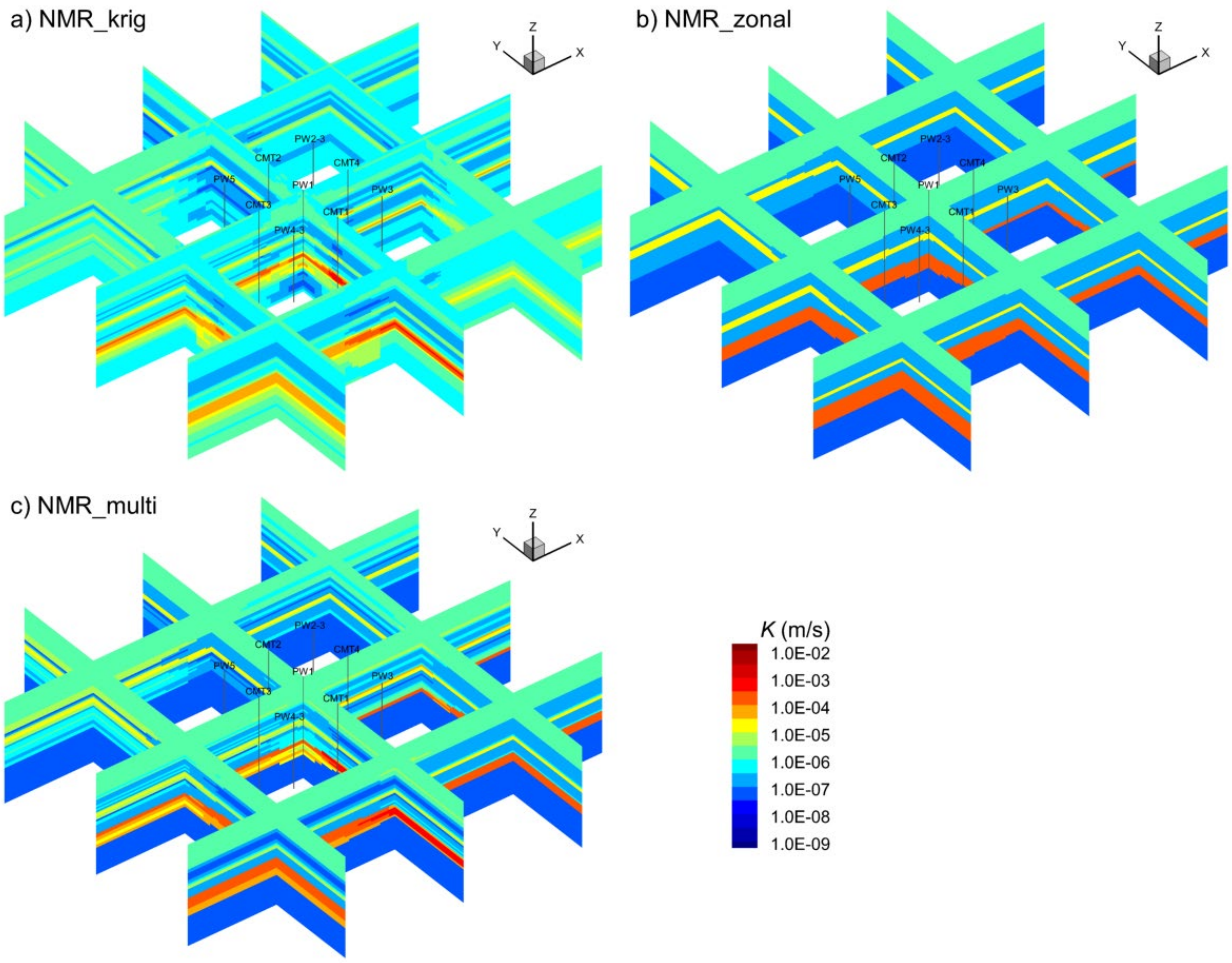


Figure C2: NMR-derived *K* fields for the three NMR cases in the simulation domain: a) *NMR_krig*, b) *NMR_zonal*, and c) *NMR_multi*.

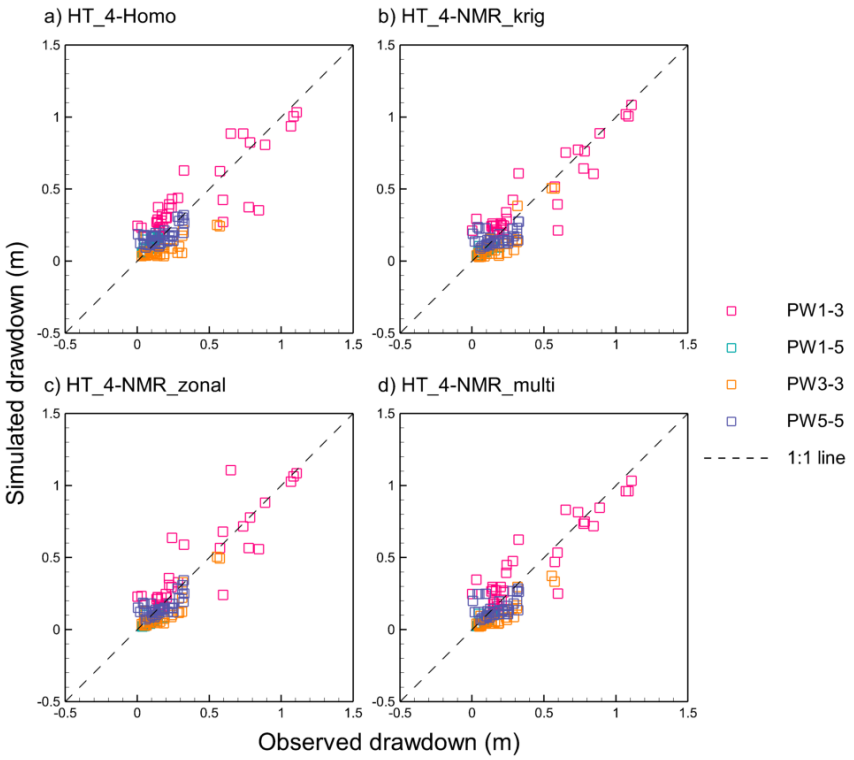


Figure C3: Calibration scatterplots of simulated versus observed drawdowns for the *HT_4* scenarios.

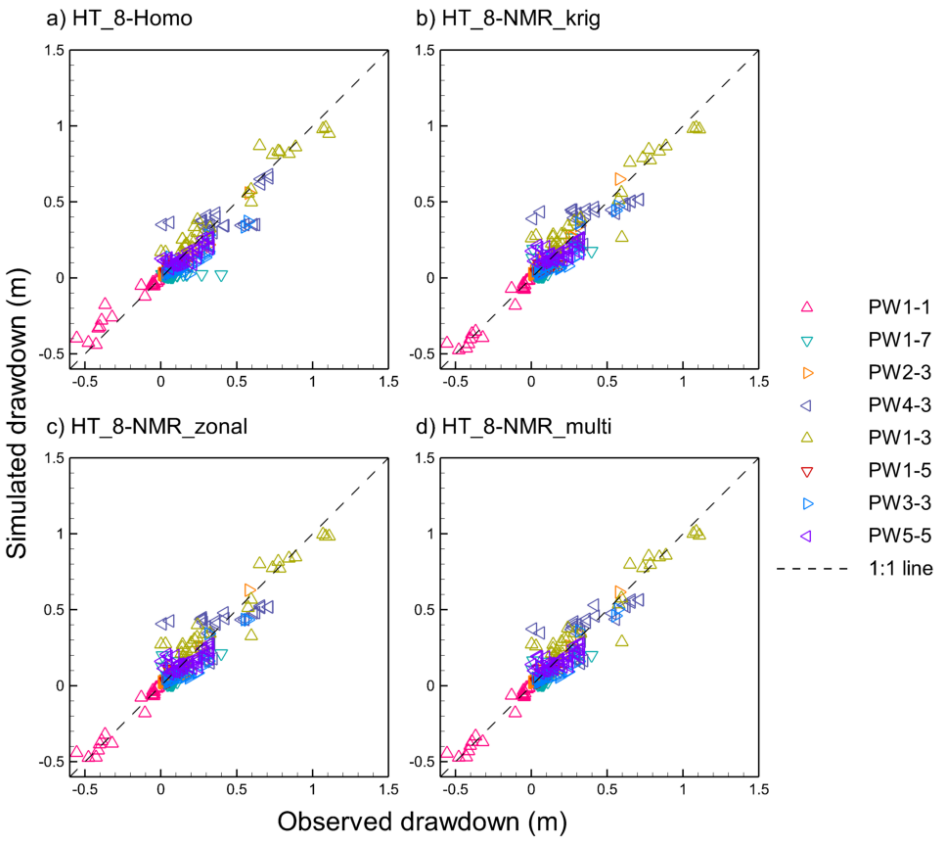


Figure C4: Calibration scatterplots of simulated versus observed drawdowns for the *HT_8* scenarios.

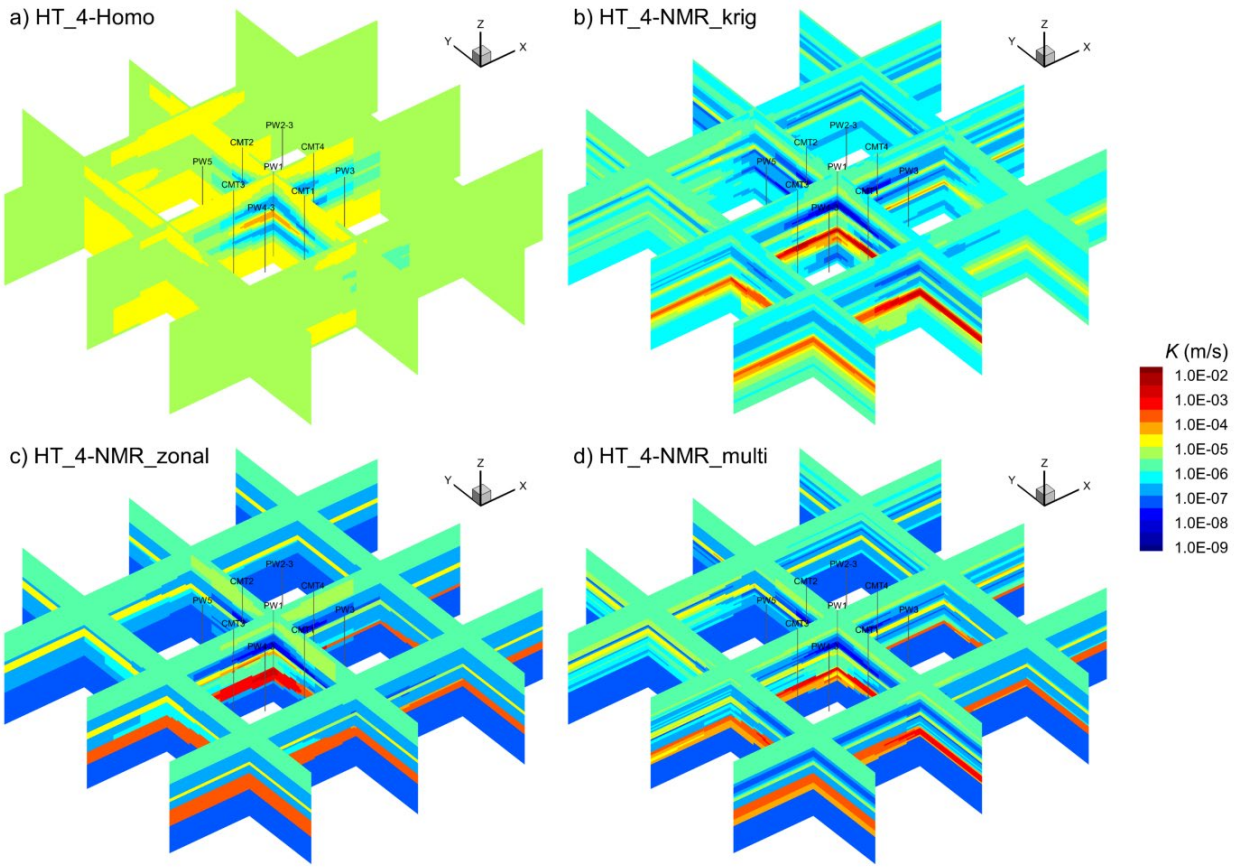


Figure C5: HT-estimated *K* fields in the simulation domain. Four pumping tests were used for model calibration. Comparative cases include: a) *HT_4-Homo*, b) *HT_4-NMR_krig*, c) *HT_4-NMR_zonal*, and d) *HT_4-NMR_multi*.

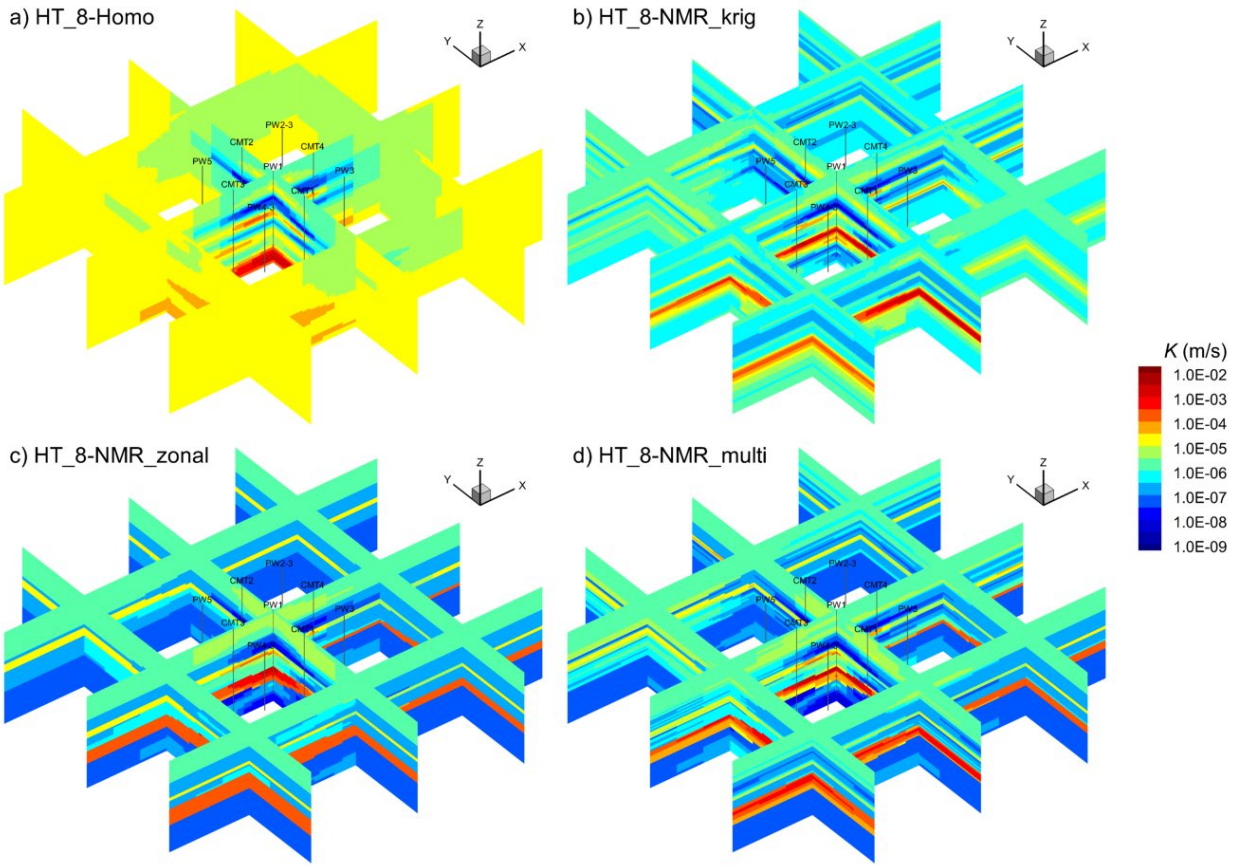


Figure C6: HT-estimated *K* fields in the simulation domain. Eight pumping tests were used for model calibration. Comparative cases include: a) *HT_8-Homo*, b) *HT_8-NMR_krig*, c) *HT_8-NMR_zonal*, and d) *HT_8-NMR_multi*.

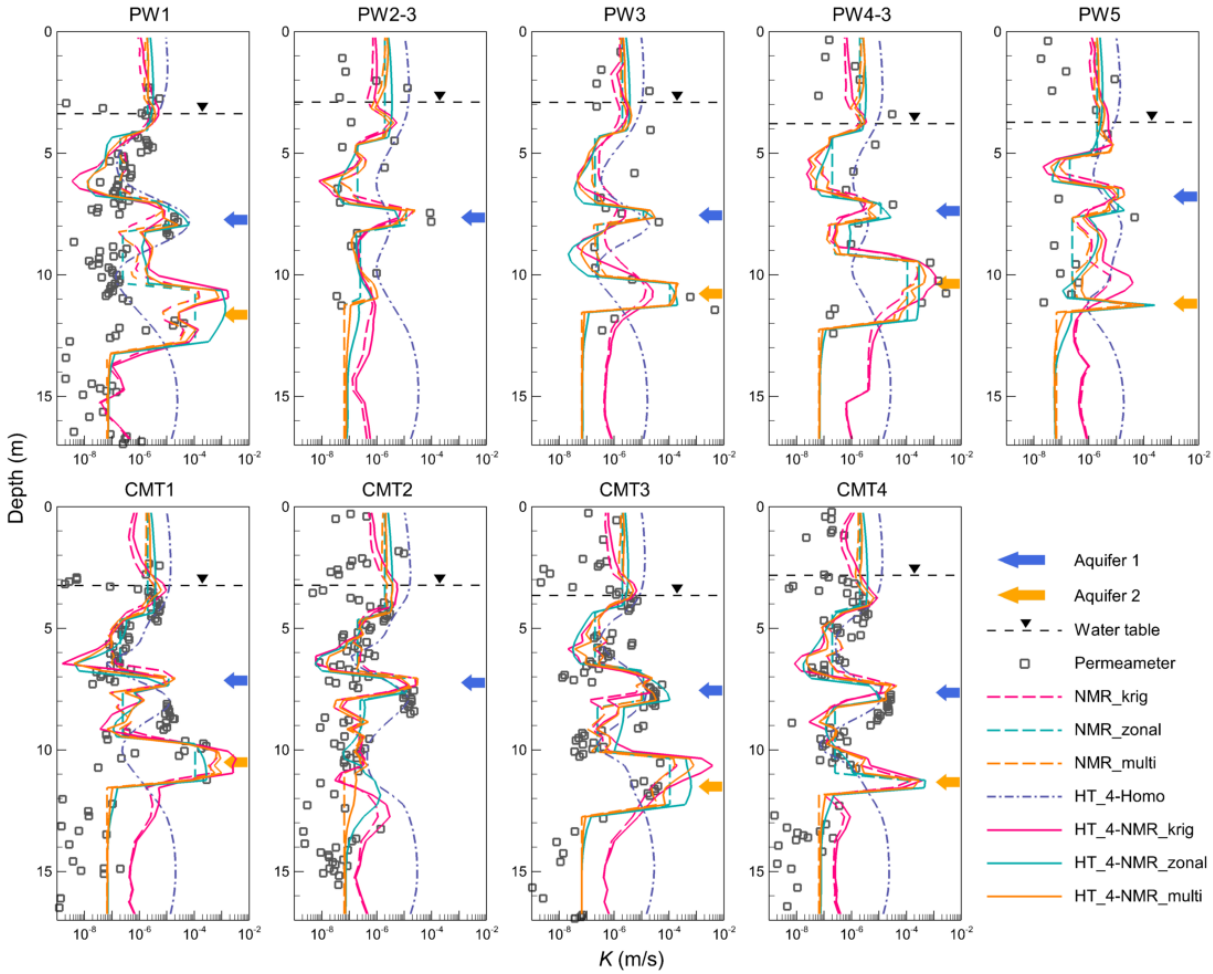


Figure C7: Comparison of *K* profiles at nine well locations between the model-derived *K* and the permeameter *K* measurements. The HT and the integrated models using the limited calibration dataset (i.e., *HT_4*), as well as the NMR models, are included in the comparison. Aquifers 1 and 2 are denoted using colored arrows.

Table C1: The drawdown records from the "-6" and "-7" observation intervals of the pumping tests used for model calibration.

ID	Drawdown (m)	Pumping locations	Stress depth (mbgs)	Pumping rate (L/min)	Observation intervals	Calibration datasets
1	0.103	PW1-5	11.46	4.4	CMT4-6	*HT_4* and *HT_8*
2	0.031	PW3-3	7.62	2.1	CMT2-6	*HT_4* and *HT_8*
3	0.02	PW5-5	11.61	8.1	CMT1-7	*HT_4* and *HT_8*
4	0.095	PW5-5	11.61	8.1	CMT2-6	*HT_4* and *HT_8*
5	0.040	PW5-5	11.61	8.1	CMT2-7	*HT_4* and *HT_8*
6	0.005	PW5-5	11.61	8.1	CMT3-6	*HT_4* and *HT_8*
7	0.046	PW5-5	11.61	8.1	CMT4-6	*HT_4* and *HT_8*
8	0.269	PW1-7	15.56	1.05	CMT1-6	*HT_8*
9	0.268	PW1-7	15.56	1.05	CMT1-7	*HT_8*
10	0.17	PW1-7	15.56	1.05	CMT2-6	*HT_8*
11	0.104	PW1-7	15.56	1.05	CMT2-7	*HT_8*
12	0.397	PW1-7	15.56	1.05	CMT3-6	*HT_8*
13	0.011	PW1-7	15.56	1.05	CMT3-7	*HT_8*
14	0.148	PW1-7	15.56	1.05	CMT4-6	*HT_8*
15	0.007	PW1-7	15.56	1.05	CMT4-7	*HT_8*
16	0.275	PW4-3	11.55	30.2	CMT1-6	*HT_8*
17	0.276	PW4-3	11.55	30.2	CMT1-7	*HT_8*
18	0.423	PW4-3	11.55	30.2	CMT2-6	*HT_8*
19	0.282	PW4-3	11.55	30.2	CMT2-7	*HT_8*
20	0.267	PW4-3	11.55	30.2	CMT3-6	*HT_8*
21	0.061	PW4-3	11.55	30.2	CMT3-7	*HT_8*
22	0.556	PW4-3	11.55	30.2	CMT4-6	*HT_8*
23	0.01	PW4-3	11.55	30.2	CMT4-7	*HT_8*

Appendix D: Supplementary Information for Study V: Three-Dimensional Analysis of Heat Tracer Transport with Hydraulic Conductivity Distributions from Various Approaches in a Complex Aquifer System

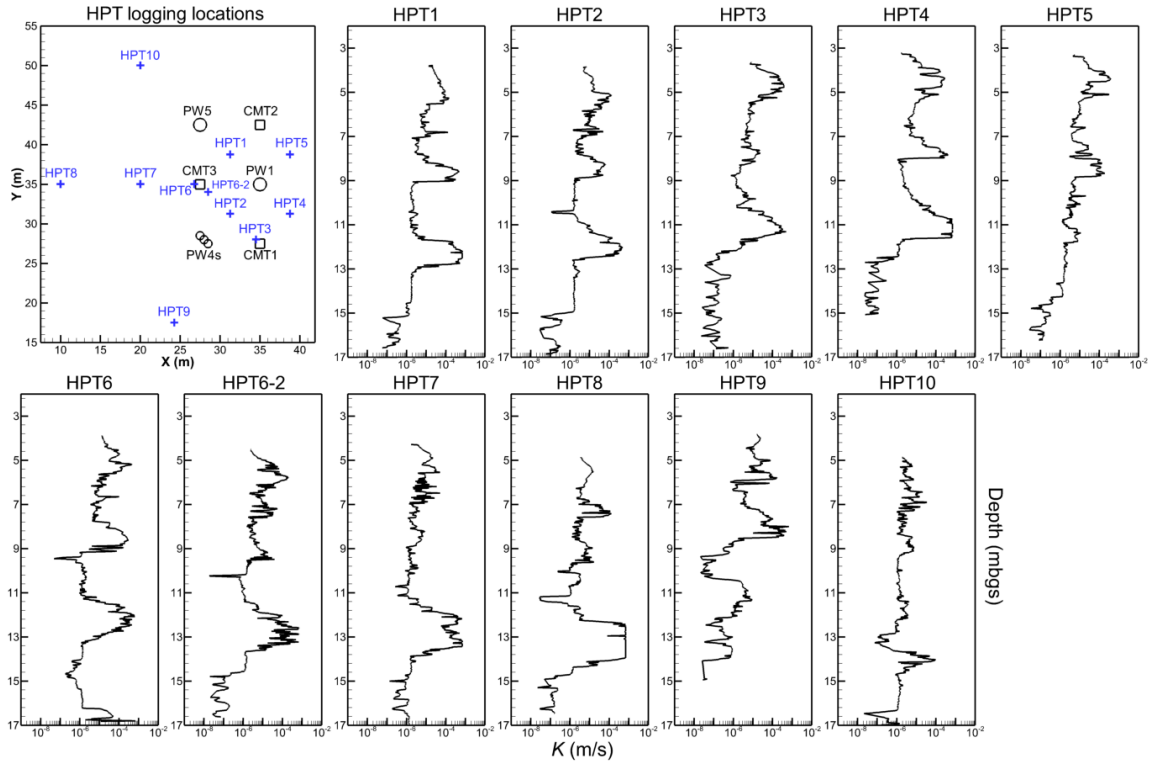


Figure D1: HPT-derived *K* profiles at 11 borehole locations, as indicated in the well map. *K* values were estimated using the site-specifically optimized equation developed by Zhao and Illman (2022b).

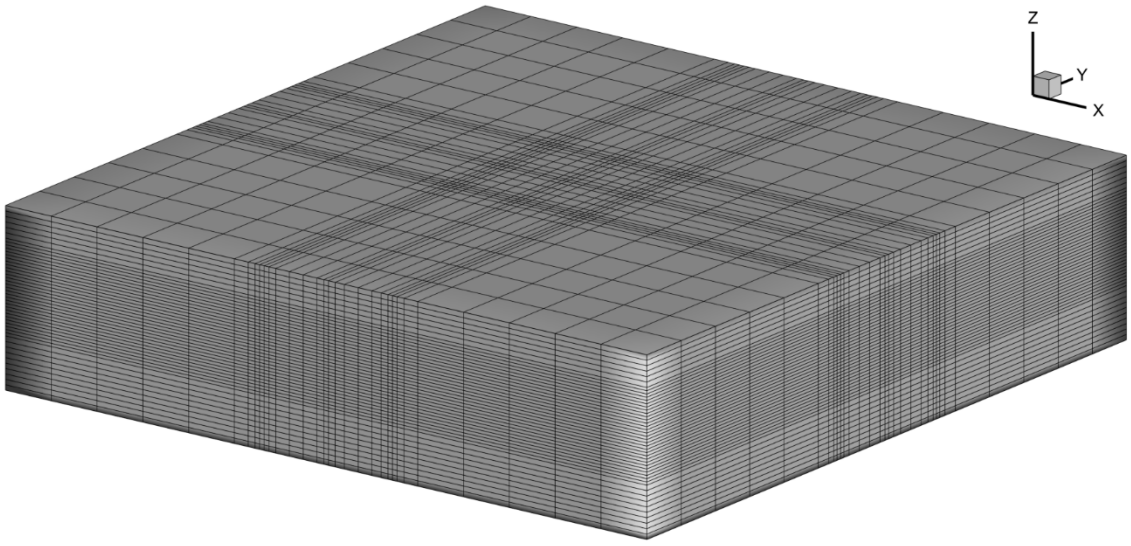


Figure D2: Grid discretization for the groundwater flow and heat transport modeling in HydroGeoSphere. The top boundary is assigned to be 340.4 m above see level, with a thick aquitard layer treated as the bottom boundary. In total, 44206 elements were generated in model discretization, and the block size decreases gradually from 5 m × 5 m × 0.5 m near the boundaries to 0.5 m × 0.5 m × 0.3 m around the pumping and observation wells.

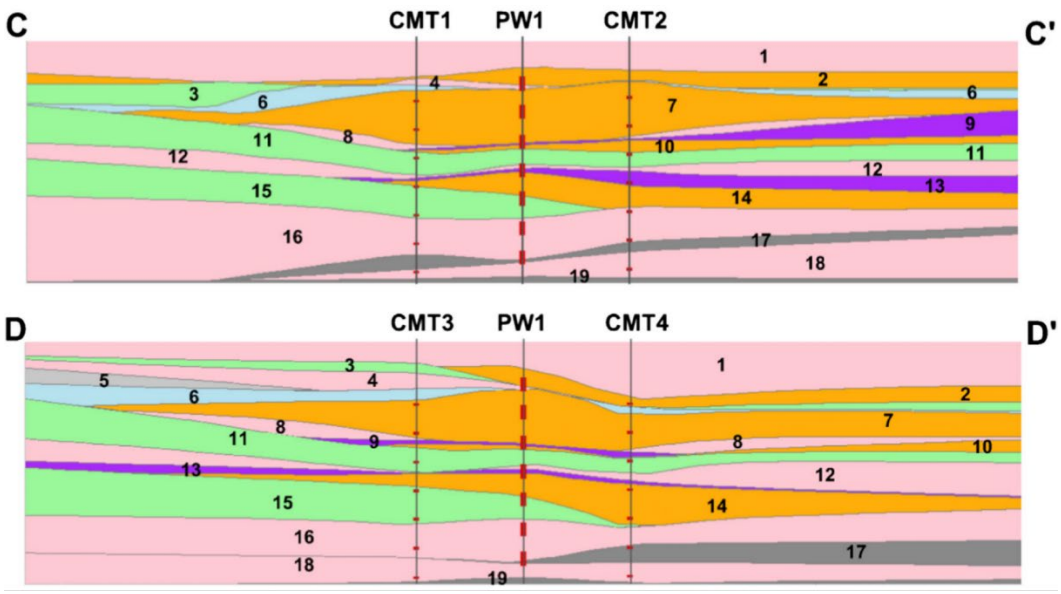


Figure D3: C-C' and D-D' cross-sectional views of the 19-layer geological zonation model by Zhao and Illman (2017). Numbers in cross-sections indicate the 19 layers of different soil materials: Clay (1, 4, 8, 12, 16, 18); Silt and Clay (17, 19); Silt (2, 7, 10, 14); Sandy Silt (6, 9, 13); Sand and Silt (5); Sand (3, 11); Sand and Gravel (15).

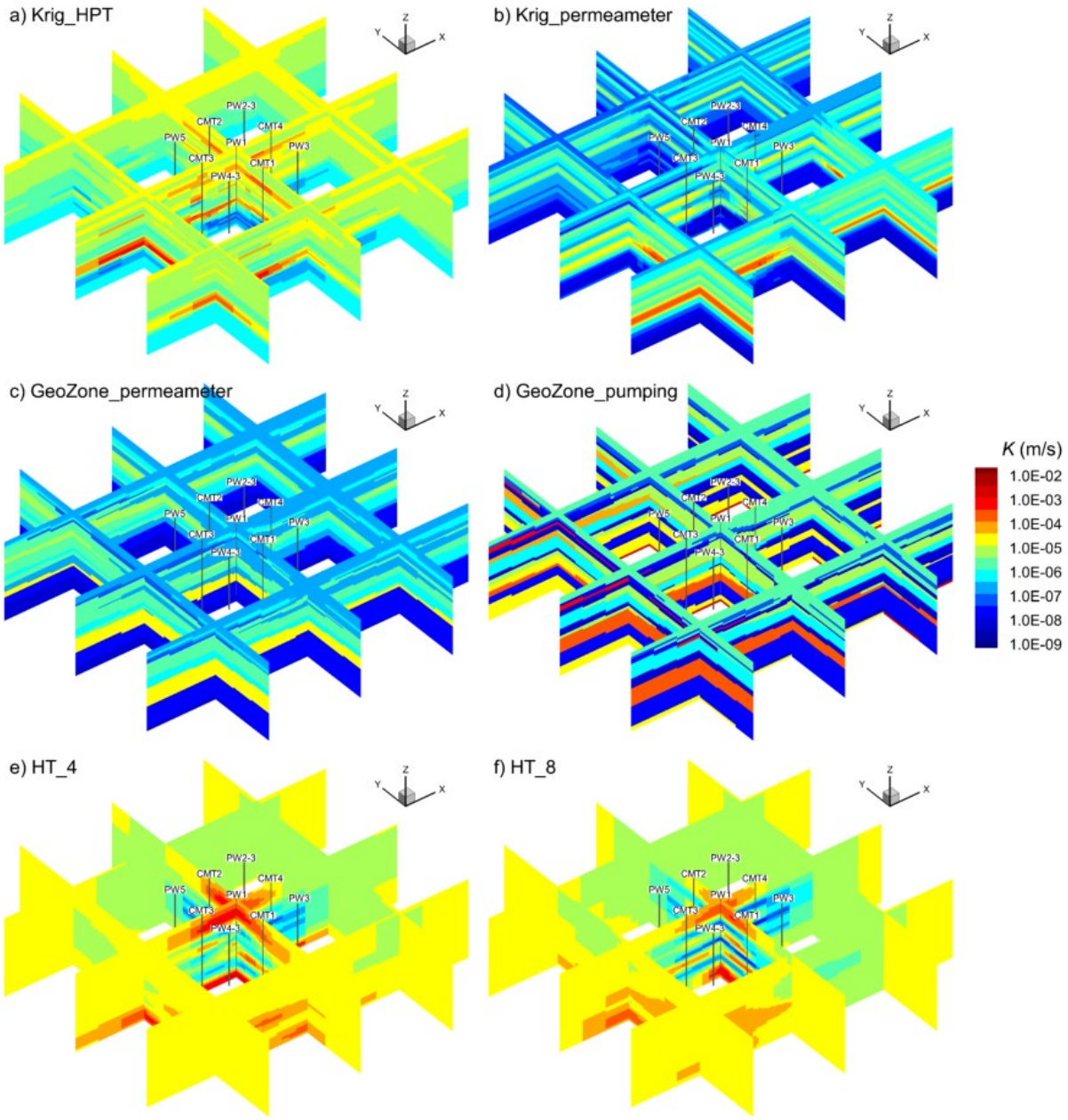


Figure D4: Fence diagrams of the comparative *K* fields in the simulation domain: a) *Krig_HPT*, b) *Krig_permeameter*, c) *GeoZone_permeameter*, d) *GeoZone_pumping*, e) *HT_4*, and f) *HT_8*.

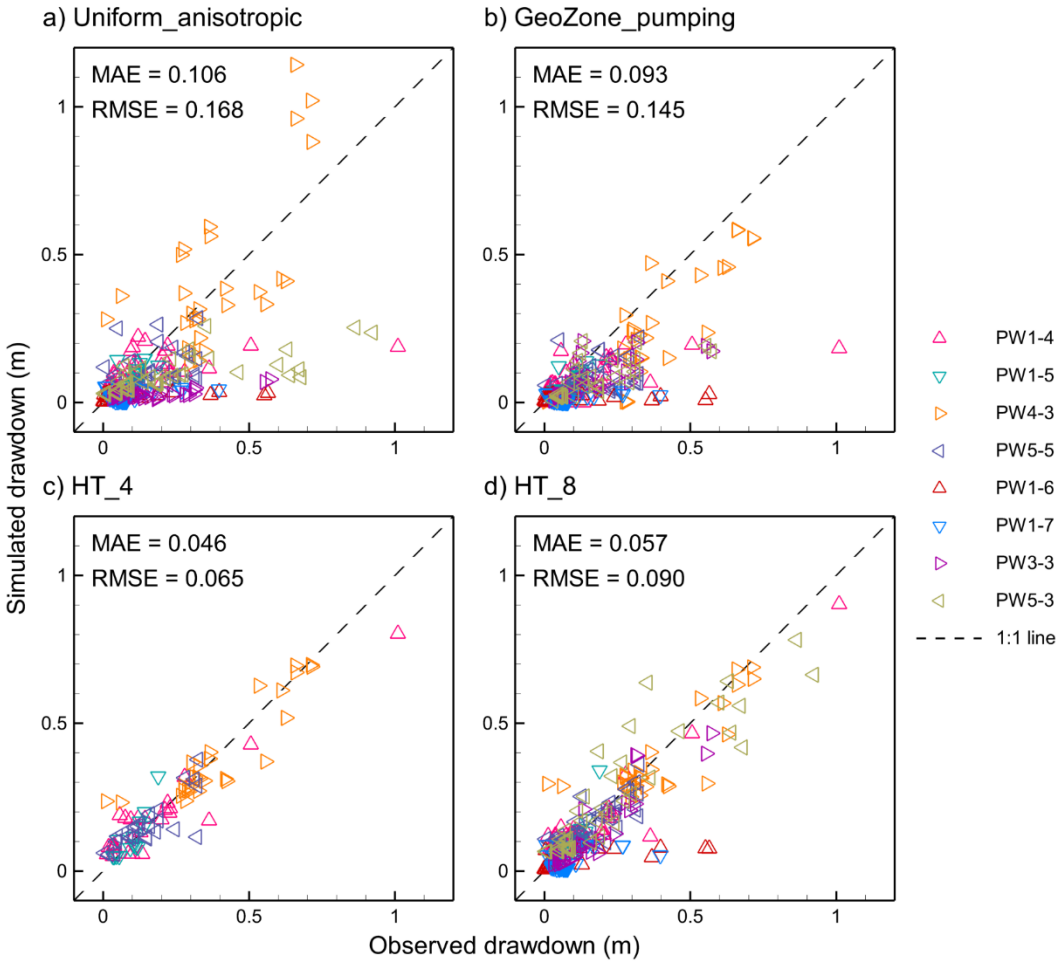


Figure D5: Scatterplots of simulated versus observed drawdowns for model calibrations for the inverse models a) *Uniform_anisotropic*, b) *GeoZone_pumping*, c) *HT_4* and the d) *HT_8* , respectively. Note that limited observation data from four pumping tests were used to calibrate the c) *HT_4* model, while all observation data from eight pumping tests were used to calibrate models a) *Uniform_anisotropic*, b) *GeoZone_pumping*, and d) *HT_8*.

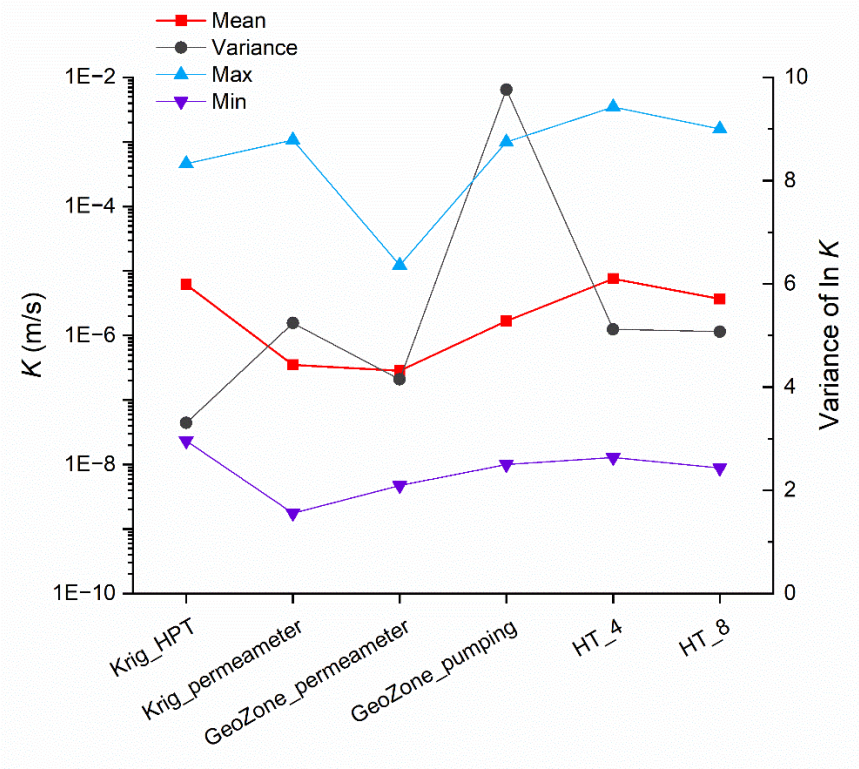


Figure D6: The statistics of the comparative *K* fields within the data domain, including the geometric mean of *K*, the variance of ln *K*, and the maximum and the minimum *K*.

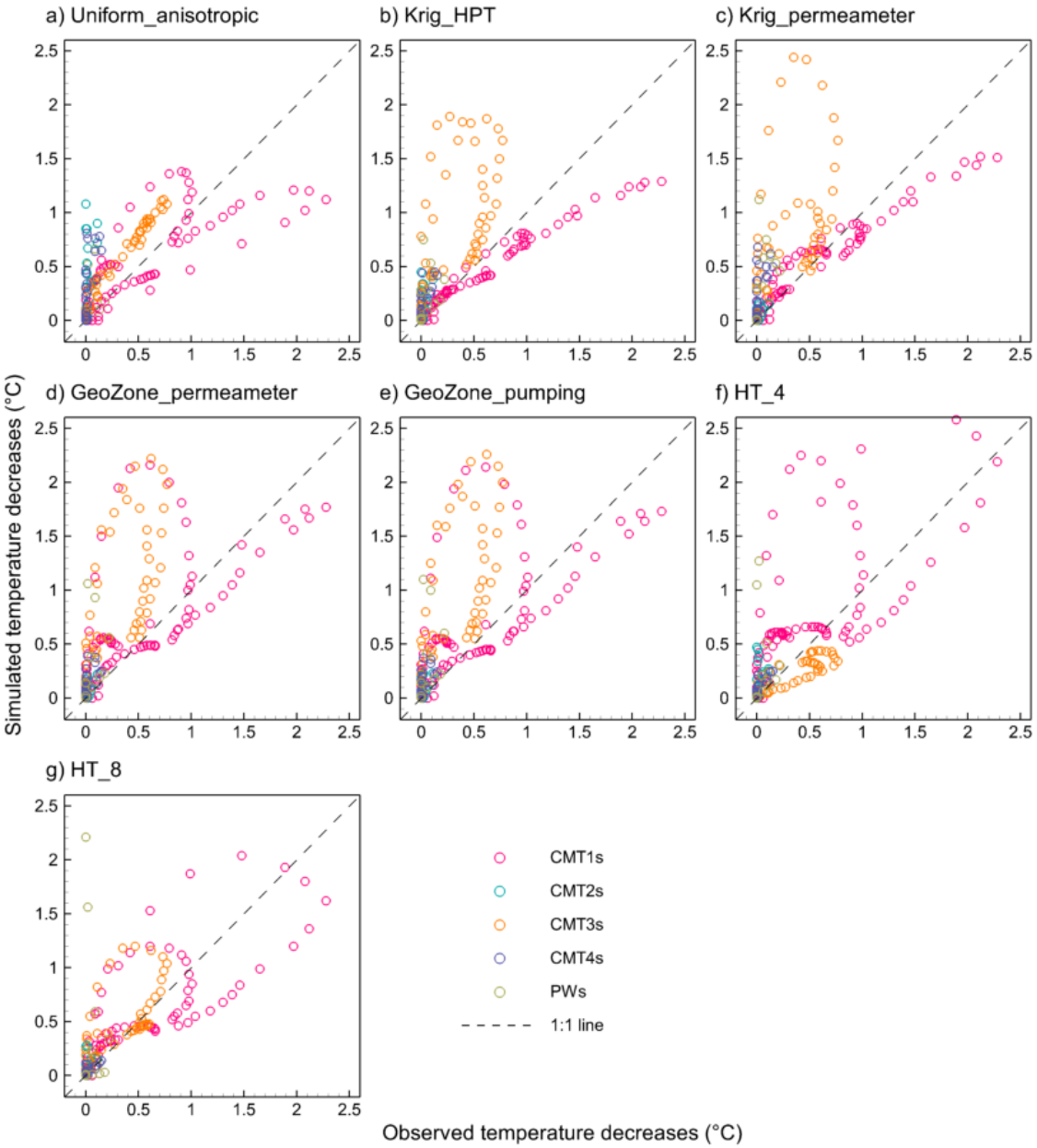


Figure D7: The scatterplots of temperature decreases between model simulations and field observations at selected times. Various *K* representations are used for numerical simulations of heat transport, including a) *Uniform_anisotropic*, b) *Krig_HPT*, c) *Krig_permeameter*, d) *GeoZone_permeameter*, e) *GeoZone_pumping*, f) *HT_4*, and g) *HT_8*.

Table D1: Detailed records of the flow rates for pumping (PW4-3) and injection (PW1-5). These detailed records of flow rates were used in numerical models of transport simulations.

Location	Start Time	End Time	Duration (h)	Flow rates (L/min)
PW4-3	2024-11-28 15:00	2024-11-29 9:00	18.0	8.014
PW4-3	2024-11-29 9:00	2024-12-03 6:45	93.8	7.867
PW4-3	2024-12-03 6:45	2024-12-03 8:50	2.1	0.000
PW4-3	2024-12-03 8:50	2024-12-07 1:45	88.9	7.935
PW4-3	2024-12-07 1:45	2024-12-07 8:45	7.0	0.000
PW4-3	2024-12-07 8:45	2024-12-07 12:30	3.8	7.888
PW4-3	2024-12-07 12:30	2024-12-07 13:05	0.6	0.000
PW4-3	2024-12-07 13:05	2024-12-07 23:38	10.5	8.122
PW4-3	2024-12-07 23:38	2024-12-08 0:15	0.6	0.000
PW4-3	2024-12-08 0:15	2024-12-12 10:00	105.8	8.103
PW4-3	2024-12-12 10:00	2025-01-16 5:55	835.9	8.190
Location	Start Time	End Time	Duration (h)	Flow rates (L/min)
PW1-5	2024-11-28 15:00	2024-11-29 9:00	18.00	8.014
PW1-5	2024-11-29 9:00	2024-12-02 1:30	64.50	7.915
PW1-5	2024-12-02 1:30	2024-12-02 6:50	5.33	2.000
PW1-5	2024-12-02 6:50	2024-12-02 11:30	4.67	8.143
PW1-5	2024-12-02 11:30	2024-12-02 12:20	0.83	0.000
PW1-5	2024-12-02 12:20	2024-12-02 13:00	0.67	7.914
PW1-5	2024-12-02 13:00	2024-12-02 14:20	1.33	0.000
PW1-5	2024-12-02 14:20	2024-12-03 6:45	16.42	7.934
PW1-5	2024-12-03 6:45	2024-12-03 8:50	2.08	0.000
PW1-5	2024-12-03 8:50	2024-12-04 18:40	33.83	8.074
PW1-5	2024-12-04 18:40	2024-12-04 19:30	0.83	0.000
PW1-5	2024-12-04 19:30	2024-12-05 14:00	18.50	8.154
PW1-5	2024-12-05 14:00	2024-12-05 14:40	0.67	0.000
PW1-5	2024-12-05 14:40	2024-12-05 23:34	8.90	8.232
PW1-5	2024-12-05 23:34	2024-12-06 2:20	2.77	0.000
PW1-5	2024-12-06 2:20	2024-12-07 1:45	23.42	7.965
PW1-5	2024-12-07 1:45	2024-12-07 8:45	7.00	0.000
PW1-5	2024-12-07 8:45	2024-12-07 12:30	3.75	7.961
PW1-5	2024-12-07 12:30	2024-12-07 13:05	0.58	0.000
PW1-5	2024-12-07 13:05	2024-12-12 7:45	114.67	8.004
PW1-5	2024-12-12 7:45	2024-12-12 10:00	2.25	0.000
PW1-5	2024-12-12 10:00	2025-01-16 5:55	835.9	8.190

Table D2: Estimated *K* values and corresponding posterior 95% confidence intervals for the *GeoZone_pumping* model.

Operation screens	Stress depth* (mbgs)	Stress type	Rates (L/min)	Total number of heads
PW1-4	9.46	Pumping	6.3	37 (27†)
PW1-5	11.46	Pumping	4.4	28 (14†)
PW4-3	11.55	Pumping	30.2	28 (28†)
PW5-5	11.61	Pumping	8.1	41 (30†)
PW1-6	13.56	Pumping	0.95	31
PW1-7	15.56	Pumping	1.05	34
PW3-3	7.62	Pumping	2.1	31
PW5-3	7.62	Pumping	7.8	42

* It indicates the mid-depths of the pumping screens for the stress depth.

† It denotes the number of head observations for the limited HT calibration dataset.

Table D3: Estimated *K* values and corresponding posterior 95% confidence intervals for the *GeoZone_pumping* model.

Layers	Estimated *K* (m/s)	95% confidence limits	
		Lower (m/s)	Upper (m/s)
1	3.14E-06	1.88E-07	5.25E-05
2	5.70E-08	1.35E-12	2.41E-03
3	2.03E-06	2.79E-10	1.48E-02
4	1.07E-08	5.21E-14	2.19E-03
5	5.39E-04	8.78E-06	3.31E-02
6	1.00E-08	3.90E-11	2.56E-06
7	7.16E-06	1.88E-06	2.73E-05
8	4.90E-06	9.94E-10	2.41E-02
9	7.94E-05	4.36E-05	1.44E-04
10	6.44E-07	4.04E-09	1.03E-04
11	5.90E-07	1.54E-07	2.26E-06
12	1.00E-08	7.04E-13	1.42E-04
13	5.56E-06	7.83E-08	3.95E-04
14	2.42E-05	4.85E-07	1.21E-03
15	1.22E-04	7.41E-05	1.99E-04
16	1.16E-08	2.15E-10	6.29E-07
17	1.00E-08	2.70E-10	3.71E-07
18	1.16E-05	2.13E-12	6.32E+01
19	1.00E-03	3.87E-05	2.58E-02

Table D4: The mean arrival time (μ) (day) of the tracer strength and the variance (σ^2) (day^2) of temperature breakthrough curves for both field observations and model simulations. The observation ports with significant observed temperature decreases (over 0.3 °C) and the effluent port (PW4-3) are incorporated. These ports are horizontally arranged in ascending order of computed values for the observed mean arrival time and in descending order for the observed variance.

Mean Arrival Time (day)	CMT1-4	CMT3-4	CMT1-5	CMT3-5	PW4-3	CMT1-3	CMT1-6	Comparison	
Field observations	28.79	31.52	33.10	33.20	36.10	37.69	38.74	MAE	RMSE
Uniform_anisotropic	31.94	30.00	31.50	31.46	33.32	36.66	36.23	2.05	2.17
Krig_HPT	29.09	26.52	31.67	25.88	26.92	36.27	37.19	3.74	4.92
Krig_permeameter	29.16	23.60	30.83	28.61	29.79	35.03	36.94	3.70	4.46
GeoZone_permeameter	27.85	26.24	25.80	27.25	28.36	35.83	33.93	4.84	5.40
GeoZone_pumping	28.61	27.13	26.88	28.17	28.65	36.20	34.43	4.15	4.77
HT_4	24.07	33.76	25.41	34.92	23.84	33.91	34.37	5.25	6.24
HT_8	23.94	26.34	27.70	32.35	24.44	33.66	35.59	5.02	5.88
Variance (day^2)	CMT1-4	CMT3-4	CMT1-5	CMT3-5	CMT1-3	CMT1-6	PW4-3	Comparison	
Field observations	91.70	86.06	78.17	75.78	66.84	56.86	54.70	MAE	RMSE
Uniform_anisotropic	84.92	93.05	86.33	88.73	64.43	66.27	80.98	10.43	12.62
Krig_HPT	95.46	100.75	86.94	101.61	66.86	60.86	97.39	14.25	20.05
Krig_permeameter	92.72	103.16	90.77	100.46	72.61	62.02	93.69	15.05	19.43
GeoZone_permeameter	92.77	95.51	93.92	95.57	67.86	78.13	89.64	14.76	18.53
GeoZone_pumping	92.06	94.97	94.24	94.88	66.09	75.91	89.40	14.14	18.01
HT_4	97.97	84.99	98.36	76.97	79.88	77.00	86.77	13.42	17.13
HT_8	105.17	117.39	101.23	91.98	82.95	71.30	90.35	21.47	22.98

Table D5: The tracer total response strength (M_0), the mean arrival time (μ), and the variance (σ^2) of the simulated heat tracer transport using the *HT_8* model for *K* representation and various values of transport and thermal parameters, including "Total porosity field", "Larger dispersion", "No dispersion", "Stronger thermal conduction", and "Weaker thermal conduction" scenarios.

Total Strength (°C·day)	CMT14	CMT15	CMT34	CMT35	CMT13	CMT16	PW43	Comparison	
Field observations	6.05	3.00	2.31	1.73	1.34	0.50	0.37	MAE	RMSE
HT_8	5.22	3.61	3.76	1.57	1.41	0.86	3.62	0.96	1.41
Total porosity field	4.65	2.88	3.07	1.01	0.72	0.37	3.52	0.99	1.38
Larger dispersion	5.45	3.97	4.83	2.17	1.64	1.06	3.56	1.23	1.62
No dispersion	5.17	3.49	2.97	1.22	1.30	0.78	3.45	0.85	1.27
Stronger thermal conduction	4.74	3.43	3.47	1.66	1.58	1.03	3.40	0.97	1.35
Weaker thermal conduction	5.87	3.80	4.20	1.41	1.11	0.60	3.87	1.01	1.54
Mean Arrival Time (day)	CMT14	CMT34	CMT15	CMT35	PW43	CMT13	CMT16	Comparison	
Field observations	28.79	31.52	33.10	33.20	36.10	37.69	38.74	MAE	RMSE
HT_8	23.94	26.34	27.70	32.35	24.44	33.66	35.59	5.02	5.88
Total porosity field	26.45	28.50	30.49	34.69	25.78	36.67	38.23	3.04	4.33
Larger dispersion	20.87	21.90	25.02	29.90	24.26	32.11	34.22	7.26	7.77
No dispersion	25.84	30.84	29.35	34.38	25.14	34.64	36.40	3.56	4.77
Stronger thermal conduction	23.98	26.15	27.17	31.46	24.53	32.50	34.37	5.57	6.21
Weaker thermal conduction	23.85	26.63	28.50	33.60	24.38	35.30	37.18	4.36	5.54
Variance (day^2)	CMT14	CMT34	CMT15	CMT35	CMT13	CMT16	PW43	Comparison	
Field observations	91.70	86.06	78.17	75.78	66.84	56.86	54.70	MAE	RMSE
HT_8	105.17	117.39	101.23	91.98	82.95	71.30	90.35	21.47	22.98
Total porosity field	102.54	111.83	91.97	78.14	64.70	54.04	92.95	13.71	18.72
Larger dispersion	110.99	119.23	112.05	104.19	93.46	81.51	93.02	29.19	29.79
No dispersion	92.72	97.19	90.58	80.84	75.84	65.13	93.35	12.22	16.68
Stronger thermal conduction	106.99	117.65	103.67	96.52	89.35	78.58	90.86	24.79	25.63
Weaker thermal conduction	102.33	116.73	97.58	84.99	73.19	61.10	89.97	16.54	20.06