Hydraulic Tomography Analyses of Different Datasets for Subsurface Heterogeneity Characterization at Various Scales

by

Ning Luo

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

Waterloo, Ontario, Canada, 2023

© Ning Luo 2023

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:

Kent S. Novakowski

Associate Vice-Principal Research / Professor of Civil Engineering, Queen's University

Supervisor:

Walter A. Illman

Professor of Earth and Environmental Sciences, University of Waterloo

Internal Members:

David L. Rudolph

Professor of Earth and Environmental Sciences, University of Waterloo

Steven J. Berg

President and CEO, Aquanty, Inc.

Internal-external Member:

Bryan A. Tolson

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.

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

Statement of Contributions

My doctoral research journey has been marked by meaningful contributions to the advancement of subsurface heterogeneity characterization through the innovative integration of diverse data for hydraulic tomography (HT) analyses. By collaborating with esteemed colleagues, I have enriched the field's understanding of subsurface properties across scales, with a focus on groundwater management and environmental assessments. This thesis consists in part of four manuscripts written for publication. Exceptions to sole authorship of material are as follows:

1. **Research presented in Chapter 3:**

   In collaboration with fellow researchers, I investigated the impact of geological information on inverse modeling. Comprehensive comparisons of different modeling approaches were carried out to evaluate their performances and guide field application of HT with geological information.

   Citation:

   Luo, N., Zhao, Z., Illman, W.A., Berg, S.J., 2017. Comparative study of transient hydraulic tomography with varying parameterization and zonations: Laboratory sandbox investigation. *Journal of Hydrology*, 554: 758-779. DOI:10.1016/j.jhydrol.2017.09.045

2. **Research presented in Chapter 4:**

   Working collaboratively, we applied HT to interpret long-term municipal wellfield records. This synthetic study demonstrates the feasibility of transient HT analysis of existing hydrographs for large-scale heterogeneity characterization.

Citation:

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. DOI:10.1016/j.jhydrol.2020.125438

3. **Research presented in Chapter 5:**

We extended the synthetic study to transient HT analysis of field records. As part of a collaborative effort, I contributed to the development of alternative data processing and analysis strategies to account for uncertain initial and boundary conditions for groundwater flow modeling, making it possible to apply HT for large-scale heterogeneity characterization using field records.

Citation:

Luo, N., Illman, W.A., Zha, Y., 2022. Large-scale three-dimensional hydraulic tomography analyses of long-term municipal wellfield operations. *Journal of Hydrology*: 127911. DOI:10.1016/j.jhydrol.2022.127911

4. **Research presented in Chapter 6:**

Collaborating with fellow researchers, I actively participated in analyzing field cross-hole flowmeter measurements and conducting inverse modeling using different types of data. Our collective efforts demonstrate the practical utility of integrating flowmeter measurements with head response data for improved characterization of heterogeneity details.

Citation:

Luo, N., Zhao, Z., Illman, W.A., Zha, Y., Mok, C.M.W., Yeh, T.-C.J., 2023. Three-dimensional steady-state hydraulic tomography analysis with integration of cross-hole flowmeter data at a highly heterogeneous site. *Water Resources Research*, 59(6): e2002WR034034. DOI:10.1029/2022WR034034

As lead author of these four chapters, I was responsible for conceptualizing study design, performing data analysis, carrying our model runs, and writing and submitting manuscripts. My coauthors provided guidance of the research and provided feedback on draft manuscripts.

Abstract

Hydraulic Tomography (HT) has been evaluated to be a robust approach for high-resolution characterization of subsurface heterogeneity. However, geostatistics-based HT may produce overly smooth distributions of hydraulic parameters such as hydraulic conductivity (K) and specific storage (S_s) when hydraulic head data used to constrain the inversion is limited. Furthermore, only a few HT studies have been performed for large-scale field problems due to the difficulty in conducting dedicated HT surveys at large-scale sites, as well as due to uncertainty regarding model conceptualization. This thesis documents four studies designed to investigate the performance of HT in characterizing the spatial distributions of K and S_s at various scales through the inclusion of different types of data for inverse modeling. Study 1 investigates the effect of prior geological information on K and S_s heterogeneity characterization through Transient Hydraulic Tomography (THT) analysis of laboratory sandbox data. Study 2 explores the feasibility of THT analysis of long-term municipal well records for large-scale heterogeneity characterization through synthetic experiments, while Study 3 extends the synthetic study to a field application utilizing data from a wellfield in Kitchener, Ontario, Canada by addressing uncertain initial and boundary conditions for inverse modeling. Finally, Study 4 evaluates the usefulness of field cross-hole flowmeter measurements in mapping spatial K distribution through Steady-State Hydraulic Tomography (SSHT) analysis at a highly heterogeneous field site located on the University of Waterloo. Results from these studies mainly reveal that: (1) the incorporation of prior geological information into geostatistical inverse models improves characterization significantly when pumping tests and drawdown measurements are sparse; however, attention must be paid when constructing geological models for reliable structure information, (2) existing municipal wellfield records could be utilized for large-scale heterogeneity characterization using the approach of HT when uncertainties regarding initial and boundary conditions are well addressed for inverse modeling, and (3) the integration of cross-hole flowmeter measurements with hydraulic head data improves characterization results in terms of revealing K

heterogeneity details and predicting independent hydraulic test data. Overall, the body of work presented in this thesis advocates the inclusion of additional datasets that carry non-redundant heterogeneity information for geostatistical inverse modeling and demonstrates the feasibility of utilizing alternative datasets in HT for subsurface heterogeneity characterization at large-scale sites.

Acknowledgements

I would like to express my sincere gratitude to all those who have supported me throughout my Ph.D. journey. First, I would like to thank my supervisor, Prof. Walter A. Illman, for his guidance, support, and encouragement throughout these research projects and for providing me the opportunity of pursuing this Ph.D. at the University of Waterloo. His expert knowledge, insightful feedback, and patient mentoring have been invaluable in shaping my research and helping me to achieve my goals.

I would also like to thank the members of my thesis committee, Prof. David L. Rudolph and Dr. Steven J. Berg, for their valuable feedback and contributions to this work. Their constructive criticism and thoughtful suggestions have helped me to refine and improve my research. Thanks also go out to my internal-external examiner Prof. Bryan Tolson and the external examiner Prof. Kent Novakowski for their time reviewing this work and providing valuable comments to improve the quality of this thesis.

I am deeply grateful to the participants who generously gave their time and energy to take part in completing all studies. Without their willingness to share their experiences and perspectives, this research would not have been possible. Special thanks to Dr. Zhanfeng Zhao for many useful suggestions on solving technical issues.

I would also like to acknowledge the support and encouragement of my family and friends, especially my wife Ka Man Chan, who have been a constant source of inspiration and motivation throughout this journey. Their unwavering belief in me has kept me going, even during the most challenging times.

Finally, I extend my heartfelt thanks to everyone who has contributed to this work in one way or another.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Groundwater Flow Modeling

Numerical modeling has been widely used in hydrogeology to understand subsurface behavior in terms of groundwater flow and solute transport. Mathematically, groundwater flow in a three-dimensional, saturated, porous medium can be described through the following equations (Batu, 1998):

$$\nabla \cdot [K(\mathbf{x})\nabla H(\mathbf{x})] + Q(\mathbf{x}_p) = 0 \qquad \text{(Eq. 1.1 } \textit{for steady-state simulation}\text{)}$$

$$\nabla \cdot [K(\mathbf{x})\nabla H(\mathbf{x})] + Q(\mathbf{x}_p) = S_s(\mathbf{x})\frac{\partial h}{\partial t} \qquad \text{(Eq. 1.2 } \textit{for transient simulation}\text{)}$$

subject to initial and boundary conditions:

$$H|_{t=0} = f(\mathbf{x}), \qquad H|_{\Gamma_1} = f_1(\mathbf{x}, t), \qquad [K(\mathbf{x})\nabla H(\mathbf{x})] \cdot \mathbf{n}|_{\Gamma_2} = q_0 \qquad \text{(Eq. 1.3)}$$

where, in Eqs. 1.1 and 1.2, ∇ is the gradient operator, $K(\mathbf{x})$ is hydraulic conductivity (L/T), $H(\mathbf{x})$ is hydraulic head (L), $Q(\mathbf{x}_p)$ is the volumetric flux per unit volume (1/T) at location $\mathbf{x}_p$, and $S_s(\mathbf{x})$ is specific storage (1/L). In Eq. 1.3, $f(\mathbf{x})$ is the initial hydraulic head (L) at location $\mathbf{x}$, $f_1(\mathbf{x}, t)$ is a constant head (L) at the Dirichlet boundary Γ_1 at location $\mathbf{x}$ and at time t, q_0 is the specific discharge (L/T) at the Neumann boundary Γ_2, and $\mathbf{n}$ is a unit vector normal to Γ_2. On the other hand, following the work of Burnett and Frind (1987), the simulation of solute transport in a porous medium can be expressed using mathematical equations as:

$$-\nabla \cdot (\mathbf{q}C - \theta_s \mathbf{D}\nabla C) \pm Q_c = \theta_s \frac{\partial C}{\partial t} \qquad \text{(Eq. 1.4a)}$$

$$D_{ij} = [\alpha_L - \alpha_T]\frac{v_i v_j}{|\mathbf{v}|} + \alpha_T |\mathbf{v}|\delta_{ij} + D_0 \delta_{ij} \qquad \text{(Eq. 1.4b)}$$

subject to initial and boundary conditions:

$$C|_{t=0} = C_0, \qquad C|_{\Gamma_1} = C_D, \qquad [-\mathbf{q}C + \theta_s \mathbf{D}\nabla C] \cdot \mathbf{n}|_{\Gamma_3} = \mathbf{q}C_0 \qquad \text{(Eq. 1.5)}$$

In Eq. 1.4a, $\mathbf{q} = -K(\mathbf{x})\nabla H(\mathbf{x})$ is the specific discharge (L/T), θ_s is the effective porosity (-), C is the solute concentration (M/L^3), Q_c is the rate (M/L^{3}T) at which solutes are injected (source) or extracted (sink), and $\mathbf{D}$ is the hydrodynamic dispersion coefficient (L^2/T) evaluated from velocity and dispersivities, as shown in Eq. 1.4b. α_L and α_T represent longitudinal and transverse dispersivities (L), respectively, v_i and v_j are velocities (L/T) at different directions (x, y, and z), $|\mathbf{v}|$ is the magnitude of the velocity, D_0 is the effective molecular diffusion coefficient (L^2/T), and δ_{ij} is the identity tensor. In Eq. 1.5, C_0 is the initial solute concentration in the system, C_D is the specified concentration at the Dirichlet boundary (Γ_1), and $\mathbf{q}C_0$ represents the mass flux (M/L^{2}T) at the Cauchy boundary (Γ_3).

In practice, groundwater modeling involves two concepts: model identification and parameter estimation, as described by (Carrera et al., 2005). Model identification (or conceptualization) applies to defining the conceptual groundwater model of specific features, such as the governing equation describing the relationship between parameters and model outputs, initial and boundary conditions, and heterogeneity patterns of hydraulic parameters. Site specific knowledge (e.g., local geology/hydrogeology) and objectives (e.g., groundwater

flow/solute transport) are normally required to yield a reliable conceptual groundwater model. On the other hand, parameter estimation refers to assigning hydraulic parameters to elements throughout the discretized domain for groundwater modeling. The performance of the constructed numerical model in simulating groundwater flow and solute transport has been found to rely heavily on the accuracy of estimated hydraulic parameters in representing the reality (e.g., Berg and Illman, 2011a; Illman et al., 2012).

Over the past eight decades, numerous efforts have been dedicated to estimate hydraulic conductivity (K) and specific storage (S_s), which are two important hydraulic parameters for simulating groundwater flow, and in particular, for predicting solute transport (Ni et al., 2009) in a confined aquifer. Traditionally, the estimation of K and S_s is performed through the analysis of aquifer tests using type curve solutions (e.g., Hantush and Jacob, 1955; Mathias and Butler, 2006; Mishra and Neuman, 2011; Moench, 1997; Neuman, 1974; Theis, 1935) depending on the characterized aquifer type (e.g., confined/unconfined, leaky, saturated/unsaturated) and boundary conditions (e.g., full/partial penetration of well screen). Due to the implemented homogeneous assumption of hydraulic parameters, the utilization of these analytical solutions ignores the inherent heterogeneous nature of porous/fractured media and yields averaged K and S_s estimates over the tested area. Although the use of analytical models may lead to good matches between observed drawdowns and type curves, the estimated hydraulic parameters may be scenario-dependent. For instance, Wu et al. (2005) demonstrated that the conventional analyses of aquifer tests yielded biased K estimates that evolved with time and depended on the location of monitoring wells, while the estimated S_s is mainly affected by the geology (i.e.,

heterogeneity) between pumping and monitoring wells.

To reveal heterogeneity of hydraulic parameters, laboratory core sample analyses and various field experiments have been developed to yield K and S_s estimates at small scales. For instance, laboratory grain size analysis, permeameter tests (e.g., Alexander et al., 2011; Sudicky, 1986) and Hydraulic Profiling Tool (HPT) (e.g., Borden et al., 2021; McCall et al., 2017) yield high-resolution estimates of small-scale K measurements at interval ranges from couple centimeters to tens of centimeters; flowmeter survey is commonly applied to reveal the vertical distribution of horizontal K along the screen at well vicinity (Molz et al., 1994); slug and single-hole tests can be used to obtain local K and S_s estimates at various well locations. These small-scale estimates can then be interpolated using geostatistical methods (i.e., kriging) to yield spatial K and S_s distributions throughout the model domain for simulating groundwater flow and solute transport (e.g., Salamon et al., 2007; Sudicky et al., 2010; Zhao et al., 2023).

Using such an interpolation approach, a sufficient number of small-scale estimates is needed to fully capture the heterogeneity patterns of hydraulic parameters (Rehfeldt et al., 1992). Furthermore, Kuhlman et al. (2008) demonstrated that the interpolated K and S_s fields strongly relied on the small-scale estimates, which, however, might be biased in representing realistic conditions due to the restricted assumptions implied in analytical solutions (e.g., Hyder et al., 1994 for slug test analysis) or empirical formulae (e.g., the power-law model developed by Zhao and Illman (2022a) for HPT data analysis) for these estimates. Furthermore, the employed geostatistical interpolation methods often ignore the information about the geologic

structure and the flow field at a site, yielding overall smooth K and S_s distributions.

To overcome these issues, numerical inverse modeling approaches were developed to derive K and S_s distributions for groundwater modeling based on field measurements (i.e., point estimates of hydraulic parameters, head pressure, flux, concentration, etc.) and corresponding governing equations. In the following section, the fundamental concepts of several inverse modeling approaches are provided.

1.2 Inverse Modeling Approaches

The earliest numerical inverse approach could be traced back to Stallman (1956), who derived transmissivity (T) values from flow nets based on head measurements (Bennett and Meyer, 1952). In this study, head data were assumed to be known throughout the domain, and the relative distribution of T was computed through solving an ensemble of first-order partial differential groundwater flow equations. This approach, termed "direct" by Neuman (1973), shows some significant drawbacks. First, the interpolation of head data from point measurements to model grid introduces smoothing and unknown artificial errors into the head distribution over the simulation domain. Second, the solution tends to be unstable, in which, small changes in head data would cause significant variations of T estimates.

Alternatively, most recently developed inverse modeling approaches follow the "indirect" method proposed by Neuman (1973), which estimates hydraulic parameters by minimizing an objective function of calibration errors (i.e., residuals between observed and simulated head data) and a plausibility term to regularize hydraulic parameter estimates. The objective function

can be written in a common form as:

$$F = F_h + \lambda F_p \qquad \text{(Eq. 1.6a)}$$

$$F_h = (\mathbf{h} - \mathbf{h}^*)^t \mathbf{C}_h^{-1} (\mathbf{h} - \mathbf{h}^*) \qquad \text{(Eq. 1.6b)}$$

$$F_p = (\mathbf{p} - \mathbf{p}^*)^t \mathbf{C}_p^{-1} (\mathbf{p} - \mathbf{p}^*) \qquad \text{(Eq. 1.6c)}$$

where F_h represents the squared errors of observation data, and λF_p represents the regularization term for parameter estimates. $\mathbf{h}$ and $\mathbf{h}^*$ represent observed and simulated data vectors, respectively, $\mathbf{p} - \mathbf{p}^*$ shows measurement errors of model parameters, $\mathbf{C}_h$ and $\mathbf{C}_p$ are the covariance matrices of measurement errors of data and model parameters, respectively, and λ is the weighting factor of model parameters with respect to the objective function. The minimization of the objective function of Eq. 1.6a then refers to finding a good match between observed and simulated data, while maintaining the stability of model parameters. Since the relation between hydraulic parameters and field observations is usually non-linear, the non-linear model is often linearized using an approximation method (e.g., Taylor series) and the objective function is minimized iteratively for parameter estimation.

The Levenberg-Marquardt (L-M) method (Marquardt, 1963) is an minimization algorithm developed to solve non-linear problems by iteratively minimizing the sum of the squared differences between the observed and simulated data (least-squares) for parameter estimates. In comparison to traditional non-linear least-squares problems, it modifies the objective function by adding a regularization term that ensures numerical stability of the inverse problem

and avoids overfitting of data points. Furthermore, the L-M method combines the features of the Taylor series method (Jacobian matrix) and the gradient-descent algorithm to achieve fast convergence and stability. Although the L-M method shows some advantages, the objective function of its traditional form only accounts for the F_h term in Eq. 1.6. This may result in physically unreasonable hydraulic parameter estimates, especially when the inverse model is parameterized by a large number of unknowns.

The L-M method is employed in the model-independent parameter estimation code PEST (Doherty, 2015) for the determination of optimal hydraulic parameter set. Instead of solely minimizing the discrepancy of observation data (F_h), PEST also minimizes a regularization criterion (F_p) that assures non-homogeneity of hydraulic parameters, yielding a global objective function similar as Eq. 1.6. The term F_p can be viewed as a regularization criterion in the sense of Tikhonov (1963), which introduces "preferred parameter relationship" or "preferred parameter state" into the calibration process. Furthermore, the inverse of covariance matrices $\mathbf{C}_h^{-1}$ and $\mathbf{C}_p^{-1}$ are replaced by two square, diagonal matrices incorporating weights assigned to observations and regularizations, respectively. However, Ackerer et al. (2023) argued that such a regularization did not include prior information of hydraulic parameter estimates for model calibration, limiting the number of hydraulic parameters to be estimated.

Carrera and Neuman (1986a; 1986b) proposed maximum likelihood estimator (MLE) to estimate groundwater model parameters by minimizing a "log-likelihood" criterion that includes not only field observations, but also prior information of model parameters. Other than

K and S_s, model parameters, such as interior recharge/leakage rate, boundary head, flow rate, and interior and boundary head-dependent recharge/leakage coefficients, are also accounted for in the inverse model. Prior information of these parameters is incorporated by adding a plausibility criterion to the objective function (similar to Eq. 1.6a), and model parameter estimates are obtained by minimizing the objective function in a sense of the least-squares method. Carrera and Neuman (1986a; 1986b) demonstrated that the inclusion of such a plausibility criterion would mitigate the risk of instability of inverse problems. The performance of the proposed MLE method was then evaluated through synthetic and field studies (Carrera and Neuman, 1986c). Results showed that the inclusion of prior information on model parameters to be estimated led to improved sum of square errors of observation data.

It should be noted that approaches described above (the L-M method and its implementation in PEST, and the MLE method) treat unknown parameters as deterministic quantities. When using these approaches for site heterogeneity characterization, the model domain should be subdivided into a set of zones, and uniform hydraulic parameter values are estimated for each zone. Although geological mapping and other field experiments (i.e., geophysical survey) can help in determining zones, the shape of identified zones (i.e., layer boundaries of geological units) may still involve great uncertainty. The effect of uncertain geological layer boundaries on K and S_s estimates was evaluated by Tong et al. (2021), who characterized a municipal water-supply wellfield using zonation models based on geological investigations from different sources. Their results showed that all zonation models were well calibrated to fit well to the observation data; however, estimated K and S_s values for a specific

geological unit varied from one zonation model to another. Furthermore, the use of zones of uniform hydraulic parameters ignores the intra-layer heterogeneity of geological units, which, however, are of critical importance in accurately predicting groundwater flow and solute transport. Relevant results are obtained in this research and presented in Chapter 4 of this thesis. To circumvent these shortcomings of a zonation model, geostatistics-based inverse modeling approaches are developed to estimate spatial distributions of hydraulic parameters for groundwater modeling.

de Marsily et al. (1984) proposed the pilot points method (PPM) for subsurface heterogeneity characterization. Basically, the PPM involves (1) estimating hydraulic parameters at pilot points, and (2) spatially interpolating parameter estimates at pilot point locations to a model grid using a geostatistical method (i.e., kriging/cokriging). The original form of the PPM (de Marsily et al., 1984), however, shows some drawbacks. For instance, pilot points were selected manually by trial and error, while the location and number of pilot points were later evaluated to show significant impacts on parameter estimation results (Hendricks-Franssen, 2001; LaVenue and Pickens, 1992). Second, pilot point estimates were not regularized, which might lead to instability of the inverse problem, especially when the model was parameterized by a great number of pilot points.

In later studies, the original form of the PPM has undergone several modifications to relieve the impact of arbitrary pilot point selection on characterization results and to overcome the issue of unstable solutions to inverse problems. For instance, RamaRao et al. (1995)

included conditional simulation to generate an initial parameter field for model prediction and sequentially increased the number of pilot points by adding new ones at locations where they showed the highest sensitivity to the objective function. Doherty (2003) introduced a regularization criterion of model parameters to the objective function, making the revised PPM possible to mitigate the risk of instability when estimating hydraulic parameters at a great number of pilot points. Furthermore, Alcolea et al. (2006) incorporated prior information about model parameters in the optimization process and formulated the inverse problem in a maximum likelihood framework.

Despite that the PPM has been widely used to characterize spatial variations of hydraulic parameters in both porous and fractured media (e.g., Certes and de Marsily, 1991; Lavenue and de Marsily, 2001; Pool et al., 2015), one critical question remains that whether the geostatistical model used to interpolate hydraulic parameter field provides a good representation of heterogeneity patterns. Huang et al. (2011) demonstrated that the interpolation based on a given spatial statistic covariance function (or variogram) ignored the flow field that contains heterogeneity information of hydraulic parameters at the site. Furthermore, Pool et al. (2015) compared the PPM to the geology-based zonation model for characterizing a heterogeneous aquifer to predict groundwater flow and solute transport. Their results revealed that, while the *T* field estimated through PPM led to a better match of head data, the calibrated geology-based zonation model yielded better prediction of seawater intrusion at the site. Such a study implies that instead of relying solely on a traditional geostatistical interpolation method (i.e., kriging/cokriging), prior structural information should also be incorporated for site

heterogeneity characterization using the PPM.

Other than the PPM, geostatistical inverse modeling approaches have been developed to estimate spatial distribution of hydraulic parameters for groundwater modeling. For instance, Kitanidis (1995) proposed the Quasi-Linear Geostatistical Approach (QLGA), in which, hydraulic parameter fields are treated as random fields and characterized by their means and covariances. Briefly, the QLGA estimates spatial distribution of hydraulic parameters by minimizing the discrepancy between the simulated and measured observation data, while constraining the covariances of the estimated parameter fields to the expected ones. In the work of Kitanidis (1995), the statistical parameters describing the expected covariance of hydraulic parameters were derived on the basis of the Maximum Likelihood and Gauss-Newton methods.

Yeh et al. (1996) proposed a geostatistical approach, named Successive Linear Estimator (SLE), to solve non-linear inverse problems for spatial T distribution estimation. Similar to the classical cokriging technique, the SLE characterizes the spatial distribution of T on the basis of the covariances of T and pressure heads (H) and their cross-covariance. The SLE approach treats the natural logarithm of T ($\ln T$) of an aquifer as a multi-Gaussian, second-order stationary, stochastic process. Given unconditional means, variance, and correlation scales of T, SLE starts with cokriging of initial T estimates and head data from aquifer test to create an initial estimate of $\ln T$ field. The initially estimated T field is then used to solve the governing equation of steady-state groundwater flow to obtain simulated head data. To account for the non-linear relation between T and H, $\ln T$ field is updated iteratively based on the discrepancy between

observed and simulated head data as well as the revised covariance and cross-covariance matrices of the estimated T and simulated H in each iteration. The fundamental concepts and main mathematical algorithms of the proposed SLE are summarized as Appendix A.

With the implementation of SLE, Yeh and Liu (2000) developed a Sequential Successive Linear Estimator (SSLE) for Steady-State Hydraulic Tomography (SSHT). In their study, steady-state head data from a set of pumping tests conducted during a HT survey were incorporated and interpreted sequentially for K heterogeneity characterization. Specifically, SSLE starts the iterative process with available K estimates and the head response data from one test. Once the inverse model converges, the estimated K field and the revised parameter covariances, conditioned on head data from the first test, are used as prior information for the next estimation based on a new set of head data from the next test. Then, Zhu and Yeh (2005) extended the formulated SSLE to Transient Hydraulic Tomography (THT), in which, K and S_s are jointly estimated using transient head responses collected during a series of aquifer tests.

Such a sequential approach allows the inverse model to interpret massive head data from a HT survey, while maintaining the covariance matrix at a manageable size for computational efficiency. However, Illman et al. (2008) illustrated that a different sequence of aquifer tests incorporated into an inverse model would lead to varying estimates of K field. They concluded that the cleanest data (with the highest signal-to-noise ratio) should be included into the SSLE algorithm first, followed progressively by noisier data.

To overcome this issue, Xiang et al. (2009) proposed the Simultaneous Successive Linear

Estimator (SimSLE), in which, head response data from all aquifer tests during a HT survey are interpreted simultaneously for K and S_s heterogeneity characterization. In this thesis, geostatistical inversions of all presented studies were carried out using the SimSLE, while code was modified by Prof. Yuanyuan Zha to account for temporally variable pumping/injection flow rates.

1.3 Review of Hydraulic Tomography Studies

Over the past two decades, hydraulic tomography (HT) has been proposed (Gottlieb and Dietrich, 1995; Yeh and Liu, 2000) and utilized as an advanced site characterization approach for revealing spatial distribution of hydraulic parameters. In a dedicated HT survey, hydraulic response data at various locations are collected during a series of spatially varying pumping/injection tests. Analogous to the concept of tomographic surveys in geophysics (Neuman, 1987), such head response data provide heterogeneity snapshots regarding hydraulic parameters at different locations between each pumping and observation ports (Yeh and Lee, 2007). The collected data during HT surveys can subsequently be interpreted using different inverse algorithms (e.g., Bohling et al., 2002; Brauchler et al., 2003; Kitanidis, 1995; Kitanidis and Lee, 2014; Xiang et al., 2009; Yeh et al., 1996; Zha et al., 2018; Zhu and Yeh, 2005; Zhu and Yeh, 2006) to map the heterogeneity of hydraulic parameters. Yeh et al. (2008) concluded that the data collected during such tomographic surveys provide many constraints for model calibration, yielding more accurate estimates of K and S_s fields with less uncertainty in comparison to traditional characterization methods.

The robust performance of HT for both porous and fractured media characterization has been evaluated through numerous numerical (Bohling et al., 2002; Yeh and Liu, 2000; Zhu and Yeh, 2005), laboratory (Berg and Illman, 2011a, 2012; Illman et al., 2008, 2010, 2015; Illman et al., 2007; Liu et al., 2002, 2007; Sharmeen et al., 2012; Zhao et al., 2015, 2016), and field (Berg and Illman, 2011b, 2013, 2015; Bohling et al., 2007; Brauchler et al., 2011; Cardiff et al., 2009, 2012, 2013; Castagna et al., 2011; Illman et al., 2009; Paradis et al., 2016; Straface et al., 2007; Zha et al., 2015, 2016; Zhao and Illman, 2017) studies.

Other than classical steady-state (e.g., Yeh and Liu, 2000) and transient (e.g., Zhu and Yeh, 2005) HT, information on water-level changes derived from drawdown curves has also been proposed to be utilized for HT analysis. For instance, to overcome the impact of uncertain boundary conditions on inverse modeling, Bohling et al. (2002) proposed a steady shape analysis of transient HT data for K heterogeneity characterization. Later, Bohling et al. (2007) validated the steady shape HT through a field study conducted in an alluvial aquifer located in Kansas, USA, and they concluded that steady shape analysis of transient HT data yielded similar performance in estimating K profiles when compared to transient HT. On the other hand, Brauchler et al. (2003) proposed a travel-time based HT approach to obtain diffusivity ($D = K/S_s$) distribution between tested wells. Hu et al. (2011) then combined the travel-time and steady shape HT analyses to estimate spatial variations of K and S_s. Furthermore, Zhu and Yeh (2006) proposed the HT analysis of temporal moments of drawdown recovery data for mapping heterogeneity of K and S_s, which was later validated through a laboratory sandbox study by Yin and Illman (2009).

A set of comparison studies have shown that geostatistical analysis of HT data is superior to traditional geostatistical interpolation methods (i.e., kriging) in accurately mapping K and S_s heterogeneities in greater details. Through a controlled sandbox study, Illman et al. (2010) illustrated that the heterogeneous K field obtained through SSHT showed a better correspondence to the synthetic aquifer and performed significantly improved predictions of independent pumping tests in comparison to that obtained through kriging of small-scale K estimates (core sample analyses and single-hole tests). Using the same sandbox, Berg and Illman (2011a) extended the comparison study of Illman et al. (2010) to THT for K and S_s heterogeneity characterization, and Illman et al. (2012) compared the performances of K fields derived from kriging of permeameter K estimates and computed through SSHT on predicting solute transport. Besides, Berg and Illman (2015) characterized a highly heterogeneous field site of glaciofluvial deposits using seven methods. Comparison results revealed that THT performed considerably better than other "traditional" methods (kriging, effective parameter model, transition probability/Markov Chain geostatistics model, geological model, and stochastic inverse model) of dealing with K and S_s heterogeneities and predicting independent pumping tests.

Despite the robust performance of HT on characterizing subsurface heterogeneity, the estimated spatial distribution (or tomogram) of hydraulic parameters through geostatistical inversion of HT data may still suffer from the issue of smoothness, especially at locations where observation data are lacking for inverse modeling (Cardiff et al., 2013; Illman et al., 2015). Ahmed et al. (2015) attributed the smoothness issue to the inherent estimation of conditional

means implied in most geostatistical inversion approaches (e.g., Kitanidis, 1995; Yeh et al., 1996). The estimated smooth K and S_s tomograms might be adequate in predicting drawdown distributions from independent pumping tests not used in the calibration effort. However, the recovery of a finer scale resolution tomogram including layer boundaries is likely needed for improved predictions of solute and contaminant transport.

To overcome the issue of smoothness, structural information was suggested to be incorporated for inverse modeling by many studies. For instance, Jiménez et al. (2013) applied the travel-time inversion of hydraulic head response data to obtain the information of domain structural features, which in turn was used to guide the pilot point-based HT inversion for mapping spatial K distribution. Using the proposed sequential procedure, Jiménez et al. (2015) characterized high-resolution K and S_S tomograms at a field site in Stegemühle, Germany, to predict solute transport. On the other hand, through a sandbox study, Zhao et al. (2016) examined the value of integrating geological information on a HT survey by constructing geology-based initial K field for SSHT analyses. They found that utilizing an accurate geological model as a prior estimate was beneficial in improving the K tomogram in terms of revealing inter/intra-layer K heterogeneity. Later, Zhao and Illman (2017) incorporated geology-based structural information for SSHT analysis to characterize the spatial K distribution at the North Campus Research Site (NCRS) located on the University of Waterloo campus in Waterloo, Canada. In a different study, prior structural information for HT analysis was incorporated by using mean and covariance of hydraulic parameters conditioned on site-specific geological features for geostatistical inverse modeling (Zha et al., 2017).

Alternatively, additional datasets that carry non-redundant heterogeneity information with respect to head measurements were suggested to be integrated for HT analysis to refine the delineation of heterogeneity details of hydraulic parameters. Li et al. (2008) jointly inverted flowmeter measurements and hydraulic head data to address the issue of depth-averaged head data for *K* heterogeneity characterization. Zha et al. (2014) integrated flux data with head pressures to map the *K* tomogram in a synthetic two-dimensional (2-D) fractured rock. Tso et al. (2016) extended the work of Zha et al. (2014) to characterize spatial *K* distribution in a synthetic three-dimensional (3-D) porous medium. Most recently, Zhao and Illman (2022a, 2022b) advocated that in-situ HPT profiles and corresponding kriged *K* distributions (Zhao et al., 2023) could be integrated with head data for highly heterogeneous unconsolidated materials characterization using the approach of HT.

Another practical issue of HT that deserves significant attention is whether the approach be applied to characterize hydraulic parameter heterogeneity at large-scale sites? To date, most of HT studies were performed at small-scale (limited to tens of square meters) sites, while only a few studies have been carried out for large-scale (several square kilometers) site characterization through HT analysis of head response data from dedicated pumping/injection tests (e.g., Illman et al., 2009; Kuhlman et al., 2008; Zha et al., 2015, 2016). This may be attributed to the reasons that (1) designing and conducting dedicated HT surveys for obtaining sufficient hydraulic response data are typically difficult, and sometimes impractical, at large-scale sites, (2) the constructed large-scale groundwater model for inverse modeling may involve great uncertainties regarding model conceptualization (i.e., initial and boundary

condition, unknown natural/artificial sources, etc.), and (3) it is always computationally intensive when a highly parameterized geostatistical model is adopted in order to characterize heterogeneity of hydraulic parameters in great details.

Instead of the traditional data collection strategy used for HT survey, alternative datasets have been proposed and utilized for large-scale heterogeneity characterization using HT (e.g., Mao et al., 2018; Wang et al., 2017; Yeh et al., 2009; Zha et al., 2019). Specifically, Yeh et al. (2008) provided an opinion of using natural stimuli (i.e., river-state variations, lightning, earthquake, barometric variations, storm events, etc.) as sources of excitations for basin-scale subsurface characterization. Unlike traditional single-well or advanced multiple-well (Kuhlman et al., 2008) pumping tests, natural stimuli can easily stress the aquifer to yield groundwater responses over the entire basin. They pointed out that groundwater variations at different scales induced by natural stimuli with frequent and spatially varying occurrence is analogous to that of HT surveys, and the monitored groundwater responses along with the characterized natural stimuli can be interpreted for hydraulic parameter estimation.

Following this thought, Yeh et al. (2009) proposed river stage tomography as a new approach for basin-scale subsurface heterogeneity characterization, which was later validated through a field experiment conducted in Zhoushui River alluvial fan, Taiwan (Wang et al., 2017). Mao et al. (2018) jointly interpreted datasets obtained from water inrush incidents and one traditional pumping test to characterize large-scale heterogeneities in K and S_s at a deep coal mining site located under the North China Plain, China. Yeh et al. (2008) contended that

natural stimuli-based HT surveys should be a future direction for large-scale subsurface characterization; however, significant challenges still exist in accurately characterizing the locations and strengths of natural stimuli.

To avoid the uncertainty associated with natural stimuli, existing hydraulic head records in a wellfield with well-characterized artificial stimuli (pumping/injection operations with known locations and rates) can be utilized as alternative datasets for subsurface heterogeneity characterization, as suggested by Yeh and Lee (2007). Such records are typically abundant and can be readily obtained from contaminant transport monitoring or municipal water-supply wellfields, but they are rarely adopted for mapping the heterogeneity of hydraulic parameters. Most recently, Zha et al. (2019) exploited the pump-and-treat system for subsurface heterogeneity characterization at the AFP44 site located in Tucson, Arizona, US.

Although numerous studies have shown that HT can be used to yield promising subsurface heterogeneity characterization results, there is still ample room for further improvement of the HT approach in aspects of applying it to characterize heterogeneity at large-scale sites and integrating additional datasets to yield high-resolution K and S_s tomograms in greater details. Thus, this thesis focuses on answering the following questions:

1. How useful will prior geological information be when spatial K and S_s distributions are jointly estimated through THT?
2. Can existing long-term municipal well records be subjected to THT analysis for characterizing K and S_s heterogeneities at a large-scale site?
3. Can cross-hole flowmeter measurements be integrated with hydraulic head data to improve the characterization of spatial K distribution?

2 Research Topics and Objectives

This thesis contains three research topics with four completed studies. The overall objective of this thesis is to investigate the performance of HT in characterizing K and S_s heterogeneities at varying scales through the inclusion of different types of data for inverse modeling.

The first topic (Study I) evaluates the effect of prior geological information on K and S_s heterogeneity characterization through THT analyses of laboratory sandbox data. The second topic examines the feasibility of THT analysis of long-term municipal wellfield records for large-scale K and S_s heterogeneity characterization. Two studies are completed for this topic: one synthetic case (Study II) and one field application (Study III). The last topic (Study IV) investigates the usefulness of field cross-hole flowmeter measurements for mapping spatial K distribution at a highly heterogeneous glaciofluvial deposit site.

In this chapter, specific objectives of each study are described. Results from four studies are presented and discussed individually in Chapters 3, 4, 5, and 6. Chapter 7 draws summary and conclusions from each study, while Chapter 8 provides some recommendations for future HT studies based on the four completed studies in this thesis.

Study I: *Geological Data for THT Analysis: Laboratory Sandbox Study*

To answer the question of what level of model complexity is required for HT analysis, Illman et al. (2015) compared HT models of varying parameterization through SSHT analyses of laboratory sandbox data of Illman et al. (2010), including: (1) isotropic and anisotropic effective parameter models, (2) a geology-based zonation model with uniform K in each layer, and (3) a geostatistical model with a spatially variable K field. All models were calibrated using two datasets of different pumping and monitoring densities. Their results showed that the geology-based zonation model with perfect knowledge of stratigraphy performed nearly as well as the highly parameterized geostatistical model, especially when the number of pumping test data utilized for model calibration was reduced. Furthermore, their comparison results revealed that details of K heterogeneity were lost when the number of hydraulic data was reduced for geostatistical inverse modeling. In particular, the estimated spatial K distributions through geostatistical inversions of head data were able to recover the major layers of high and low K values, but distinct layer boundaries were not recovered.

In a different study, Zhao et al. (2016) investigated the impact of integrating geological information of varying accuracy on SSHT using the laboratory sandbox data of Illman et al. (2010). They found that utilizing an accurate geological model as a prior estimate for geostatistical inversion was beneficial in improving the K distribution or tomogram, especially when hydraulic head data were sparsely distributed.

These two studies were performed through SSHT analyses of sandbox data; however,

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

Study II: *THT Analysis of Long-term Municipal Well Records: Synthetic Study*

Zha et al. (2019) exploited the pump-and-treat system for large-scale heterogeneity of hydraulic parameters characterization at the AFP44 site located in Tucson, Arizona, US. In their study, hydraulic head changes during four distinct events (i.e., system shutdown and resumption, changes in pumping/injection operations, and significant variations of flow rates) were extracted from the existing head records and utilized traditional aquifer test data for inverse modeling. However, it should be noted that the characteristics of well hydrographs might be quite different from wellfield to another, depending on the associated pumping/injection regime. Without presenting apparent changes of hydraulic head with distinct events, the well hydrographs in a municipal wellfield appear to be highly variable due to the continuous operation of water-supply wells with variable flow rates. The extracted hydraulic

head records within selected periods are affected by prior pumping/injection operations at the same site, resulting in significant difficulties in interpretation of head records due to unknown initial conditions for groundwater modeling. On the other hand, Sun et al. (2013) proposed a temporal sampling strategy for HT analysis using dedicated pumping test data; however, questions remain as to which data points should be extracted from the long-term head records and utilized for subsurface heterogeneity characterization.

In this study, a series of numerical experiments that mimic the hydraulic conditions at the Mannheim East site, a municipal water-supply wellfield located in the southwest area of the City of Kitchener, Ontario, Canada, was performed. Specifically, a synthetic 3-D multi-aquifer/aquitard system was developed and characterized using different modeling approaches with different head records for groundwater flow and solute transport predictions. Fundamentally, such a synthetic study with minimized sources of errors (i.e., model identification and head measurement) yields a general framework for subsurface heterogeneity characterization using existing long-term pumping/injection and water-level records.

The main objectives of this synthetic study are to: (1) explore the feasibility of utilizing municipal well data for subsurface heterogeneity characterization using THT; (2) evaluate the performance of three different modeling approaches (homogeneous, geology-based zonation, and geostatistical models) for THT analyses with records from a municipal wellfield; and (3) investigate the effect of data selection for inverse modeling. The computed K and S_s tomograms from the different models were validated through forward numerical simulations of additional

pumping tests not used during model calibration and nonreactive tracer migration through the municipal wellfield.

Study III: *THT Analysis of Long-term Municipal Well Records: Field Application*

This study extended Study II to a field investigation. Through a synthetic case study, we showed that long-term municipal well records could be subjected to THT analysis for site characterization. The comparison results revealed that the highly parameterized geostatistical model with prior geological information was the best approach to interpret the exiting hydraulic data at municipal well sites for large-scale heterogeneity characterization, yielding K and S_s tomograms that could provide much improved predictions of not only groundwater level variations, but also solute transport compared to effective parameter and geology-based zonation models.

However, questions remain on whether THT analysis of field records using a geostatistics-based method would yield accurate delineation of large-scale heterogeneity patterns of hydraulic parameters due to the facts that: (1) field observation data always involve measurement errors, (2) estimated/identified initial and boundary conditions for groundwater modeling may involve great uncertainties, especially for large-scale inverse problem with long-term observation data, and (3) it is difficult to exclude water-level variations induced by unknown/unidentifiable natural/artificial stimuli.

In the previous work, the effect of uncertain initial condition on model calibration was accounted for and minimized by incorporating a certain period of pumping/injection information prior to the observation data for model spin-up, as suggested by Yu et al. (2019). However, the effect of uncertain boundary conditions and the uncertainty from other aspects on inverse modeling were not addressed.

In this study, the subsurface heterogeneity of K and S_s at the Mannheim water-supply wellfield was characterized through THT analyses of long-term field records. Specifically, a geological model was firstly constructed based on the available borehole logs at the study site and initially calibrated with uniform K and S_s values in each geological unit. The calibrated geology-based zonation model was then utilized as initial guesses of K and S_s fields for highly parameterized geostatistical model calibration. Alternative data processing and analysis strategies were proposed to minimize the effect of uncertain initial and boundary conditions on inverse modeling for both models. The estimated K and S_s tomograms from both the geology-based zonation and geostatistical models were validated through the prediction of municipal water-level records and dedicated pumping test data that have not been used for calibration efforts. Furthermore, the calibration and validation results were assessed to evaluate the uncertainties associated with model conceptualization, which may help in the future to improve the groundwater model. Overall, this study can be used as a guide to characterize large-scale heterogeneities in hydraulic parameters through HT analysis of long-term field water-level records from municipal wellfield operations.

Study IV: *Integration of Cross-hole Flowmeter Measurements with Hydraulic Head Data for SSHT Analysis: Field Study at NCRS*

In this study, borehole flowmeter measurements were proposed as a new type of data and integrated into HT analysis to improve heterogeneity characterization results at the North Campus Research Site (NCRS), which has been classified as highly heterogeneous glaciofluvial deposits through previous site investigations (Alexander et al., 2011). Flowmeter surveys have been widely used in porous media to reveal vertical distribution of horizontal K at well locations (e.g., Molz et al., 1989), and in fracture media to identify flow zones in single wells as well as their spatial connections (e.g., Paillet, 2000; Williams and Paillet, 2002).

Despite the technique of flowmeter has been developed to capture a wider range of flow rates and have a higher precision of the measured flow (from impeller and heat-pulse to electromagnetic), the accuracy of measured flow log in representing true heterogeneity condition as well as its interpretation for K estimates could strongly be influenced by borehole geometry. For instance, Bomana et al. (1997) and Dinwiddie et al. (1999) reported the influence of sand/gravel-pack around well screen, which would cause a portion of groundwater to bypass the meter through vertical flow in the surrounding high permeable materials. To reduce the magnitude of bypass flow, pumping/injection at a low flow rate (i.e., less than 10 L/min) is suggested when performing flowmeter survey at sand/gravel-packed boreholes (Dinwiddie et al., 1999). On the other hand, Paradis et al. (2011) proposed the utilization of a direct-push well without any high permeable packs around well screen for flowmeter surveys.

At the NCRS, flowmeter tests were performed mainly in multi-screen wells with sand-pack around each screen interval. The collected vertical flow logs in such boreholes had not been subjected to analysis in any previous flowmeter related studies, which rendered the usefulness of such data for accurate *K* heterogeneity characterization still in doubt. As a result, the first objective of this study was to examine the flowmeter measurements collected at the NCRS through preliminary analysis of vertical flow logs and comparing them to the investigated lithology at borehole locations.

Several conventional methods have been developed to interpret flowmeter measurements in porous media to estimate *K* variations as a function of vertical position at well vicinity (Kabala, 1994; Molz et al., 1989; Molz and Young, 1993; Rehfeldt et al., 1989). However, it should be noted that all proposed conventional interpretation methods assume that the aquifer is composed of a series of stratified and horizontal layers in the vicinity of the well, and groundwater flows into (or away from) borehole horizontally. Such implemented assumptions limit the estimated *K* values to properly characterize *K* heterogeneity in 3-D space (Molz et al., 1989). Furthermore, any violation of the horizontal flow assumption may result in biased estimation of horizontal *K* distribution in the vertical direction (Bomana et al., 1997; Ruud et al., 1999). As a result, more sophisticated interpretation methods are still required to analyze flowmeter measurements for reliable *K* estimates.

Fienen et al. (2004) analyzed flowmeter data using Bayesian inverse method. Briefly, with prior knowledge of depth-averaged *K* value at well location, they formulated the measured

vertical flow at certain depth as a linear function of the integration of unknown *K* values for unit depth from the bottom to that depth. The formulated governing equation was then adopted to estimate *K* distribution along well screen using the Bayesian inverse method. In addition to *K* estimates, the utilization of geostatistical inverse method also provided uncertainty analysis regarding estimated *K* values. Later, Li et al. (2008) jointly inverted flowmeter and hydraulic head data for *K* heterogeneity characterization. In their study, the information of relative *K* distribution at each well location along the screen interval was obtained from flowmeter measurements, which was then incorporated as prior information for geostatistical inversion of steady-state drawdown data from pumping tests to map 3-D spatial *K* distribution. Li et al. (2008) demonstrated that the incorporation of flowmeter data resolved the issue of depth-averaged hydraulic properties estimated using head response data collected from fully screen wells, yielding improved vertical resolution of *K* heterogeneity at well locations with reduced uncertainty.

These studies showed that flowmeter data could be analyzed using sophisticated inverse methods (e.g., geostatistical model) for *K* estimates; however, to our best knowledge, no study has been performed to characterize subsurface heterogeneity through directly inverting in-situ flowmeter measurements. As mentioned previously, pumping-induced flux data in porous media had been evaluated to carry non-redundant information about *K* heterogeneity in comparison to head response data (Tso et al., 2016), while the measured vertical flow rates during flowmeter survey were analogous to the concept of flux data collected during a hydraulic test. Thus, the second objective of this study is to simulate in-situ flowmeter

measurements using numerical model, and then jointly inverting flowmeter and steady-state head data using a geostatistical inversion method for subsurface heterogeneity characterization. To further investigate the value of flowmeter measurements, SSHT analyses of head data with prior geological information were also performed for comparison purposes.

For this study, we first described the study site in terms of local geology, well construction, as well as the collected flowmeter and steady-state head data at the site. Preliminary analysis of flowmeter data was then provided to examine the quality of measured flow logs and to select proper data points for inverse modeling. Subsequently, geostatistical inversions of different combinations of three data types (geological information, flowmeter measurements, and steady-state head data) were performed to characterize spatial K distribution at the site. Characterization results from all cases were illustrated in terms of model calibration and validation, followed by discussion and conclusion sections. Primarily, this study was performed to investigate the feasibility of integrating flowmeter data for highly heterogeneous site characterization using the approach of HT.

3 Study I: Comparative Study of Transient Hydraulic Tomography with Varying Parameterizations and Zonations: Laboratory Sandbox Investigation

3.1 Experimental Setup

3.1.1 Sandbox Description and Collected Data

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

To create a realistic heterogeneity pattern, Illman et al. (2010) built this synthetic aquifer through the cyclic deposition of sediments under varying water flow and sediment feed rates, mimicking the interfingering of natural fluvial deposit. Specifically, a uniform flow rate and a specific sand type were chosen for a given layer deposition, producing small-scale heterogeneities within each layer and larger-scale heterogeneities of different layers. Through a sediment transport process, 18 layers of varying size sands were deposited, as shown in Fig. 3.1. Such a synthetic heterogeneous aquifer with exactly known stratigraphy is necessary to investigate the effect of geological information on groundwater flow modeling. Upon

completing the deposition of the layers, 48 ports were constructed throughout the aquifer along six columns with eight ports each (see Fig. 3.1). Each port has a diameter of 1.3 cm and fully penetrates the aquifer width. These ports can be utilized for pumping and injection of water, as well as monitoring head levels with a pressure transducer instrumented at each port.

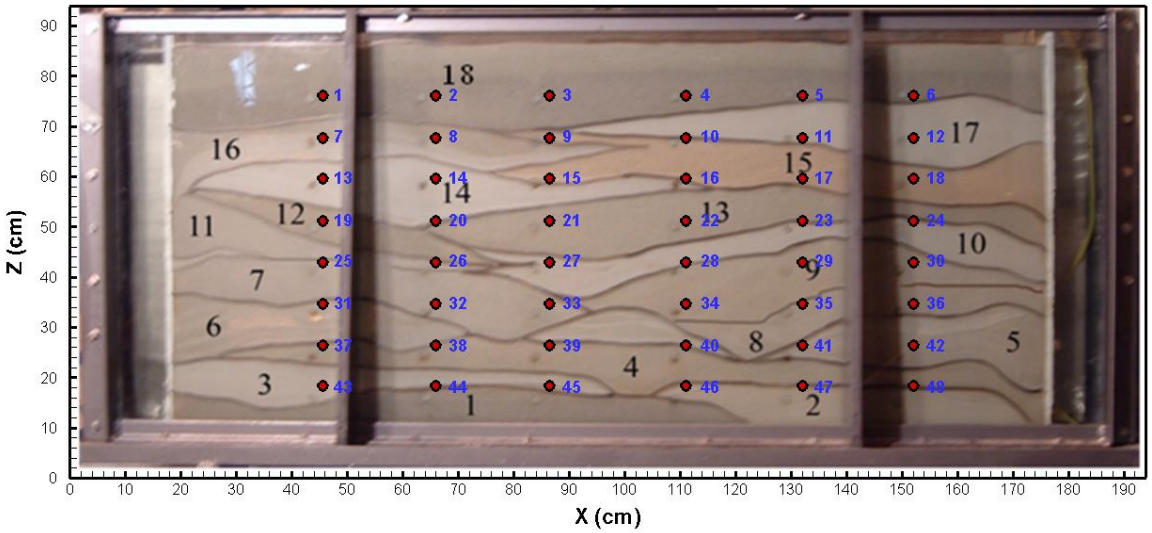

Fig. 3.1: Photograph of synthetic heterogeneous aquifer showing the layer (black) and port (blue) numbers, modified after Illman et al. (2010). Red circles indicate the 48 ports installed in the aquifer.

The synthetic aquifer was then characterized with different techniques. Core samples extracted from these ports were subjected to grain size analyses and permeameter tests to estimate local K values. In addition, single-hole pumping tests were performed at each port to obtain small-scale estimates of K and S_s. The d_{50}, which is the particle diameter for which 50% of the weight is finer, as well as K and S_s estimates of the different layers are summarized in Table 3.1. When multiple ports are available for a given layer, the geometric means of their estimates are provided.

Table 3.1: Sand type, d_{50}, K and S_s estimates for each deposited layer in the synthetic heterogeneous aquifer[a].

Layer	Sand	d_{50} (mm)	K (cm/s) Shepherd (1989)	Core Permeameter K^b (cm/s)	Single-Hole K^b (cm/s)	Single-Hole Ss^b (/s)
1	20/30	0.75	1.03×10^{-1}	3.20×10^{-2}	5.32×10^{-2}	2.12×10^{-4}
2	4030	0.35	2.99×10^{-2}	5.29×10^{-2}	5.67×10^{-2}	2.60×10^{-4}
3	F-85	0.15	7.28×10^{-3}	7.14×10^{-2}	5.70×10^{-2}	5.00×10^{-4}
4	20/40	0.58	6.68×10^{-2}	5.68×10^{-2}	5.10×10^{-2}	2.22×10^{-4}
5	mix	0.46	N/A	N/A	N/A	N/A
6	mix	0.46	N/A	8.16×10^{-2}	5.00×10^{-2}	4.00×10^{-4}
7	#12	0.52	5.70×10^{-2}	1.27×10^{-1}	7.35×10^{-2}	4.20×10^{-4}
8	F32	0.5	5.33×10^{-2}	1.34×10^{-1}	4.50×10^{-2}	1.75×10^{-4}
9	20/40	0.58	6.68×10^{-2}	8.69×10^{-2}	4.60×10^{-2}	2.15×10^{-4}
10	F-65	0.2	1.20×10^{-2}	1.13×10^{-1}	8.25×10^{-2}	1.14×10^{-3}
11	#12	0.52	5.70×10^{-2}	1.37×10^{-1}	2.05×10^{-1}	2.15×10^{-4}
12	16/30	0.87	1.32×10^{-1}	3.40×10^{-2}	4.95×10^{-2}	6.32×10^{-4}
13	20/30	0.75	1.03×10^{-1}	2.60×10^{-1}	1.05×10^{-1}	9.80×10^{-4}
14	f-75	0.17	9.22×10^{-3}	9.79×10^{-2}	5.70×10^{-2}	9.80×10^{-4}
15	20/40	0.58	6.68×10^{-2}	8.58×10^{-2}	7.50×10^{-2}	2.00×10^{-3}
16	mix	0.46	N/A	4.16×10^{-2}	2.68×10^{-2}	7.11×10^{-4}
17	F-85	0.15	7.29×10^{-3}	4.51×10^{-2}	4.47×10^{-2}	1.14×10^{-3}
18	20/30	0.75	1.03×10^{-1}	1.45×10^{-1}	1.16×10^{-1}	3.38×10^{-3}

[a] Data from Illman et al. (2010).

[b] If multiple ports are in the same layer, then the geometric mean is presented

Twenty-four cross-hole pumping tests were conducted in the synthetic aquifer with constant pumping rates ranging from 2.50 to 3.17 mL/s. These tests were conducted at 16 ports along columns 2 (ports 2, 8, 14, 20, 26, 32, 38, and 44) and 5 (ports 5, 11, 17, 23, 29, 35, 41, and 47), as well as at eight additional ports (ports 13, 15, 16, 18, 37, 39, 40, and 42). Prior to each pumping test, all pressure transducers were calibrated to ensure accurate data collection and head levels in all ports were monitored over several minutes to establish a static, initial condition. During each pumping test, hydraulic head responses in all 48 ports were recorded

until the aquifer reached a steady state condition which was determined by observing the stabilization of all pressure head measurements within the aquifer. The pump was then turned off to allow for the full recovery of hydraulic heads.

3.1.2 Data Used for Modeling

For the presented analyses, transient head data obtained from eight cross-hole pumping tests were utilized for inverse modeling and the remaining 16 tests were reserved for validation purposes. The distribution of eight tests selected for model calibration accommodates the upper, middle, and bottom portions of the synthetic aquifer. Prior to extracting data points from the drawdown records, the presence of drift in transducers was accounted for using the scheme discussed by Illman et al. (2007). Furthermore, head records from pumped ports were excluded from the analysis because these data were found to be excessively noisy due to the use of a peristaltic pump and skin effects. Previous research (Illman et al., 2007; Xiang et al., 2009) has shown that inclusion of data from the pumped port would lead to biased results and poor hydraulic parameter estimates. In order to smooth the data, pressure head data from observation ports were then fit with a fifth- or sixth-order polynomial curve (Liu et al., 2007), and five data points that represent the early, intermediate, and late times of aquifer responses were extracted from each curve. For ports where the pressure head curve could not be properly fit with a polynomial, five data points were extracted manually to represent the overall behavior of pressure heads. In total, 235 data points were extracted from each pumping test.

Two cases, depending on the number of pumping tests and the density of observation ports, were chosen for this study to assess the performance of different models. For Case 1, eight

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

3.2 Groundwater Flow Modeling Approaches

3.2.1 Numerical Model Setup

The synthetic aquifer was discretized into 741 elements and 1,600 nodes with element dimensions of 4.1 cm × 4.1 cm × 10.2 cm for all forward and inverse groundwater flow models considered. A finer mesh was also tested in a previous study (Illman et al., 2012), but the results did not show significant changes in comparison to the coarser one. Therefore, for consistency with previous studies (Berg and Illman, 2011a; Illman et al., 2010; Zhao et al., 2016) who used this discretization, this coarse grid was utilized here.

For groundwater flow modeling, the side and top boundaries of the sandbox were set as constant head, while the bottom, front, and back boundaries were considered to be no-flow. In this study, the transient flow Eqs. 1.2 and 1.3 were solved by a 3-D finite element model

VSAFT3 (Yeh et al., 1993) which is implemented in the geostatistical inversion code SimSLE (details are provided in later sections). To be consistent with different modeling approaches, this forward model was utilized to generate head distributions of pumping tests for all cases described in following sections.

3.2.2 Effective Parameter Model

The synthetic aquifer was first characterized as a homogeneous, isotropic medium to estimate the effective K and S_s values by coupling the groundwater flow model VSAFT3 (Yeh et al., 1993) with the parameter estimation code PEST (Doherty, 2015). We did not consider the case of anisotropic hydraulic parameters because previous research by Illman et al. (2015) showed little difference in the isotopic and anisotropic results for this laboratory sandbox aquifer.

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

For each case, all pumping test data were included for the estimation of effective parameters. In total, 1,880 data points were used for Case 1, while 300 data points were utilized for Case 2. The forward model was then automatically calibrated to obtain an optimal set of K and S_s by simultaneously matching all data points. The initial values of K and S_s input into

PEST are 0.06 cm/s and 6.1×10^{-4} /cm, respectively, which are the geometric means from the kriged K and S_s fields based on the estimates from single-hole pumping tests (Berg and Illman, 2011a). In PEST, the minimum and maximum bounds were set as 1×10^{-4} and 10 cm/s for K, and 1×10^{-8} and 1.0 /cm for S_s.

3.2.3 Geology-based Zonation Modeling Approach

The synthetic aquifer was then characterized using various geology-based zonation models. In this approach, the synthetic aquifer was divided into different zones based on available geological information. In each zone, the porous medium was treated to be homogeneous as well as isotropic, and a uniform set of K and S_s was estimated and assigned to describe its hydraulic properties. To assess the impact of accuracy of geological information on groundwater flow modeling, four geology-based zonation models with various accuracy and resolution (GOOD, POOR1, POOR2, and POOR3, as shown in Fig. 3.2) were considered, following the work of Zhao et al. (2016).

These geological models were constructed using Leapfrog Hydro (ARANZ Geo Ltd.) through the inclusion of borehole information obtained along the six columns of ports. In comparison to directly mapping the stratification through the sandbox glass (Fig. B1 in Appendix B: the "perfect" geological model utilized in Berg and Illman (2011a) and Illman et al. (2015)), interpolation of borehole logs is more consistent in constructing geological models from field data. Among the four geological models, GOOD (Fig. 3.2a) was constructed based on the accurate stratigraphy information along the six well columns, representing the ideal

scenario for the construction of geological model. For POOR1 (Fig. 3.2b), we introduced random errors to the layer thickness while maintaining layer sequences, mimicking the field case of poor recovery of core samples during drilling. The geological model POOR2 (Fig. 3.2c) was constructed based on the GOOD (Fig. 3.2a) model by merging some layers with similar materials. This simplified geological model with only five zones was constructed to represent the scenario with simplified description of the stratigraphy, but with well identified layer boundaries. For POOR3 (Fig. 3.2d), we provided incorrect stratigraphy information for model construction, yielding the poorest description of geological layers among four models.

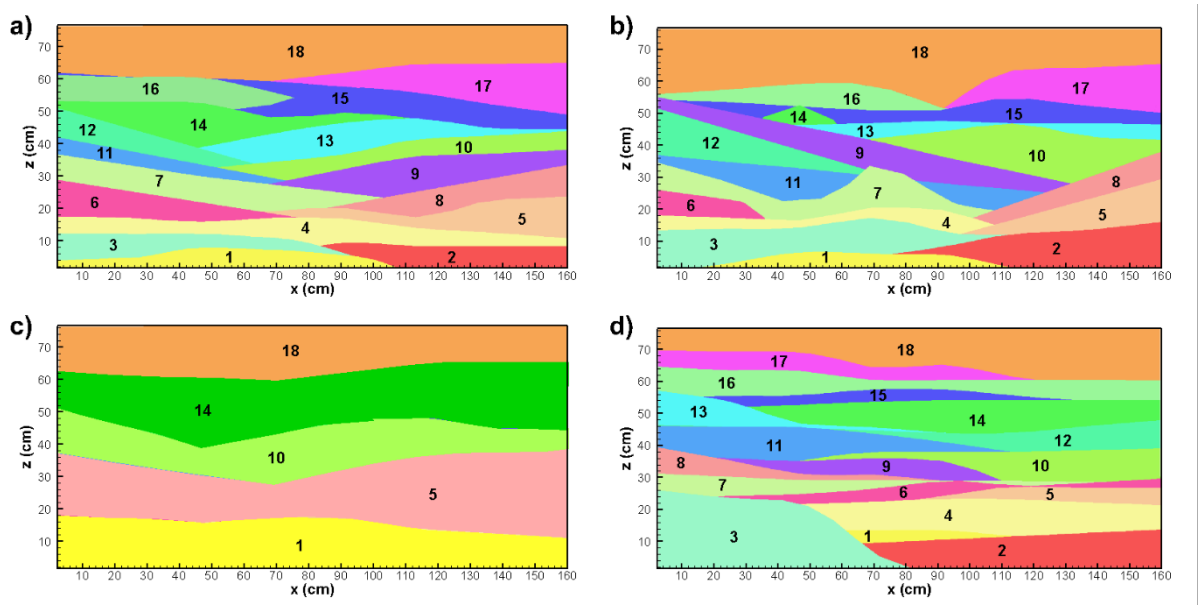

Fig. 3.2: Geological Models of varying accuracy and resolution: (a) GOOD; (b) POOR1; (c) POOR2; (d) POOR3.

The four geological models were then discretized using the grid described above to construct geology-based zonation models for aquifer characterization. In a similar fashion to the effective parameter model, all geology-based zonation models were calibrated using PEST coupled with VSAFT3 by simultaneously matching all data points. For each case study, the *K*

and S_s values obtained from the effective model were used as the initial guesses of hydraulic parameters for model calibration, while the bounds of K and Ss were set to be the same as those in the effective parameter model. In total, 36 parameters were estimated for geology-based zonation models GOOD, POOR1, and POOR3, while only 10 parameters were estimated for the POOR2 model.

3.2.4 Geostatistical Inverse Modeling Approach

All geostatistical inversions were conducted using the Simultaneous Successive Linear Estimator (SimSLE), developed by Xiang et al. (2009). This inversion approach provides an efficient way to include all data points from multiple pumping tests simultaneously for hydraulic parameter estimation. In comparison to the Sequential Successive Linear Estimator (SSLE) developed for THT analysis (Yeh and Liu, 2000), SimSLE provides more constraints to the inverse problem, resulting in faster convergence (Xiang et al., 2009). Additionally, SimSLE avoids the computation of varying final estimates when HT data are analyzed in different sequences with SSLE (Illman et al., 2008). The fundamental concept of SimSLE and the mathematical description of the implemented SLE are discussed in Section 1.2 and Appendix A, respectively.

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

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

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

3.3 Results and Discussion

3.3.1 Effective Parameter Model

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

The estimated effective K and S_s from the two cases are close to each other, while the 95% confidence intervals for both K and S_s increase after reducing the number of data points used for inverse modeling. This result implies that more pumping tests with dense observation locations may still be required to obtain reliable estimates of effective parameters.

Previously, Berg and Illman (2011a) determined the effective parameters of the synthetic aquifer by taking the geometric means of 48 single-hole K and S_s estimates ($K = 6.0 \times 10^{-2}$ cm/s and $S_s = 6.1 \times 10^{-4}$ /cm). These values, however, were found to be poor in predicting drawdowns from independent pumping tests, suggesting that the effective parameters obtained from local estimates may not be representative of the aquifer. Illman et al. (2015) reached the same conclusion based on the effective K that they estimated for the same synthetic aquifer. In particular, they found that the effective K estimated by simultaneously analyzing eight pumping tests provided improved results in terms of model calibration and validation in comparison to the work of Illman et al. (2010), in which, local estimates (e.g., core and single-hole test results) were utilized to generate the effective K value of the aquifer.

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

3.3.2 *Geology-based Zonation Models*

Four different geology-based zonation models were calibrated for each dataset case. Fig. 3.3 and Fig. 3.4 show the estimated K and S_s tomograms from different zonation models for Cases 1 and 2, respectively. Examination of these two figures reveals that the estimated locations of high and low K zones vary from one zonation model to another, when the same number of head data were included for model calibration. This is because these models are calibrated with fixed zones of parameters. With a fixed geological model, PEST focuses on the estimation of parameter values of each zone to fit the simulated to observed data as close as possible. These results imply that attention should be paid when constructing zonation models for aquifer heterogeneity characterization, since the inaccurate identification of structural features could lead to unrealistic parameter estimates. In addition, the comparison of results from Cases 1 and 2 (Fig. 3.3 and Fig. 3.4) when the same zonation model was calibrated reveals that the identification of high and low K zones varies when different numbers of head data were included for model calibration.

The estimated S_s tomograms do not show distinct structural features, except for the simplified zonation model POOR2 (Fig. 3.3f and Fig. 3.4f), revealing that the heterogeneity of S_s in this synthetic aquifer is milder in comparison to that of the K. Moreover, the estimated S_s values decrease from the top to the bottom for most zonation models. Such a decreasing trend of S_s was explained by Berg and Illman (2011a) that the upper sands were less compressed compared to the deeper sand bodies. Although the spatial variance of S_s estimates is relatively small, differences in estimated S_s tomograms can still be observed.

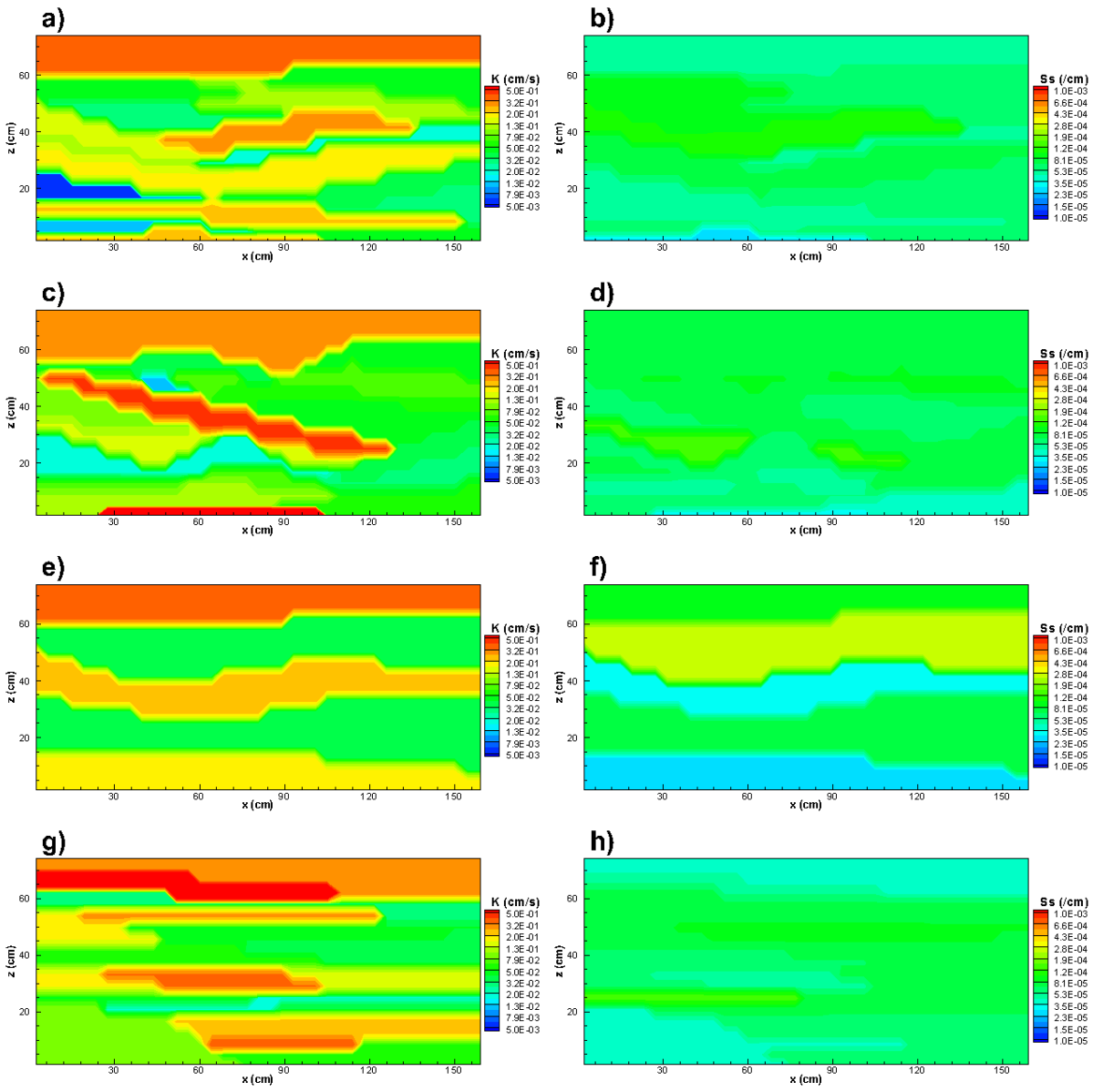

Fig. 3.3: K and S_S tomograms estimated from geology-based zonation models for Case 1. K tomograms: (a) GOOD; (c) POOR1; (e) POOR2; (g) POOR3. S_s tomograms: (b) GOOD; (d) POOR1; (f) POOR2; (h) POOR3.

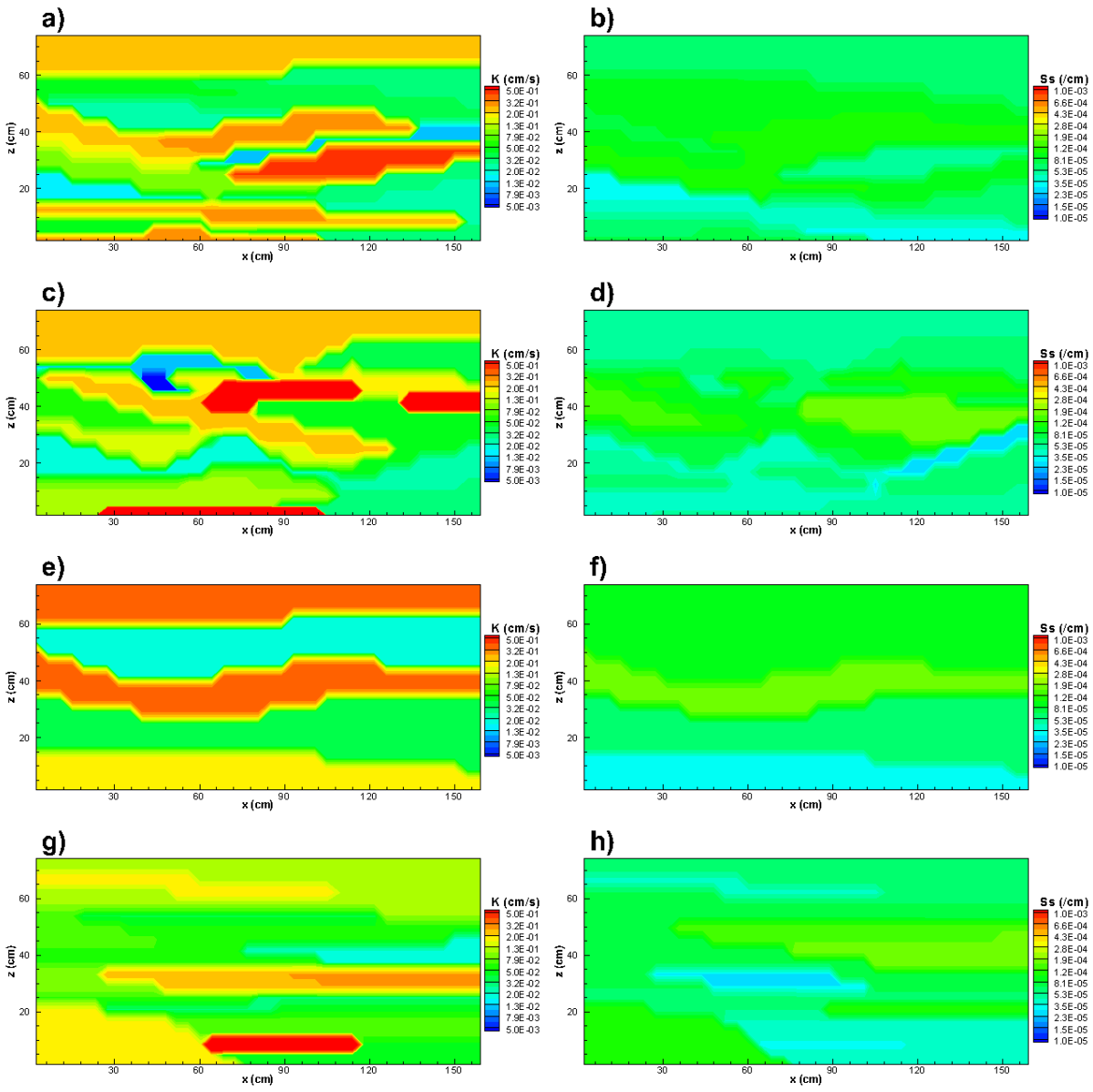

Fig. 3.4: K and S_S tomograms estimated from geology-based zonation models for Case 2. K tomograms: (a) GOOD; (c) POOR1; (e) POOR2; (g) POOR3. S_s tomograms: (b) GOOD; (d) POOR1; (f) POOR2; (h) POOR3.

Within each zone, a uniform set of K and S_s was estimated to describe its hydraulic properties. Fig. 3.5 shows the estimated K values as well as their corresponding 95% confidence intervals from all zonation models for both cases, while Fig. 3.6 shows the same, but for S_s estimates. The estimated K and S_s values as well as their 95% confidence intervals are provided in the Supplementary Material section as Table B1 to Table B4 for the GOOD, POOR1, POOR2, and POOR3 models. Fig. 3.5 reveals that when calibrating geology-based zonation models with eight pumping tests (Case 1), all K estimates have narrow confidence intervals, suggesting the high confidence of these estimates. However, when the number of head data is reduced, noticeable increases in the confidence intervals of K estimates are observed in some zones; in particular, Zone 6 for the POOR1 model and Zones 8, 13, and 18 for the POOR3 model. The main reason for this is that no observation data are available in these zones when observation ports are reduced from 47 (Case 1) to 15 (Case 2).

It is interesting to note that all S_s estimates from the calibrated geology-based zonation models result in narrow confidence intervals (Fig. 3.6), except for the case in which the simplified geology-based zonation model (POOR2) is calibrated using fewer data (as shown in green areas of Fig. 3.6c). With given structural features, the obtained S_s estimates for different zones are close to each other. This is also the case when comparing S_s estimates from different zonation models. These results suggest that the estimation of S_s for this synthetic aquifer is less likely to be affected by structural errors.

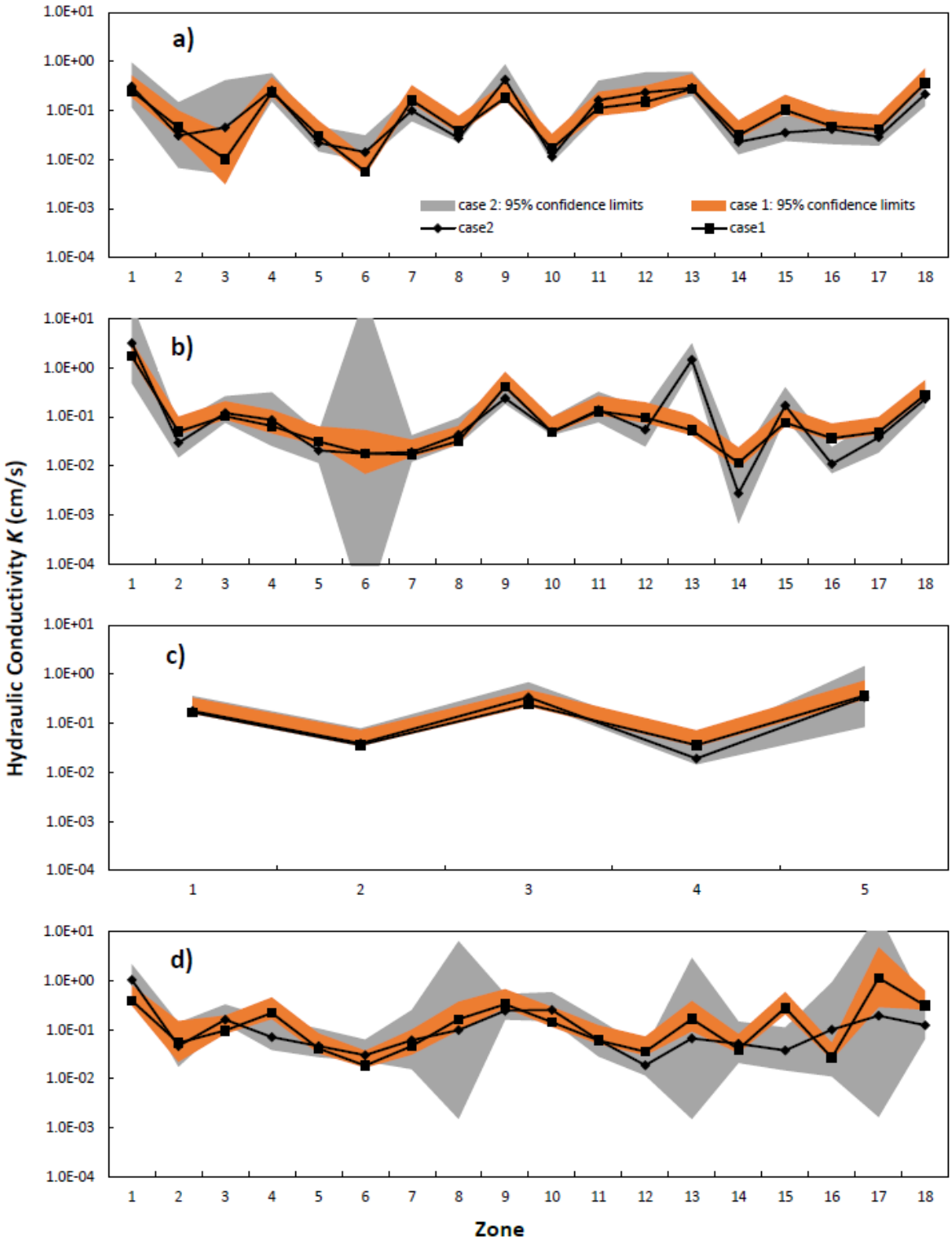

Fig. 3.5: Estimated *K* values and corresponding 95% confidence intervals of Cases 1 and 2 for four different geology-based zonation models: (a) GOOD; (b) POOR1; (c) POOR2; (d) POOR3.

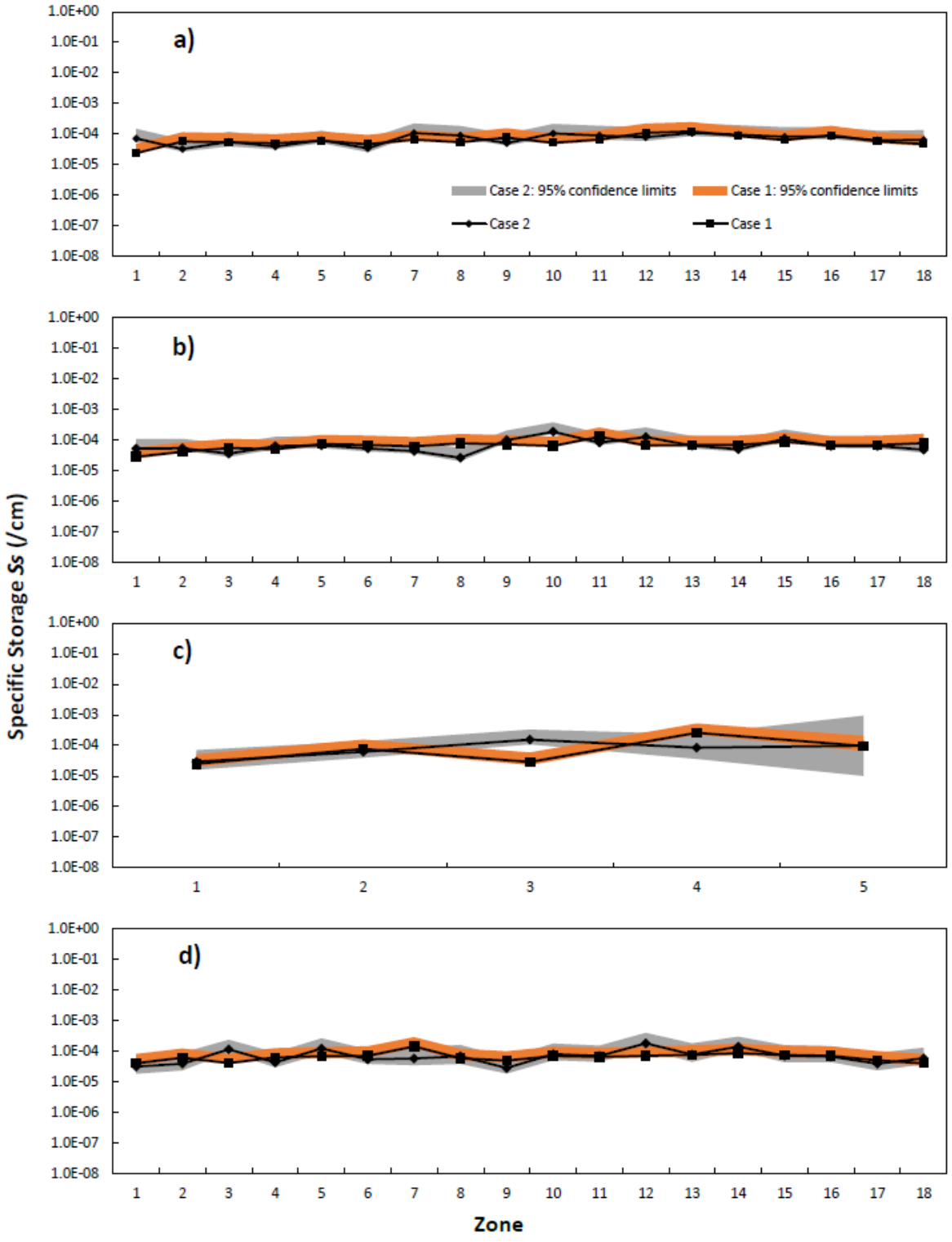

Fig. 3.6: Estimated *K* values and corresponding 95% confidence intervals of Cases 1 and 2 for four different geology-based zonation models: (a) GOOD; (b) POOR1; (c) POOR2; (d) POOR3.

3.3.3 *Geostatistical Inverse Model with Homogeneous Initial K and Ss Fields*

Without providing additional prior information, the geostatistical inversion of THT data using SimSLE started with homogeneous initial K and S_s fields. In SimSLE, the L_2 norm between the simulated and observed head is computed for each iteration. Fig. B2 in Appendix B shows how the L_2 norm evolves with the iteration number. As suggested by Xiang et al. (2009), we selected the inversion results when the L_2 norm stabilized. Here, stabilization is meant when the variation of L_2 between two successive iterations is smaller than 3×10^{-4} cm^2.

Fig. 3.7 shows the estimated K and S_s tomograms as well as the corresponding ln$K$ and lnS_s variance maps for Case 1, while Fig. 3.8 shows the same, but for Case 2. Black lines in K tomograms indicate the exact layer boundaries of the synthetic aquifer, which were delineated based on the photograph of the deposits (see Fig. 3.1). Such stratigraphic information is not included for the estimated S_s tomograms because the spatial variation of S_s of this synthetic aquifer does not reveal distinct structural features. Different from the zonation modeling approach, the geostatistical inversion of THT data estimates hydraulic parameters for each finite element, resulting in relatively smooth distribution of K and S_s estimates. Through the simultaneous inversion of transient head data from eight pumping tests with 47 observation ports (Case 1), the estimated K tomogram (Fig. 3.7a) reveals considerable details to aquifer heterogeneity. The estimated high and low K zones show significant agreement to most visible layers of the synthetic aquifer in terms of their positions. When fewer pumping tests with fewer observation ports were utilized for the geostatistical inversion (Case 2), the estimated K tomogram (Fig. 3.8a) shows a similar pattern of high and low K zones, but with great loss of

detail in heterogeneity, particularly on both sides of the aquifer where observation data were removed for model calibration.

Although the estimated S_s of this synthetic aquifer is much less variable when compared to K, the loss of detail in heterogeneity can still be observed in the estimated S_s tomograms (Fig. 3.7c and Fig. 3.8c for Cases 1 and 2, respectively). The comparison of Cases 1 and 2 reveals that a large number of pumping tests with dense observation intervals is required to capture most heterogeneity features, which is in line with the conclusion of previous studies (e.g., Cardiff et al., 2013; Illman et al., 2015). This study shows that the inclusion of additional hydraulic head data through transient analysis does not negate the necessity of a large number of monitoring points to accurately depict the structural features and their boundaries.

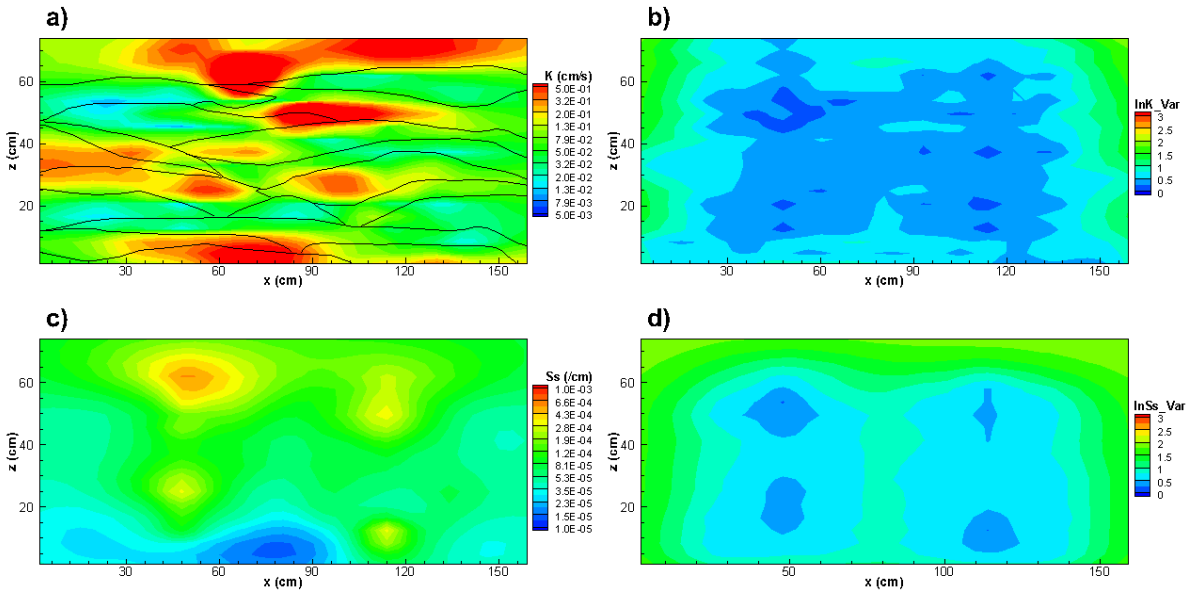

Fig. 3.7: K and S_s tomograms and their corresponding variances for Case 1 with homogeneous initial K and S_s fields. (a) K tomogram; (b) ln$K$ variances; (c) S_s tomogram; (d) lnS_s variances. Black lines in the K tomogram represent the accurate stratification of the synthetic aquifer.

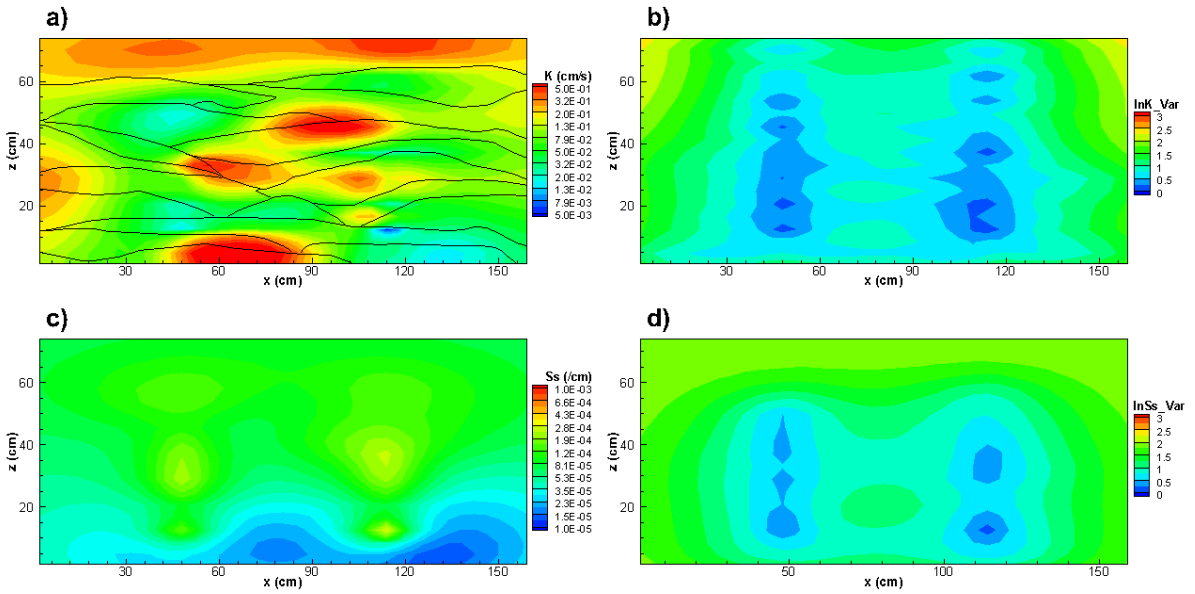

Fig. 3.8: K and S_s tomograms and their corresponding variances for Case 2 with homogeneous initial K and S_s fields. (a) K tomogram; (b) ln$K$ variances; (c) S_s tomogram; (d) lnS_s variances. Black lines in the K tomogram represent the accurate stratification of the synthetic aquifer.

The corresponding ln$K$ and lnS_s variances computed by SimSLE indicate the uncertainty of parameter estimates, with larger variance values indicating higher uncertainty. For each case, small ln$K$ and lnS_s variances were obtained around pumping ports, while variances become larger when moving away from the ports. In general, for both cases, the lnS_s variances are larger than those of ln$K$, revealing that it is more difficult to estimate S_s. Comparing variance maps from Cases 1 and 2, a significant increase in values is observed for both ln$K$ and lnS_s variances when including fewer head data for geostatistical inversions. These results indicate that more accurate K and S_s tomograms will be obtained in areas where there are available head data in comparison to areas where head data are lacking, which again emphasizes the importance of the availability of head data for aquifer heterogeneity characterization using geostatistical models.

The geostatistical inverse modeling of transient head data using SimSLE is demonstrated

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

3.3.4 Geostatistical Inverse Model with Heterogeneous Initial K and S_s Fields

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

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

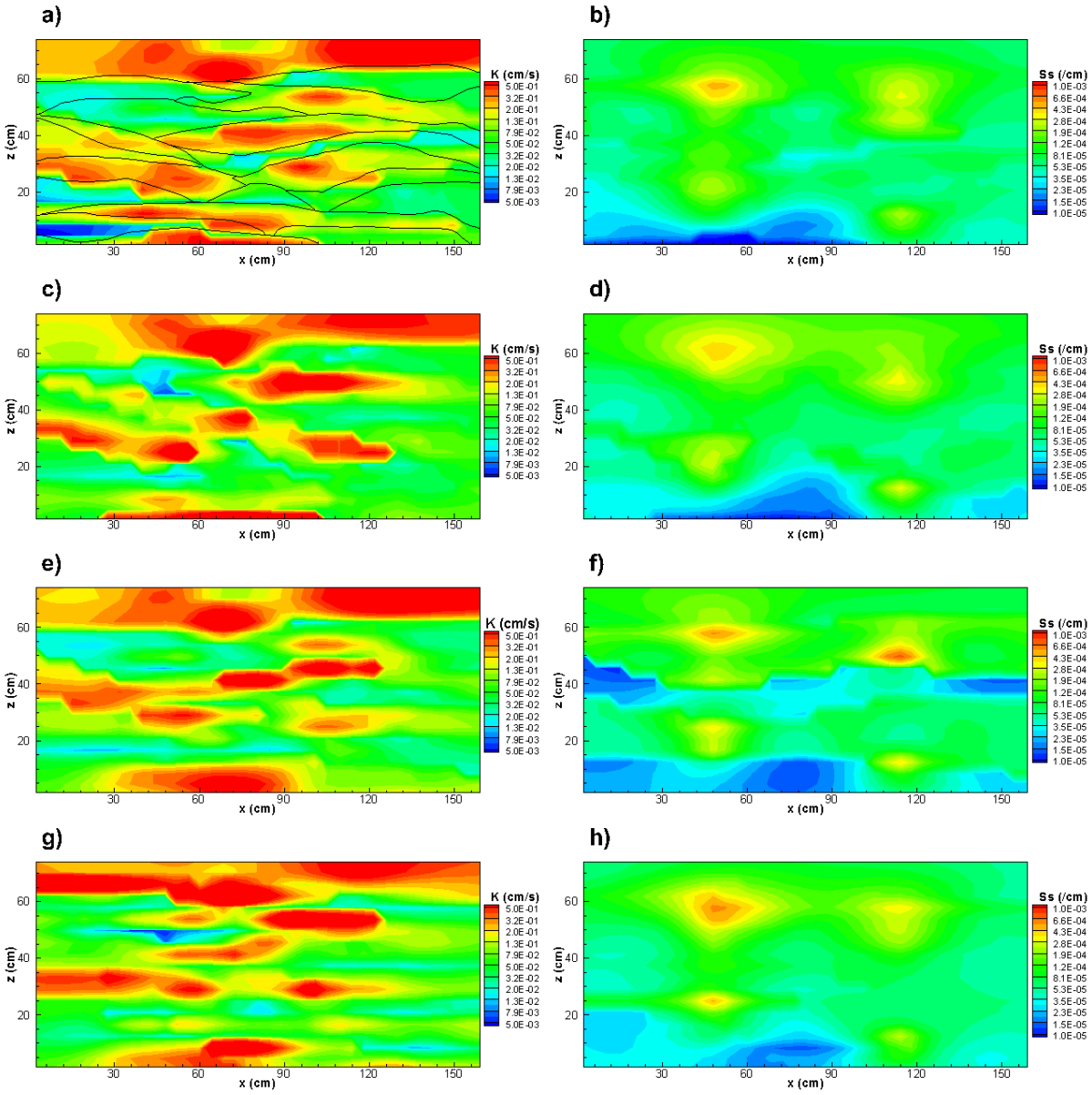

Fig. 3.9: K and S_S tomograms estimated from geostatistical models with heterogeneous initial K and S_s fields for Case 1. K tomograms: (a) GOOD; (c) POOR1; (e) POOR2; (g) POOR3. S_s tomograms: (b) GOOD; (d) POOR1; (f) POOR2; (h) POOR3. Black lines in (a) represent the accurate stratification of the synthetic aquifer.

Fig. 3.9 illustrates the estimated K and S_s tomograms for Case 1 (see Fig. B2a for L_2 norms). In particular, Fig. 3.9a and b show the K and S_s tomograms, respectively, when the GOOD geological information is incorporated into the geostatistical inversion of THT data. Black lines that represent the accurate stratification of the synthetic aquifer are also included in the K

tomogram. In comparison to the estimated K and S_s tomograms with homogeneous initial fields (Fig. 3.7), the incorporation of good prior stratigraphic information preserves more heterogeneity features and layer boundaries regarding the K tomogram; however, S_s tomograms are similar to each other.

The estimated K and S_s tomograms when including inaccurate geological information as initial guesses are shown as Fig. 3.9c - 3.9h. The estimated K tomograms are similar in terms of the patterns of estimated high and low K zones, while the shapes and the continuity of these zones are slightly different among each other. Differences can also be observed in the estimated S_s tomograms when geological information of varying resolution and accuracies were incorporated into inverse modeling. In particular, when the simplified geological information (POOR2) was included as an initial guess, unexpected low S_s zones were estimated across the middle of the simulation domain (Fig. 3.9f).

Fig. 3.10 shows the estimated K and S_s tomograms for Case 2 (see Fig. B2b for L_2 norms) when four types of geological information were introduced separately for inverse modeling. Through the incorporation of the GOOD geological model as prior information, the estimated K tomogram (Fig. 3.10a) reveals more heterogeneity details compared to the one obtained with homogeneous initial parameter fields (Fig. 3.8a), particularly at the fringes of the aquifer where observation ports are lacking. Even with limited head data, the estimated high and low K zones still show significant agreement with most layers. However, the estimated S_s tomogram does not show distinct change in comparison to the one without geological information (Fig. 3.8c).

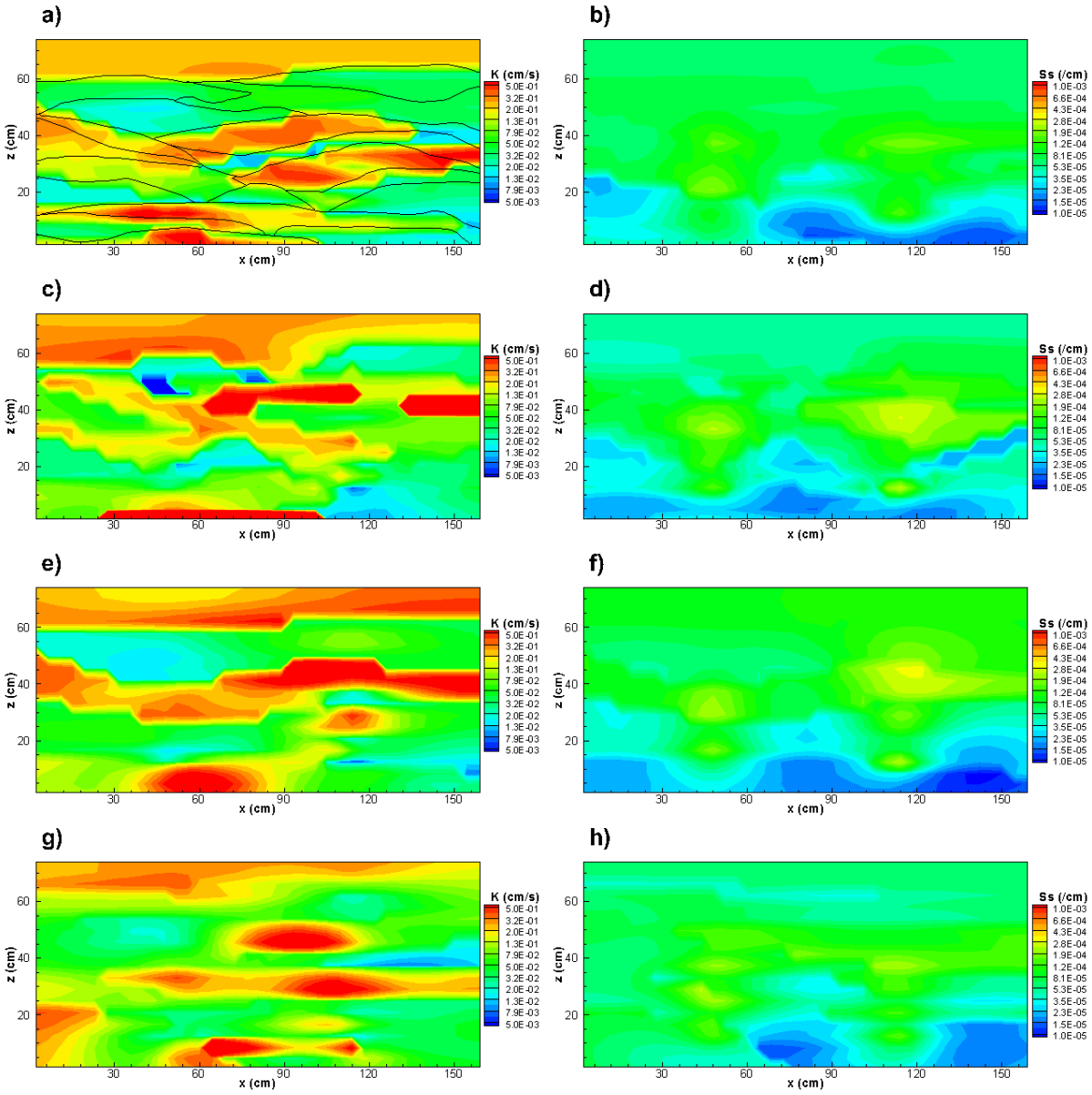

Fig. 3.10: K and S_S tomograms estimated from geostatistical models with heterogeneous initial K and S_s fields for Case 2. K tomograms: (a) GOOD; (c) POOR1; (e) POOR2; (g) POOR3. S_s tomograms: (b) GOOD; (d) POOR1; (f) POOR2; (h) POOR3. Black lines in (a) represent the accurate stratification of the synthetic aquifer.

In contrast to Case 1 results, the estimated K tomograms in Case 2 are quite different among each other when various geological information are incorporated. Similar differences in estimated S_s tomograms are also visible in terms of the pattern of high and low S_s zones as well as their shapes. This suggests that when pumping and observation densities are high, hydraulic head data will dominate the inversion process and lead to similar K and S_s tomograms

rather than reflect the prior geological information. However, the effects of prior geological information on inverse modeling become more significant when fewer pumping test data are available for calibration. The accuracy of these K and S_s tomograms are examined in later sections through the investigation of their abilities in predicting drawdowns from independent pumping tests.

3.3.5 Model Calibration and Validation

The calibration and validation results associated with different models were first assessed qualitatively by plotting the scatterplots of simulated versus observed drawdowns, which provides visual information of the spatial distribution of errors in terms of scatter and bias. Then, quantitative evaluation of model error was performed by computing the mean absolute errors (L_1), mean square errors (L_2), and coefficient of determination (R^2) between simulated and observed drawdown values using:

$$L_1 = \frac{1}{n}\sum_{i=1}^{n} |\, x_i - \hat{x}_i| \qquad (Eq.\,3.1)$$

$$L_2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \qquad (Eq.\,3.2)$$

$$R^2 = \left[\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(\hat{x}_i - \mu_{\hat{x}})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2 \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - \mu_{\hat{x}})^2}}\right]^2 \qquad (Eq.\,3.3)$$

where n is the total number of drawdown data, i indicates the data number, x_i and $\hat{x}_i$ represent *i-th* simulated and observed drawdown values, respectively, μ_x and $\mu_{\hat{x}}$ represent

averaged simulated and observed head data, respectively. The L_1 norm is calculated to analyze the discrepancy between simulated and observed drawdowns, while the L_2 norm tends to magnify large discrepancies and allow one to better assess the performance of different models. R^2 shows the correspondence between the simulated and observed drawdowns.

The calibration scatterplots of all investigated models are provided in Appendix B and illustrated as Fig. B3 and Fig. B4 for Case 1, and as Fig. B5 and Fig. B6 for Case 2. A linear model fits to all scatters and the corresponding coefficient of determination (R^2) values are provided. These scatterplots reveal that the calibration result improves when a larger number of estimated parameters are considered in the inverse model, and the geostatistical model yields the best result. This makes sense since the highly parameterized geostatistical model allows for the adjustment of K and S_s estimates in each element to fit the observation data.

For Case 1, the validation scatterplots that compare the simulated drawdown values from different models against the observed drawdowns from 16 independent pumping tests are illustrated in Fig. 3.11 and Fig. 3.12. In each scatterplot, the linear model of the fit, as well as the coefficient of determination (R^2) are provided at the bottom. Examination of Fig. 3.11 shows that, by comparing the three different modeling approaches, the geostatistical model performs the best in predicting drawdowns for the entire domain, closely followed by the GOOD geology-based zonation model, while the utilization of effective homogeneous model yields biased predictions of drawdowns with relatively larger scatter. This is consistent with the SSHT results of Illman et al. (2015). On the other hand, it is interesting to note that the

zonation models based on inaccurate geological information yield slightly better prediction results in comparison to the effective homogeneous model. The main reason for this is that more parameter sets are estimated for zonation models (18 for POOR1 and POOR3, and 5 for POOR3), which in turn emphasizes the importance of parameterization for groundwater flow modeling.

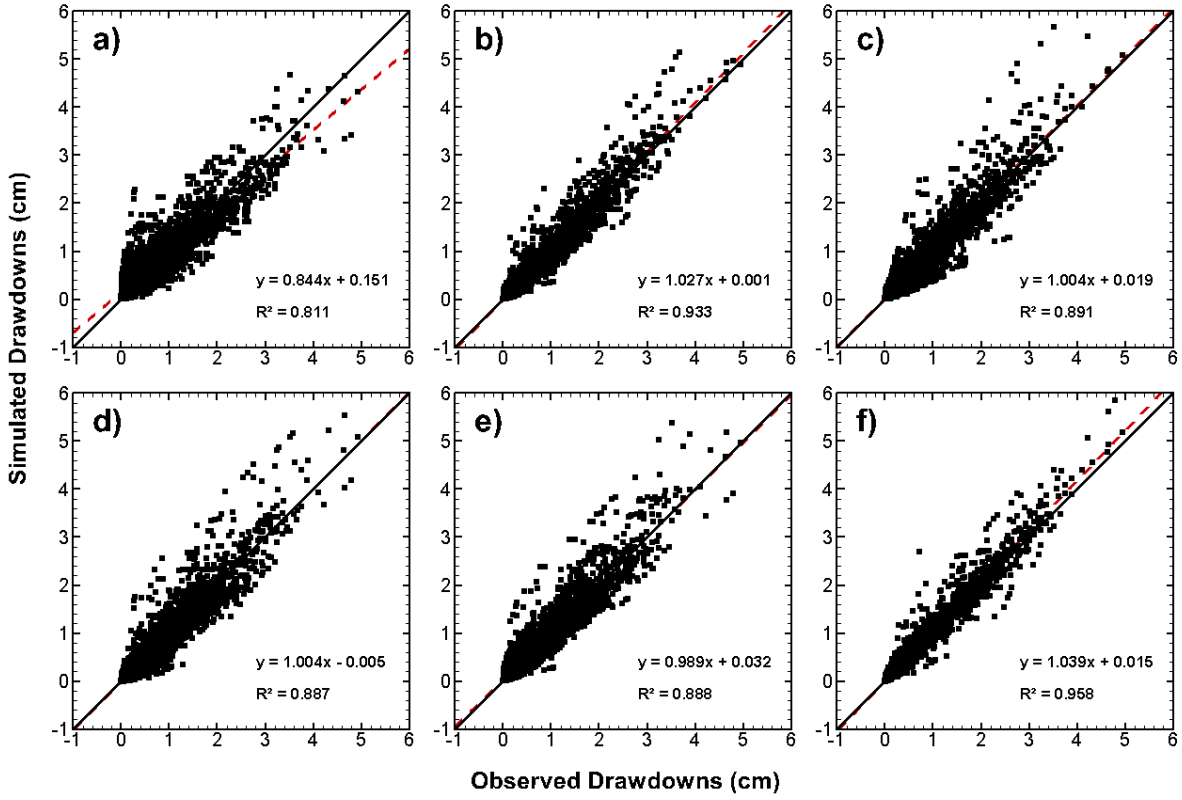

Fig. 3.11: Validation scatterplots (Case 1) of simulated versus observed drawdowns for different modeling approaches. (a) effective model, (b)-(e) four geology-based zonation models: (b) GOOD, (c) POOR1, (d) POOR2, (e) POOR3; and (f) geostatistical model with homogeneous initial parameter fields.

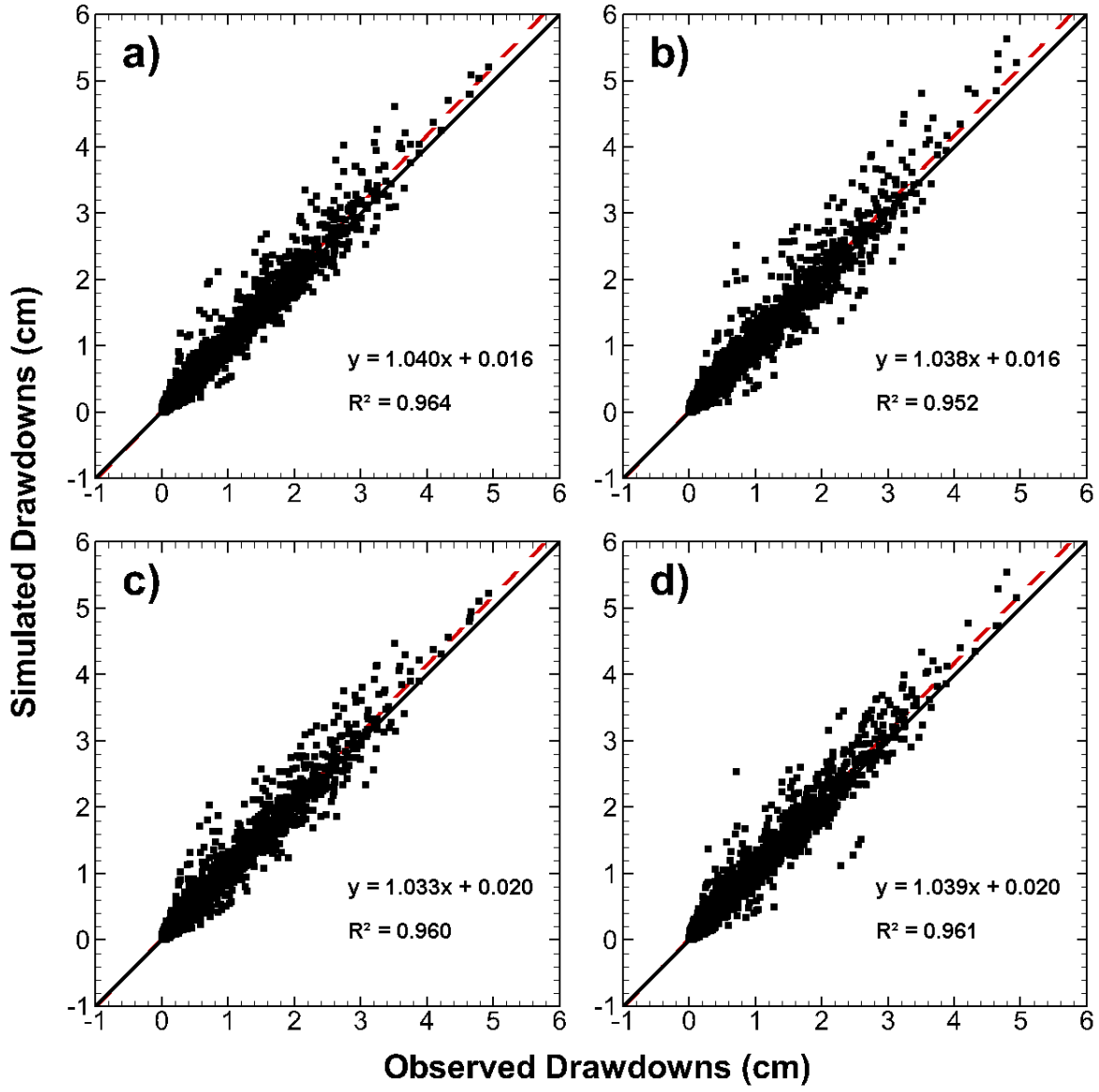

Fig. 3.12: Validation scatterplots (Case 1) of simulated versus observed drawdowns for geostatistical models incorporated with four different types of geological information. (a) GOOD, (b) POOR1, (c) POOR2, and (d) POOR3.

Examination of Fig. 3.12 reveals that the estimated K and S_s tomograms from the geostatistical inverse model with GOOD geological information (Fig. 3.12a) as initial parameter fields yield minor improvements in predicting drawdown values for the entire domain compared to the case with homogeneous initial parameter fields (Fig. 3.11f). On the other hand, the incorporation of inaccurate geological information does not significant impact

the prediction results. This makes sense because the inversion results reflect more about hydraulic head information rather than prior geological information when abundant head data are available for inverse modeling, and the resulting K and S_s tomograms perform similarly in predicting independent pumping test data.

For Case 2, the validation scatterplots are illustrated in Fig. 3.13 and Fig. 3.14. The effective homogeneous model still performs the worst in predicting drawdowns from independent pumping tests. However, it is surprising to find that the GOOD geology-based zonation model provides prediction results that are indistinguishable to the geostatistical inverse model with homogeneous initial K and S_s fields. This finding suggests that when the number of head data is limited, the utilization of geological model with good knowledge of stratification yields results that are comparable to those obtained by the geostatistical model, which is in line with the conclusion provided by Illman et al. (2015), who only analyzed steady state head data. However, it should be noted that the GOOD geology-based zonation model utilized in this study is constructed based on a large amount of borehole data with accurate identification of stratifications, which is difficult to obtain in the field. On the other hand, the validation results associated with other geology-based zonation models (POOR1, POOR2, and POOR3) are not as good as the result provided by the geostatistical inverse model.

After reducing head data for inverse modeling, some differences in validation scatterplots are evident for geostatistical inverse models with different heterogeneous initial parameter fields, as shown in Fig. 3.14. Through the incorporation of GOOD geological information, the

estimated K and S_s tomograms provide improved prediction results (Fig. 3.14a) for the entire domain with higher correlation between simulated and observed drawdowns in comparison to the case with homogeneous initial parameter fields (Fig. 3.13f). Slight improvements in terms of bias and scatter are also observed in Fig. 3.14c, in which simplified geological information (POOR2) is incorporated. On the other hand, when inaccurate stratigraphy and layer thickness information are introduced during model calibration, the estimated K and S_s tomograms provide worse prediction results (Fig. 3.14b and Fig. 3.14d for POOR1 and POOR3, respectively) in comparison to the case with homogeneous initial parameter fields.

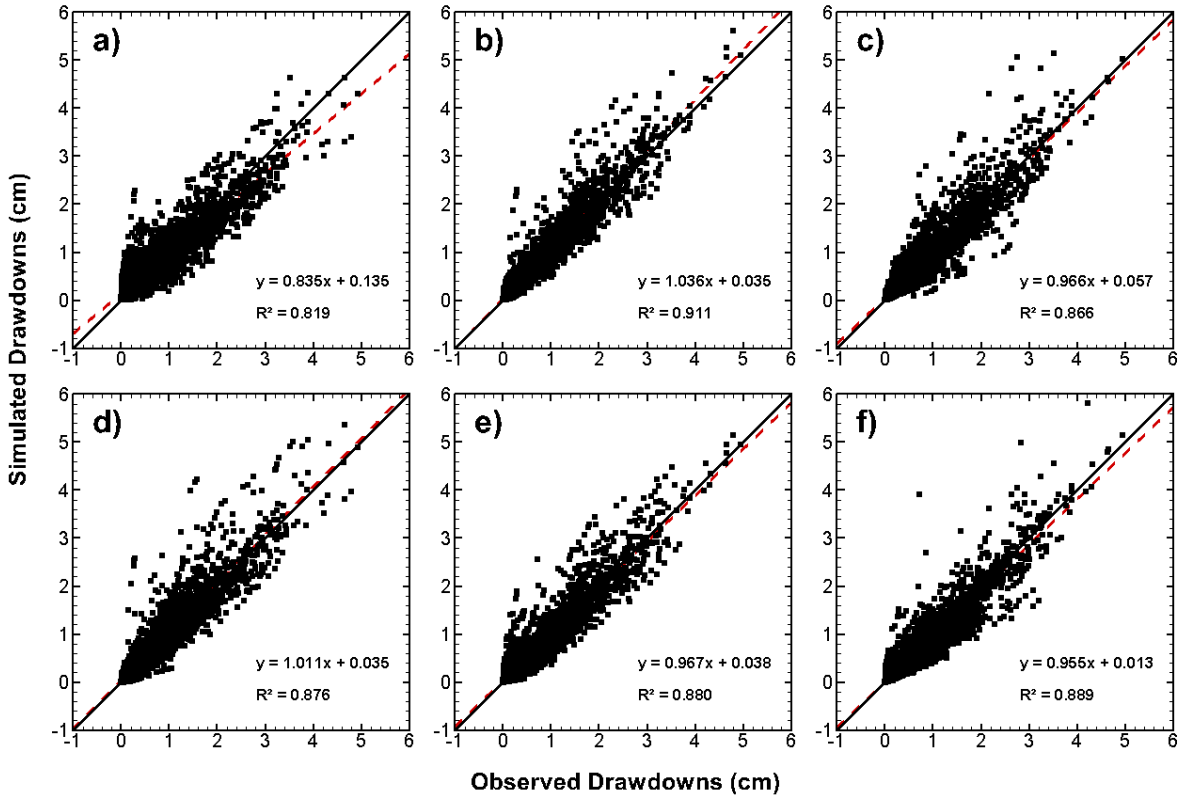

Fig. 3.13: Validation scatterplots (Case 2) of simulated versus observed drawdowns for different modeling approaches. (a) effective model, (b)-(e) four geology-based zonation models: (b) GOOD, (c) POOR1, (d) POOR2, (e) POOR3; and (f) geostatistical model with homogeneous initial parameter fields.

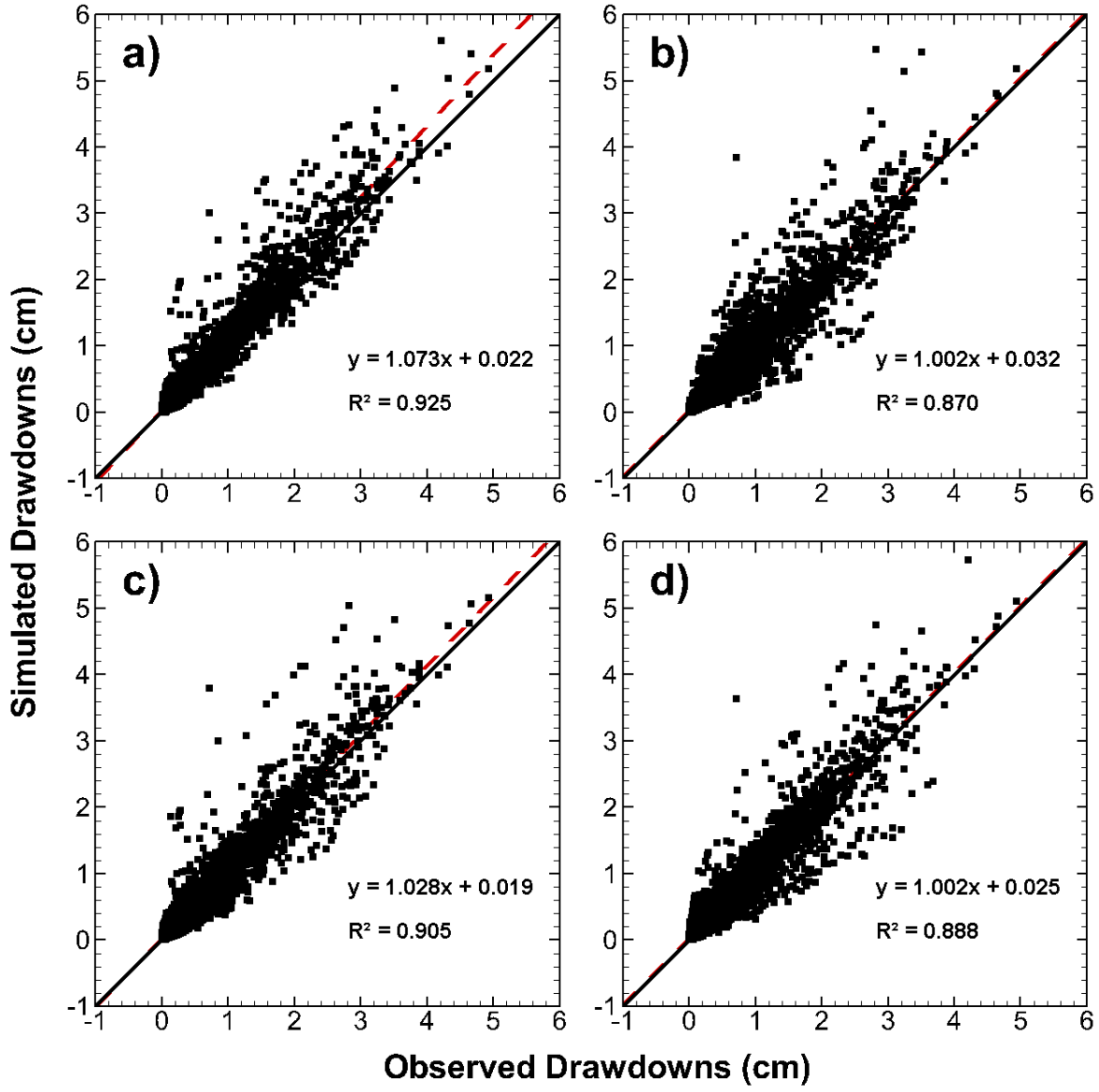

Fig. 3.14: Validation scatterplots (Case 2) of simulated versus observed drawdowns for geostatistical models incorporated with four different types of geological information. (a) GOOD, (b) POOR1, (c) POOR2, and (d) POOR3.

The L_1 and L_2 norms of calibration and validation results are summarized in Table 3.2 and Table 3.3 for Cases 1 and 2, respectively. These values were obtained by averaging the results from different pumping tests for each model calibration and validation, while the norms associated with individual pumping tests are presented in Appendix B as Table B5 - Table B8 for Case 1 and as Table B9 - Table B12 for Case 2. For both cases, the highly parameterized

geostatistical model performs consistently across the different pumping tests, suggesting that the approach is more consistently reliable in characterizing aquifer heterogeneity and predicting drawdowns in comparison to the effective homogeneous and geology-based zonation modeling approaches. However, the highly parameterized geostatistical model may suffer from the issue of over-parameterization and lead to ill-posed inversion problems. In this case, a large number of dataset (e.g., transient head responses obtained from HT) is required for the geostatistical inversion model to estimate reliable spatial distributions of hydraulic parameters (Schöniger et al., 2015). After incorporating geological information into geostatistical model calibration, L_1 and L_2 norms are found to be comparable for Case 1, while significant differences are observed for Case 2, especially for validation results.

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

Table 3.2: Summary of L_1 and L_2 norms of calibration and validation results for Case 1.

		Calibration		Validation	
		L_1	L_2	L_1	L_2
Effective Parameter Model		0.250	0.124	0.223	0.096
Geology-based Zonation Model	**GOOD**	0.102	0.027	0.118	0.038
	POOR1	0.112	0.029	0.152	0.061
	POOR2	0.130	0.052	0.151	0.064
	POOR3	0.135	0.046	0.159	0.061
Geostatistical Model with Homogeneous Initial K and S_s Fields		0.046	0.005	0.091	0.026
Geostatistical Model with Heterogeneous Initial K and S_s Fields	**GOOD**	0.050	0.006	0.087	0.023
	POOR1	0.051	0.005	0.096	0.030
	POOR2	0.050	0.005	0.090	0.024
	POOR3	0.050	0.005	0.091	0.025

Table 3.3: Summary of L_1 and L_2 norms of calibration and validation results for Case 2.

		Calibration		Validation	
		L_1	L_2	L_1	L_2
Effective Parameter Model		0.317	0.167	0.220	0.095
Geology-based Zonation Model	**GOOD**	0.079	0.016	0.139	0.059
	POOR1	0.096	0.020	0.168	0.077
	POOR2	0.179	0.090	0.156	0.077
	POOR3	0.159	0.043	0.175	0.067
Geostatistical Model with Homogeneous Initial K and S_s Fields		0.042	0.004	0.149	0.061
Geostatistical Model with Heterogeneous Initial K and S_s Fields	**GOOD**	0.048	0.006	0.128	0.057
	POOR1	0.050	0.006	0.159	0.080
	POOR2	0.048	0.005	0.129	0.060
	POOR3	0.047	0.005	0.149	0.067

3.3.6 *Predictability of Transient Drawdown Curves*

To further investigate the performance of different models in predicting independent pumping tests, simulated drawdown curves at 16 selected ports were plotted against actual data. Fig. 3.15 and Fig. 3.16 illustrate the simulated drawdown curves using K and S_s tomograms from different models when conducting a pumping test at port 40 for Case 1, while Figs. 3.17 and 3.18 illustrate the same, but for Case 2. Results for all other pumping tests used for model validation are provided in Appendix B as Fig. B7 - Fig. B66. In each subplot, the observation data are expressed as green dots, while the simulated drawdown curves from the various models are plotted with different colors and types.

Fig. 3.15 shows that when a great number of head data is used for inverse modeling, the utilization of K and S_s tomograms from the geostatistical model with homogeneous initial parameter fields is able to predict drawdowns at most of the ports, followed by the zonation model with GOOD geological information. The performance of the effective model, as well as other geology-based zonation models in predicting drawdowns, in general, are poorer and vary from one port to another. Upon incorporating geological information as initial guesses into geostatistical models, Fig. 3.16 illustrates that the estimated K and S_s tomograms perform similarly among each other, and the predictions are excellent for most ports.

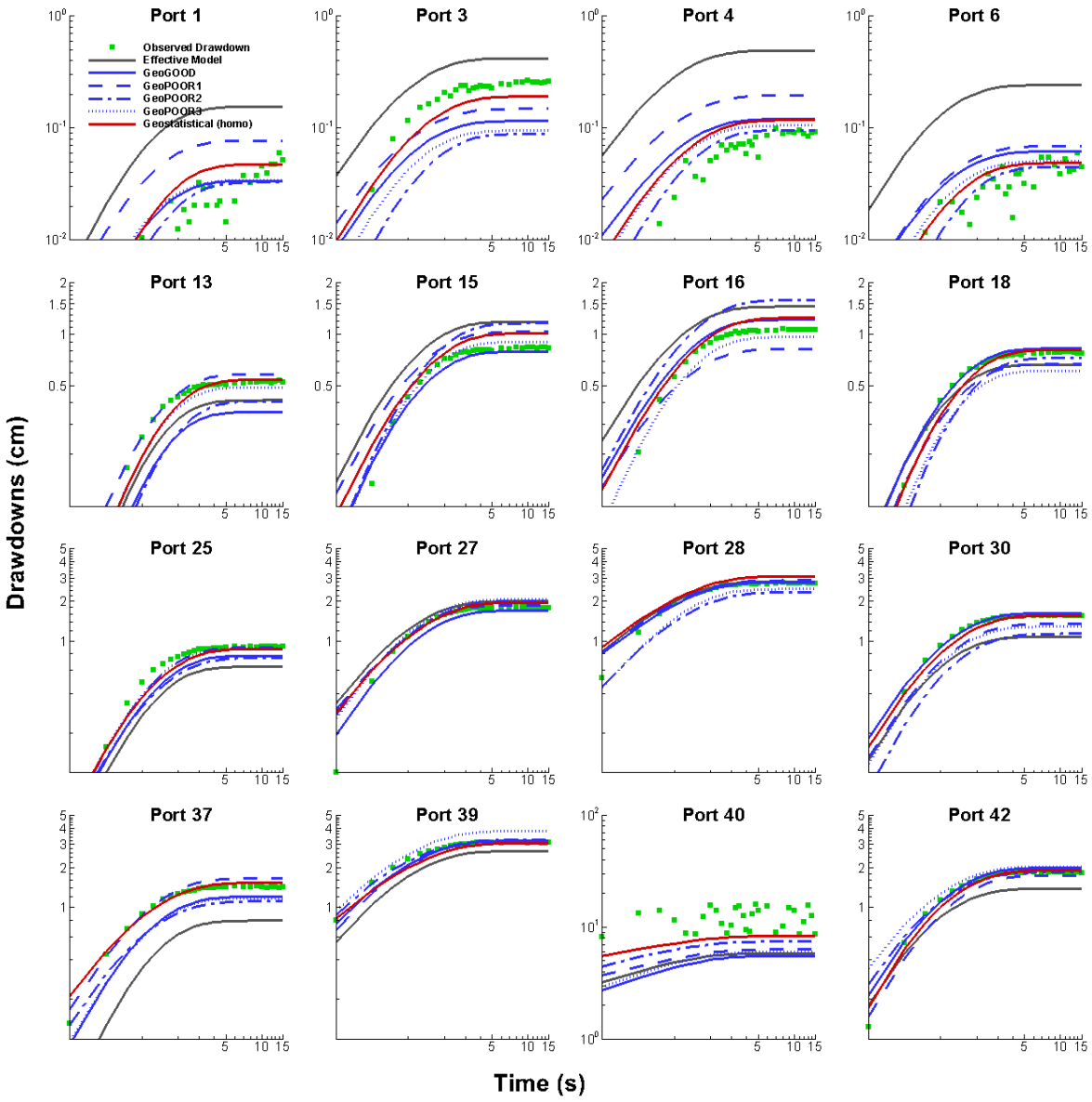

Fig. 3.15: Prediction of drawdown curves at 16 selected ports when conducting pumping test at port 40. K and S_s tomograms are obtained from different modeling approaches with eight pumping tests and 47 observation ports (Case 1).

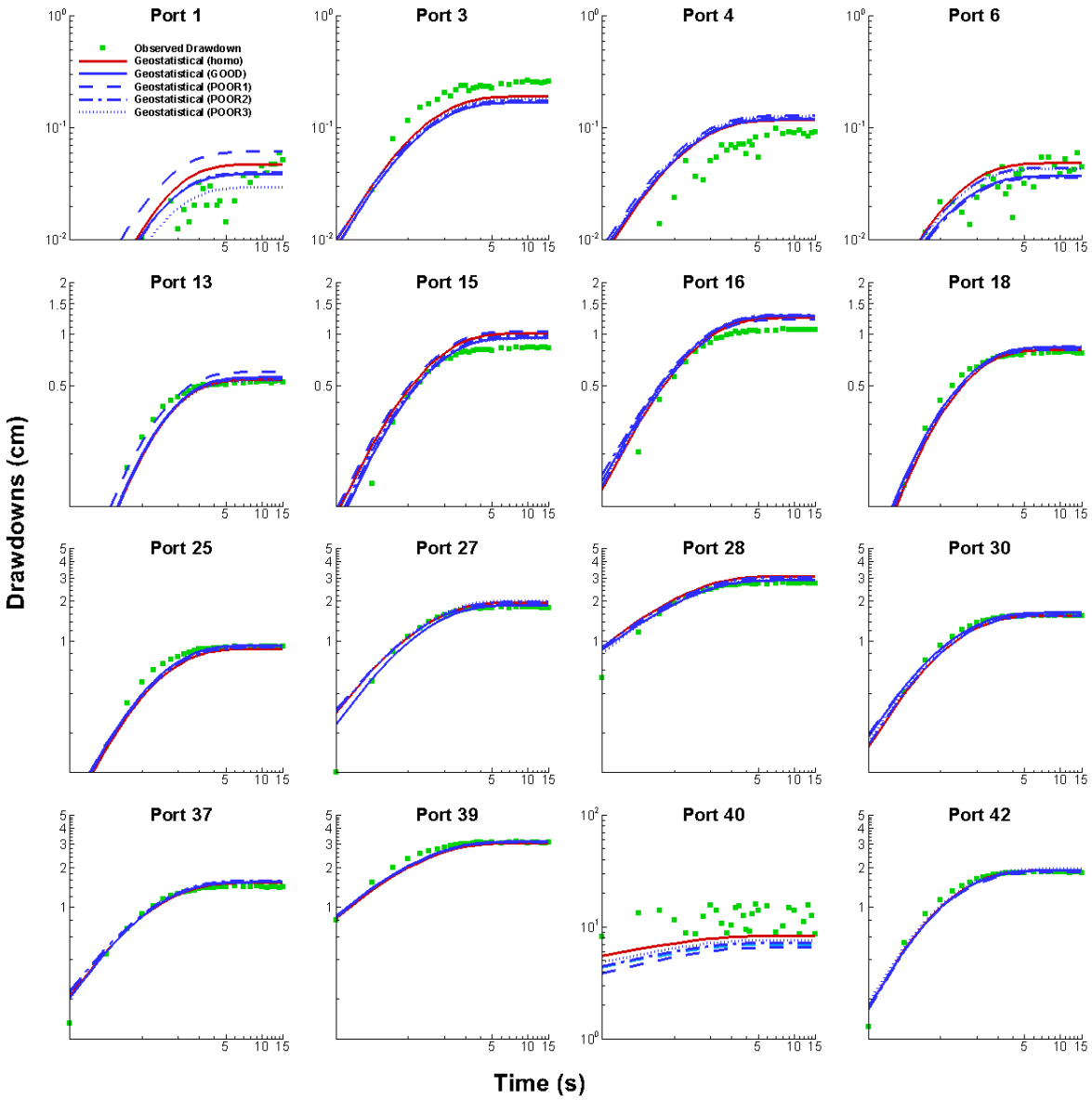

Fig. 3.16: Prediction of drawdown curves at 16 selected ports when conducting pumping test at port 40. K and S_s tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from eight pumping tests and 47 observation ports (Case 1).

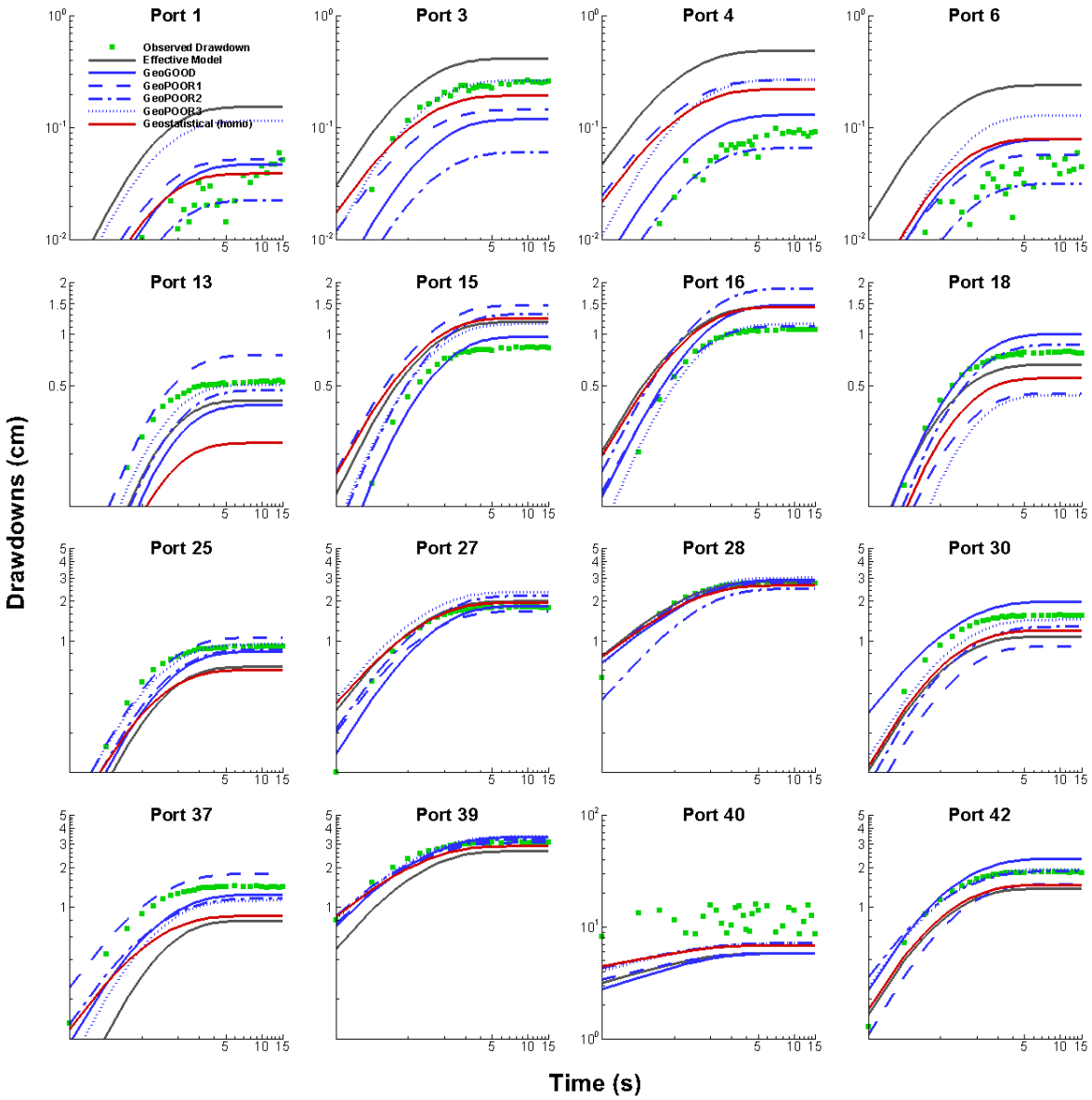

Fig. 3.17: Prediction of drawdown curves at 16 selected ports when conducting pumping test at port 40. K and S_s tomograms are obtained from different modeling approaches with eight pumping tests and 47 observation ports (Case 2).

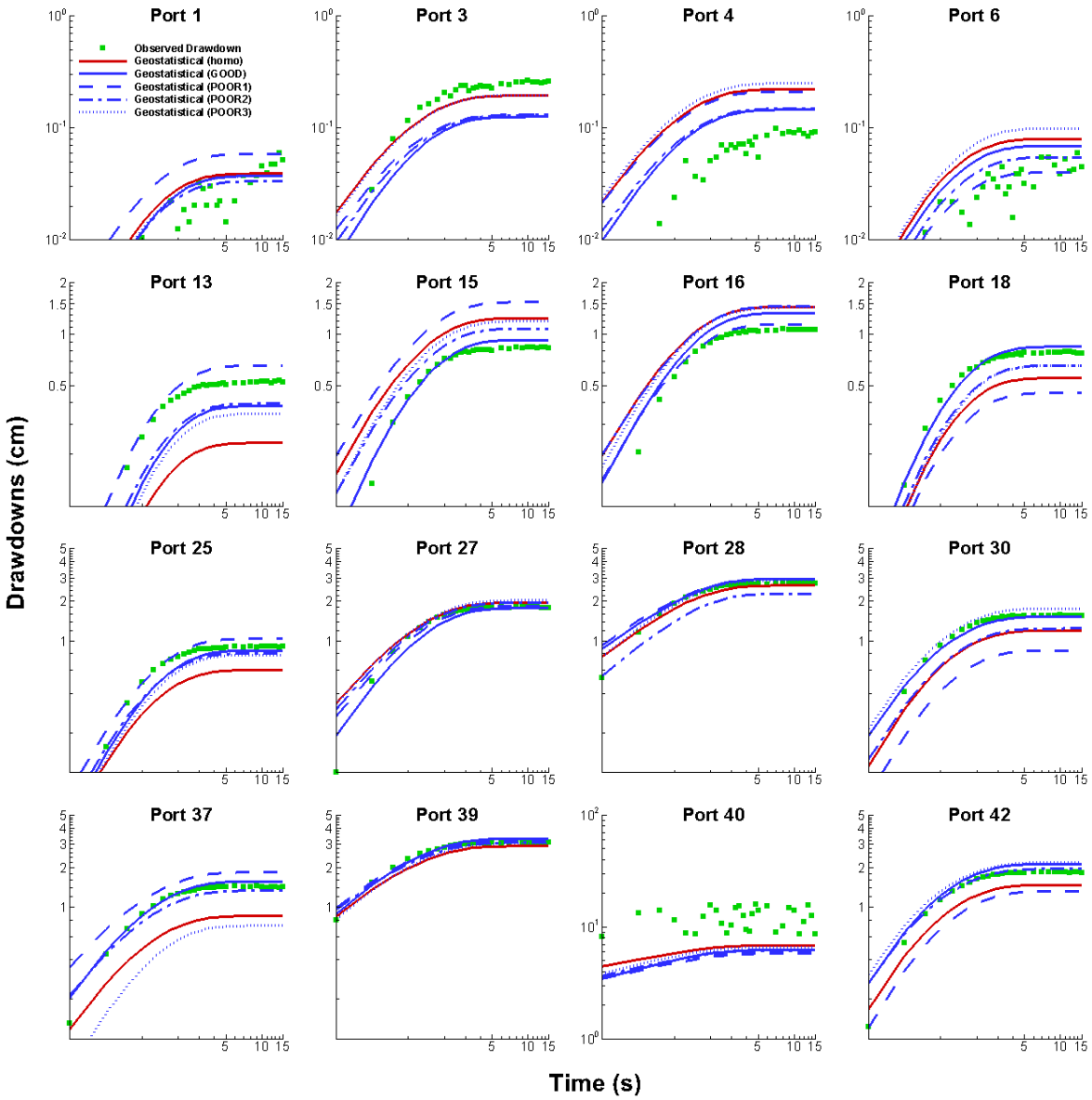

Fig. 3.18: Prediction of drawdown curves at 16 selected ports when conducting pumping test at port 40. K and S_s tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from eight pumping tests and 47 observation ports (Case 1).

When both the number of pumping tests and observation ports were reduced for inverse modeling, Fig. 3.17 shows that the geostatistical model with homogeneous initial K and S_s fields fails to capture drawdowns at some ports, particularly for the ports located at the fringes of the aquifer (e.g., ports 13, 18, 25, 30, 37, and 42). The main reason for this is that observation data at these ports are removed, and the estimated K and S_s tomograms fail to reveal the details

of heterogeneity in these areas, as shown in Fig. 3.7a and Fig. 3.7c. In contrast, by providing accurate stratifications, the K and S_s tomograms associated with geostatistical model provides better predictions of drawdown curves at these ports (Fig. 3.18). This result again suggests that when the number of head data is limited for aquifer characterization, a good knowledge of stratification is quite important, and it can be incorporated into geostatistical models to reveal more details in heterogeneity and provide more accurate prediction results. Overall, integration of geological information into hydraulic tomography is a good practice and should result in better results when accurate geological data are available.

3.4 On the Value of Transient Analysis of Hydraulic Tomography Data

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

To investigate this issue, two additional cases were performed by conducting SSHT, using

the same pumping and observation densities (Cases 1 and 2) utilized for THT analyses. Results from the SSHT were then compared to those from THT presented earlier. In particular, K tomograms from SSHT with a geometric mean of 48 S_s values ($S_s = 6.1 \times 10^{-4}$ /cm) from single-hole tests were used to conduct forward simulations of 16 independent pumping tests. Simulation results (Fig. 3.19) reveal that the drawdown predictions are significantly biased for both cases, suggesting that in order to achieve accurate predictions of transient drawdowns, transient inversions are necessary.

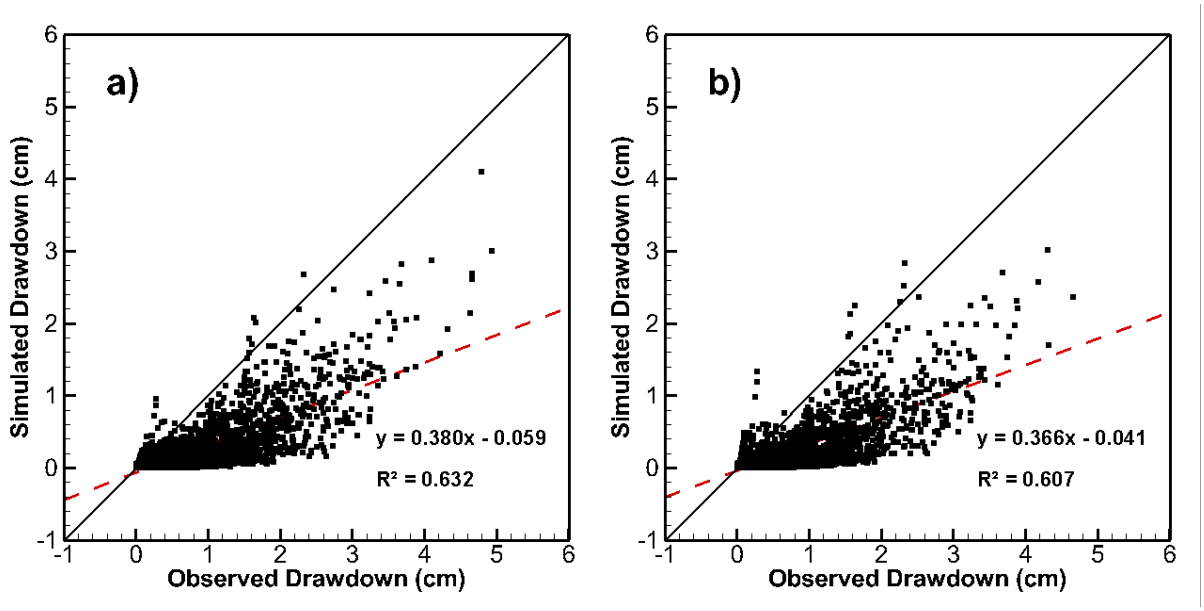

Fig. 3.19: Validation scatterplots of simulated versus observed transient drawdowns utilizing K tomograms obtained from SSHT coupled with the geometric mean of 48 S_s values (S_s = 6.1 x 10^{-4} /cm) obtained from single-hole tests. (a) Case 1, (b) Case 2.

4 Study II: Three-dimensional Hydraulic Tomography Analysis of Long-term Municipal Wellfield Operations: Validation with Synthetic Flow and Solute Transport

4.1 Experimental Setup

The Mannheim East wellfield is located within the core area of the Waterloo Moraine, which is classified as a kame deposit with three main aquifers separated by two glacial tills (Karrow, 1993). To mimic the multi-aquifer/aquitard system of the study site, a layer-cake geological model was constructed for this synthetic study, as shown in Fig. 4.1. The dimensions of the model are 5000 m, 5000 m, and 200 m in *x*-, *y*-, and *z*-directions, respectively. In total, seven geological layers were identified beneath the study site with AT and AF representing aquitard and aquifer, respectively. These layers were identified following the conceptual hydrogeological model of the Waterloo Moraine constructed by Bajc and Shirota (2007), whereas some layers with thin thicknesses were merged and irregular layer boundaries were not considered. In each geological layer, random K and S_s fields were generated by assuming Gaussian distributions of $\ln K$ and $\ln S_s$ fields with known information of their means, variances, and correlation lengths using the spectral approach (Robin et al., 1993). The mean values of $\ln K$ and $\ln S_s$ were obtained based on the predominant materials in each geological layer, while the variances and correlation lengths were estimated according to the statistical properties of spatial K and S_s distributions in natural geological formations. The generated nonstationary "true" K and S_s fields for the synthetic study are illustrated as Fig. 4.3d and Fig. 4.4d,

respectively. Table C1 in Appendix C summarizes the statistical details for random fields generation. To better evaluate the results, the entire simulation domain was subdivided into three zones (ZONE 1, ZONE 2, and ZONE 3) based on the density of well screens.

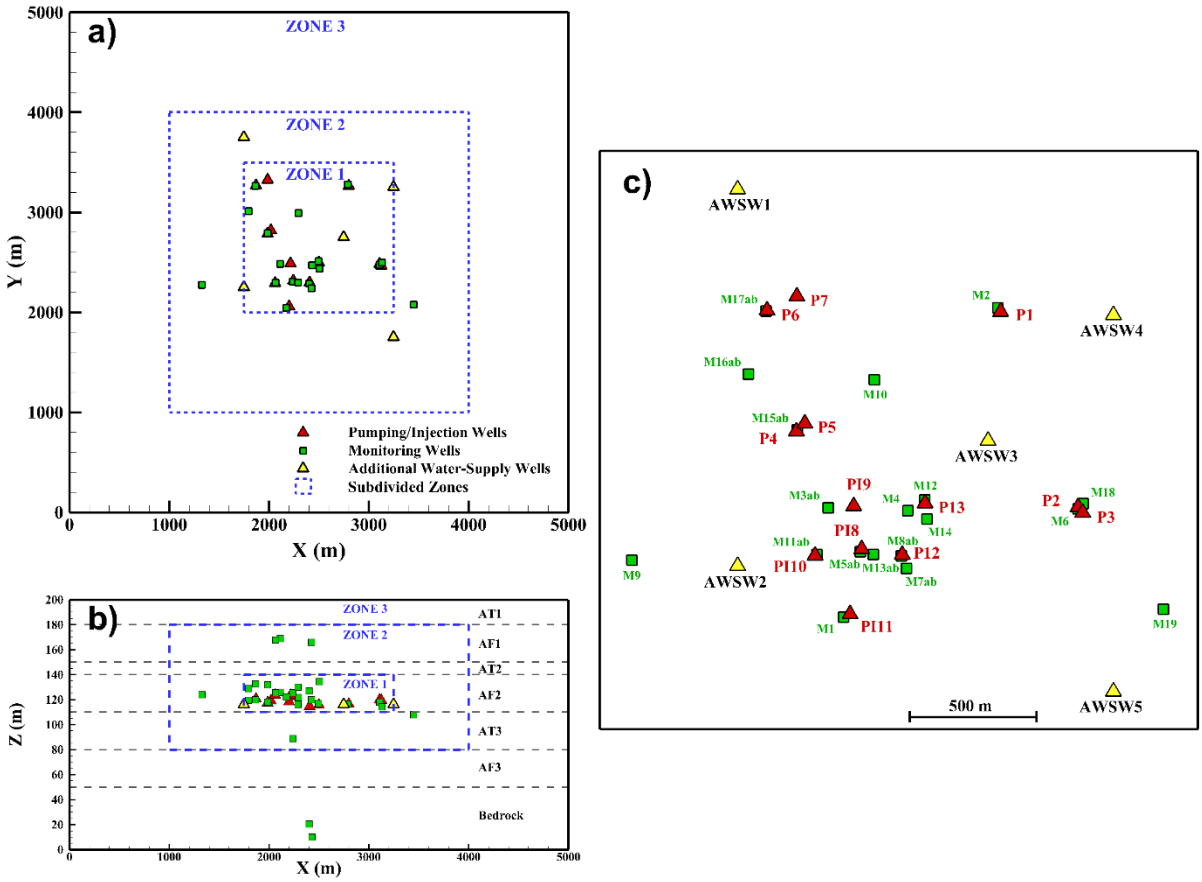

Fig. 4.1: The synthetic layer-cake geological model domain along with the distribution of pumping and monitoring wells. a) and b) illustrate the plan-view and cross-section of the simulation domain, respectively, c) shows spatial distribution of assigned wells with IDs.

To mimic the hydraulic condition within the municipal water-supply wellfield, the same well configuration as the Mannheim East wellfield was applied for the synthetic study. Within the wellfield, a well site with Aquifer Storage and Recovery (ASR) system was designed to inject and store treated surface water during low water demand periods and extract the stored water during high demand periods. In total, 13 pumping/injection municipal wells screened in

the water-supply aquifer (AF2) and 28 water-level monitoring wells screened at different layers were included in the model. The spatial distribution of these wells and their screens are illustrated in Fig. 4.1. Other than the existing wells, five additional water-supply wells (AWSWs 1-5) were included for the purpose of model validation using independent pumping test data.

The synthetic model was discretized into 33,072 triangular prism elements with 18,050 nodes, as shown in Fig. C1. The mesh was refined around wells but becomes coarser when moving towards boundaries. The four lateral boundaries of the model were set as constant head boundaries of 340 m, while the top and bottom boundaries were set as no-flow. Transient groundwater flow was then considered for the generation of synthetic head data, and the governing equations are provided as Eqs. 1.2 and 1.3. In this study, the transient flow equation was solved using the forward simulation code HydroGeoSphere (HGS) (Aquanty, 2019) to generate synthetic head data for analyses. In particular, variable pumping/injection records in all 13 water-supply wells during the years of 2012 and 2013 (Fig. 4.2) were extracted from Water Resources Analysis System (WRAS+) (Regional Municipality of Waterloo, 2014) database and included in the forward model. Simulated head values were report at all 28 monitoring well locations during the year of 2013 (as shown in Fig. C2) for the analyses presented in this study.

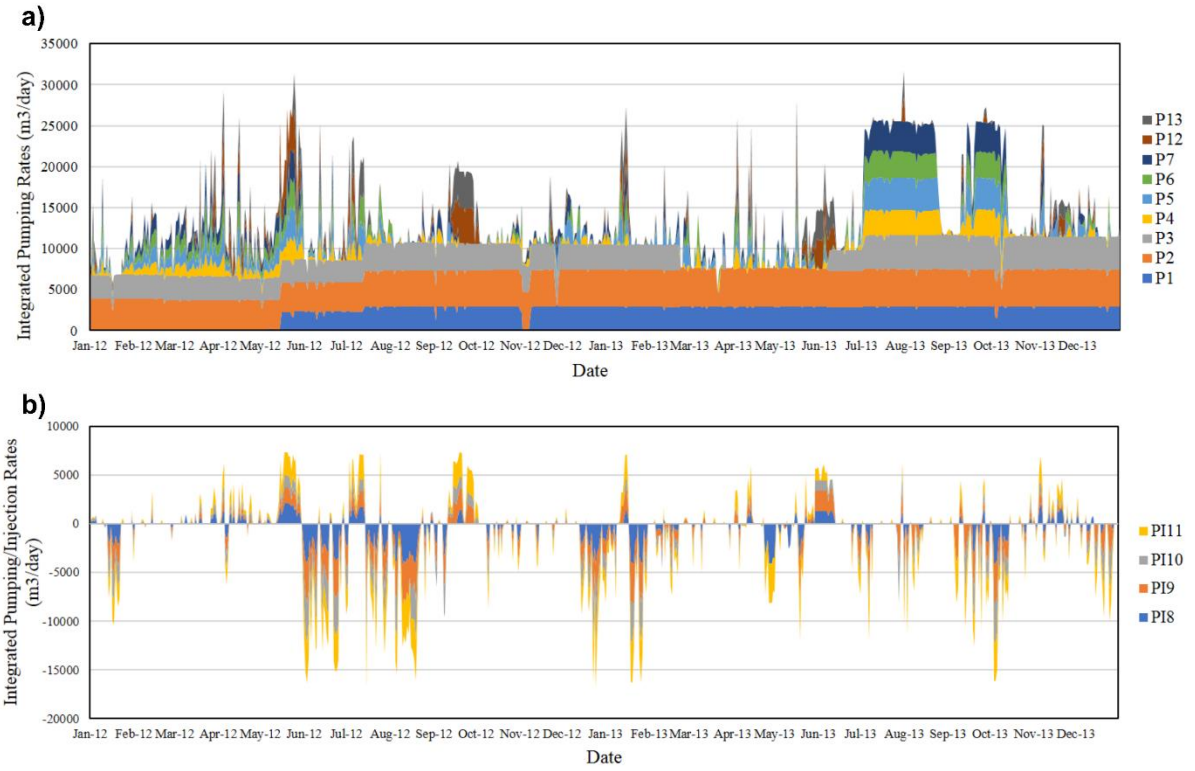

Fig. 4.2: Extraction (positive) and injection (negative) rate records at all 13 municipal wells during the years of 2012 and 2013 from the WRAS+ database.

The purpose of including pumping/injection information prior to the observation data is to mimic the pumping history of the system prior to the calibration period, which leads to uncertain initial conditions at the beginning of observation data. In addition to municipal well data, dedicated pumping test data from additional water-supply wells (AWSWs 1-5) were generated as independent pumping test data and utilized for model validation. In particular, a constant pumping rate of 8,000 m^3/day was assigned to each additional well, and drawdown data in all 28 monitoring wells were simulated (as shown in Fig. C3). The pumping rate was obtained based on the recorded daily pumped/injected volume of water at the site. The utilization of these independent pumping test data is to assess the ability of the obtained hydraulic parameter (K and S_s) fields in guiding the construction of new water-supply wells.

4.2 Data Utilized for Inverse Modeling

Instead of including all simulated head data (0.1-day interval) for analysis, daily observation data at the beginning of each day (12:00 am) were extracted and utilized for model calibration (0 – 120 days) and validation (180 – 365 days). To investigate the effect of data selection on inverse modeling, five datasets with different durations and periods were selected for model calibration in the synthetic study, as shown in Fig. C4, with their properties summarized in Table C2. All datasets satisfy the premise that long-term operation records are available at the site, while the scenario of lack of operation information (e.g., pumping/injection records) is not considered in this study. According to Yeh et al. (2009), the lack of necessary and sufficient information (e.g., poorly characterized stimuli sources and/or sparsely collected temporal and spatial response data) for large-scale heterogeneity characterization would render the inverse problem ill-posed, yielding hydraulic parameter estimates with greater uncertainty.

As mentioned previously, the interpretation of municipal well data still suffers an issue of uncertain initial conditions for groundwater modeling due to the continuous operation of municipal water-supply wells. The effect of uncertain initial condition on groundwater modeling has been investigated by Yu et al. (2019). Based on their results, the proposed spin-up method was adopted for the presented analyses to minimize the effect of uncertain initial conditions. In particular, pumping/injection rate records prior to the observation data were utilized for model spin-up and incorporated for model calibration. The model spin-up time was determined by incorporating different lengths of prior pumping/injection records for forward

simulations with known hydraulic parameter (K and S_s) fields. The simulated head variations at monitoring locations were then compared quantitatively to the observed ones. The comparison results, as shown in Fig. C5, reveal that the discrepancy between simulated and observed head data decreases significantly as the spin-up period increases and stabilizes in magnitude after incorporating pumping/injection records 180 days prior to the observation data. As a result, pumping/injection rate records for 180 days prior to the observation data were extracted and incorporated for model calibration in this synthetic study.

4.3 Groundwater Flow Modeling Approaches

4.3.1 Case 1: Effective Parameter Model

The synthetic multi-aquifer/aquitard system was first characterized as a homogeneous, isotropic medium to estimate the effective K and S_s values by coupling the groundwater flow model HGS (Aquanty, 2019) with the parameter estimation code PEST (Doherty, 2015), and is referred to as the 'effective parameter' model. The effective parameter model provides zero-resolution on subsurface heterogeneity; however, it may still be able to describe the overall behavior of groundwater flow in the system. Furthermore, the estimated effective K and S_s values could be used as the initial estimate of hydraulic parameters to guide the calibration of more sophisticated groundwater flow models. For each dataset, an optimal set of K and S_s was estimated by simultaneously matching all data points. The initial values of K and S_s input into PEST are 0.99 m/day and 1.88×10^{-4} /m, respectively, which are the calculated geometric means of the entire "true" K and S_s fields. In PEST, the lower and upper bounds were set as 1

$\times 10^{-8}$ and 1×10^{5} m/day for K, and 1×10^{-10} and 1×10^{1} /m for S_s.

4.3.2 Case 2: Geology-based Zonation Model

The response from the synthetic multi-aquifer/aquitard system was then used to calibrate the geology-based zonation model, which is normally adopted for groundwater flow modeling at large scales (e.g., regional or basin scales). In this approach, each geological layer was characterized as a homogeneous, isotropic medium, and a uniform set of K and S_s was estimated and assigned to describe its hydraulic properties. The effect of the accuracy of constructed geological models on inverse modeling has been previously investigated through sandbox experiments (Chapter 3). To avoid the uncertainty associated with model identification, it is assumed that the hydrostratigraphic contacts are perfectly known for the geological model.

In a similar fashion to the effective model, the geological model was calibrated using PEST coupled with HGS by simultaneously matching all data points. For each geological layer, the geometric means of K and S_s from the "true" fields were utilized as initial guesses of hydraulic parameters for model calibration. Additional cases were conducted by using the calibrated effective model as initial K and S_s guesses for geological model calibration; however, unrealistic values of hydraulic parameters were obtained in layers where no hydraulic head data was available, a phenomenon noted in previous studies (e.g., Berg and Illman, 2011b; Luo et al., 2017; Zhao et al., 2016). As a result, the heterogeneous K and S_s fields based on the stratigraphic information were applied as initial guesses, with the lower and upper bounds set the same as those for effective parameter models. In total, seven sets of K and S_s were estimated for subsurface heterogeneity characterization using the geology-based zonation model.

4.3.3 Case 3: Geostatistical Models

As a third case, the head response to the synthetic municipal well data were interpreted with highly parameterized geostatistical models for subsurface heterogeneity characterization. All geostatistical inversions were conducted using the Simultaneous Successive Linear Estimator (SimSLE), developed by Xiang et al. (2009) and modified for this study to account for variable pumping/injection records with a reduced covariance matrix storage technique to save on memory requirements to handle the large number of nodes and data for inversion.

Based on the differences in initial K and S_s fields, two geostatistical inversion cases were investigated. For Case 3a, homogeneous initial K and S_s fields were used for model calibration, representing the scenario of calibrating hydraulic data only. In this case, the K and S_s values obtained from the effective parameter model (Case 1) were utilized as initial guesses and assigned to the entire domain. For Case 3b, geological information was incorporated for model calibration. Geostatistical inversions in this case started from the heterogeneous initial K and S_s fields same as those utilized for geological model calibration. For both cases, the variances of $\ln K$ and $\ln S_s$ ($\sigma^2_{\ln K}$, $\sigma^2_{\ln Ss}$) were initially set to be 4.0 and 2.0, respectively, while the correlation scales were set to be λ_x = 400 m, λ_y = 400 m, and λ_z = 5 m for both K and S_s. Due to the fact that the geostatistical parameters of heterogeneous K and S_s fields are commonly unknown for field studies, the input parameters for geostatistical inversions are different from those utilized for "true" K and S_s fields generation. On the other hand, the values of these initial statistical properties have found to have negligible effects on the results when sufficient head data are available for geostatistical inversions using SimSLE (Yeh and Liu, 2000). This is

because the error covariance matrices of ln$K$ and lnS_s are updated during each iteration by gradually assimilating information from observation data (Zha et al., 2017).

4.4 Results and Discussion

In this study, five datasets (Datasets A-E) were interpreted with four different models (Cases 1, 2, 3a and 3b). Results from all investigated models were summarized and examined. In particular, K and S_s values estimated from different models were first compared to the "true" fields to illustrate the accuracy of these estimates. Then, calibration and validation results were assessed qualitatively and quantitatively by plotting scatterplots of simulated versus observed head variations and evaluating model errors, respectively. The evaluation of model errors was performed by computing the mean absolute error (L_1), mean square error (L_2), and coefficient of determination (R^2) between simulated and observed head values using Eqs. 3.1 to 3.3 in Section 3.3.5.

In the following sections, calibration and validation results associated with Dataset A are first presented to evaluate the performance of different models in revealing large-scale heterogeneities and predicting groundwater flow, while the summarized results associated with other datasets (Datasets B-E) are provided in Appendix C. Then, statistical summary (L_1, L_2, and R^2) of the validation results obtained from all investigated models is examined to give insight into the effect of data selection on inverse modeling.

4.4.1 *Model Calibration*

Through the interpretation of Dataset A, the effective parameter model (Case 1) yields K and S_s estimates as well as their 95% confidence intervals of $K = 9.87 \pm 0.14$ m/day and $S_s = 2.56 \times 10^{-4} \pm 2.5 \times 10^{-5}$ /m. Compared to the calculated geometric means from the "true" fields (0.99 m/day and 1.88×10^{-4} /m for K and S_s, respectively), the effective K and S_s obtained from the municipal well data are more representative of the effective hydraulic parameters of the layers, where most monitoring wells are screened (AF1, AT2, and AF2). This implies that more observation ports in upper and lower geological layers are required to obtain unbiased effective hydraulic parameters for the entire multi-aquifer/aquitard system.

Fig. 4.3 illustrates the obtained K tomograms from the geology-based zonation (Case 2) and geostatistical models (Cases 3a and 3b) through the interpretation of Dataset A. The "true" K field is included on the bottom right as a reference for comparison. Fig. 4.4 illustrates the same, but for the estimated S_s tomograms.

As shown in Fig. 4.3a and Fig. 4.4a, the estimated K and S_s values from the zonation model (Case 2) are found to roughly describe the average hydraulic properties of each geological layer in comparison to the "true" field (Fig. 4.3d and Fig. 4.4d for K and S_s, respectively). Here, it should be noted that geological models in this synthetic study were calibrated with known stratigraphic information and well estimated initial K and S_s values (the geometric means of random K and S_s fields in each geological layer). However, typically there is great uncertainty in the geological model constructed with sparse boreholes and utilized as prior information.

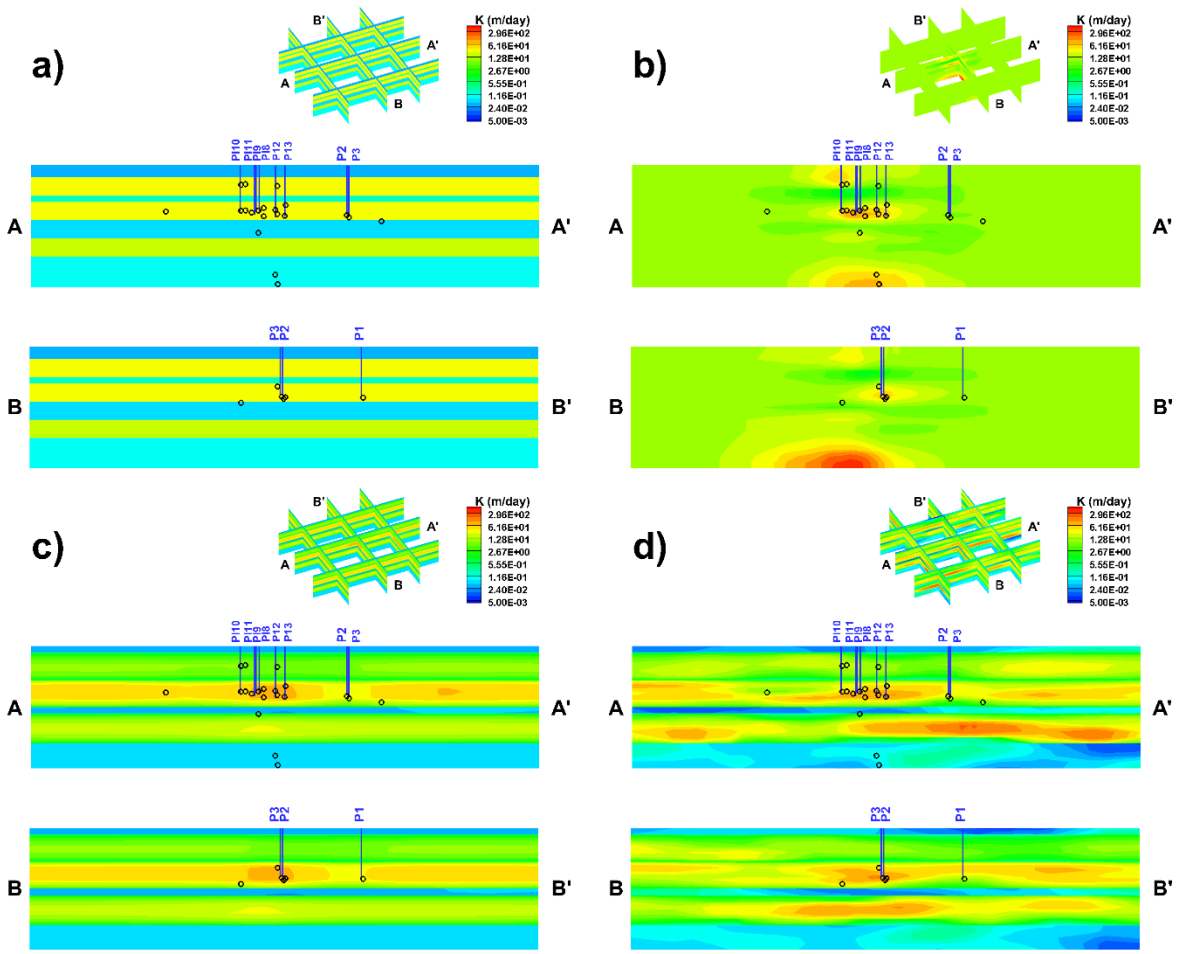

Fig. 4.3: Estimated *K* tomograms from three model cases through the interpretation of Dataset A as well as the "true" *K* field. a) Case 2, b) Case 3a, c) Case 3d, d) "true" *K* field. In each contour map, small black circles represent the location of monitoring well screens.

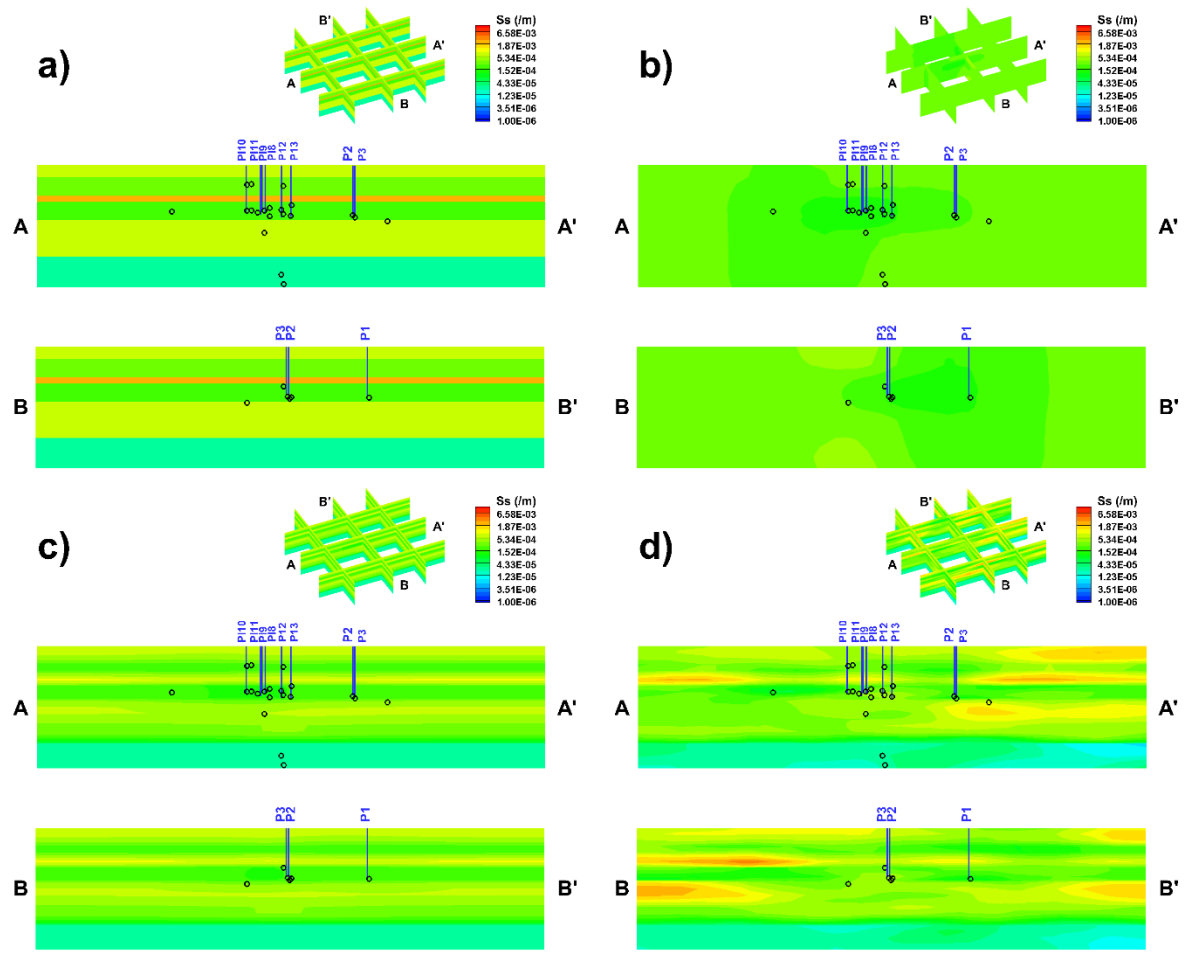

Fig. 4.4: Estimated S_s tomograms from three model cases through the interpretation of Dataset A as well as the "true" K field. a) Case 2, b) Case 3a, c) Case 3d, d) "true" S_s field. In each contour map, small black circles represent the location of monitoring well screens.

To further evaluate these estimates, the natural logarithm of these K and S_s estimates as well as their 95% confidence intervals are plotted as Fig. C14 in Appendix C. Results reveal that the narrowest confidence intervals of K estimates are obtained in the water-supply aquifer (AF2), where all pumping/injection and monitoring wells are screened, suggesting the high confidence of the K estimate of this layer. In contrast, when we examine the upper and lower layers, the confidence of K estimates decreases resulting in larger confidence intervals due to fact that fewer observation data are available in these layers for estimating reliable K values.

This is in line with the conclusion derived from the study described in Chapter 3 that when using a zonation model for subsurface characterization, hydraulic head data in each identified zones are required to yield reliable estimates of hydraulic parameters of these zones.

In comparison to K estimates, the estimated S_s values are found to have larger confidence intervals, suggesting higher uncertainty associated with these S_s estimates. This may be attributed to the fact that daily observation data were extracted and interpreted for hydraulic parameters estimation in this study. The utilization of such data points ignores early-time water-level variations right after the change of pumping/injection rates which are of critical importance for obtaining reliable S_s estimates (Sun et al., 2013). The interpretation of datasets in a denser fashion (e.g., hourly observation points) may improve the estimation of S_s. However, due to the computationally intensive nature of geostatistical inversions and the limitation of current computational resources, such a scenario of including a dense dataset is not included in this study.

The obtained K and S_s tomograms from the geostatistical models are illustrated as Fig. 4.3b-c and Fig. 4.4b-c, respectively. As shown in Fig. 4.3b, the geostatistical inversion of hydraulic head data only (Case 3a) is able to reveal heterogeneity details, where wells are concentrated with sufficient head data. However, the estimated K tomogram results in great loss of heterogeneity details in comparison to the "true" K field. Although some major zones are delineated, the overall smooth patterns fail to capture the precise shapes of stratigraphic features. Different from the K tomogram, the S_s tomogram estimated from Case 3a does not

show any distinct heterogeneity details, as shown in Fig. 4.4b. This result again implies that the selected head data for model calibration are restrictive for S_s estimations.

After incorporating the geological information for geostatistical inversion (Case 3b), significant improvement in revealing heterogeneity details is observed for both K and S_s tomograms. In particular, greater details in K heterogeneity are revealed within the water-supply aquifer (AF2), resulting in the spatial distribution of K in this layer to be comparable to that of the "true" field. For upper and lower layers, the loss of heterogeneity details is still observed due to the lack of hydraulic information in these layers. We believe that the estimated K tomogram can be further enhanced if more monitoring wells are available for head response records at different layers. The improvement in the S_s tomogram is not as distinct as that in the K tomogram; however, slight patterns of S_s heterogeneities are more revealed after the incorporation of geological information (as shown in Fig. 4.4c).

The estimated K and S_s values from geostatistical models (Case 3a and 3b) were then evaluated by analyzing the uncertainty associated with these estimates and comparing them to the "true" values. For uncertainty analysis, the corresponding ln$K$ and lnS_s variance maps are plotted (shown as Fig. C19), with larger variances indicating higher uncertainty of the estimates. For both cases, relatively small ln$K$ variances are obtained in the central area of the simulation domain, where wells are concentrated for hydraulic head data, while variances become larger when moving away from the wells. In general, the lnS_s variances are computed to be larger than those of ln$K$, suggesting the higher uncertainty of these S_s estimates in comparison to the

K estimates. This may again be attributed to the temporal resolution of observation data for model calibration, as discussed above.

The estimated K and S_s values were then compared to the "true" values by plotting the scatterplots of corresponding estimated versus "true" ln$K$ and lnS_s values (shown as Fig. C24). This comparison reveals that the geostatistical inversion of hydraulic head only (Case 3a) is still able to yield relatively reliable K and S_s estimates in the area with sufficient hydraulic head data (ZONE 1), while large discrepancies of these estimates are observed in ZONEs 2 and 3. After incorporating the stratigraphic information, significant improvements are observed for both K and S_s estimates in all three zones. These results indicate that the municipal well data can be used to characterize subsurface heterogeneity with HT. However, since such hydraulic data are typically concentrated in the pumping area, accurate stratigraphy information is of critical importance for geostatistical inversions to accurately reveal heterogeneity patterns and yield reliable estimates of hydraulic parameters. Earlier studies by Zhao et al. (2016) and Luo et al. (2017) have shown that the inclusion of inaccurate stratigraphy information will have deleterious impacts on parameter estimates.

The performance of four different models is further assessed qualitatively and quantitatively by plotting the scatterplots of calibration results, as shown in Fig. 4.5. In each scatterplot, data points corresponding to three subdivided zones are distinguished with different colors. A linear model that fits all data points is provided along with the corresponding coefficient of determination (R^2), as well as calculated L_1 and L_2 norms. Examination of Fig. 4.5 reveals that the calibration results improve when a larger number of estimated parameters

are accounted for in inverse modeling (from Cases 1 to 3). This makes sense since the highly parameterized geostatistical model allows for the adjustment of K and S_s estimates in each element to fit the data. After incorporating prior geological information, the geostatistical model (Case 3b) yields the best fit of simulated and observed head variations (Fig. 4.5d).

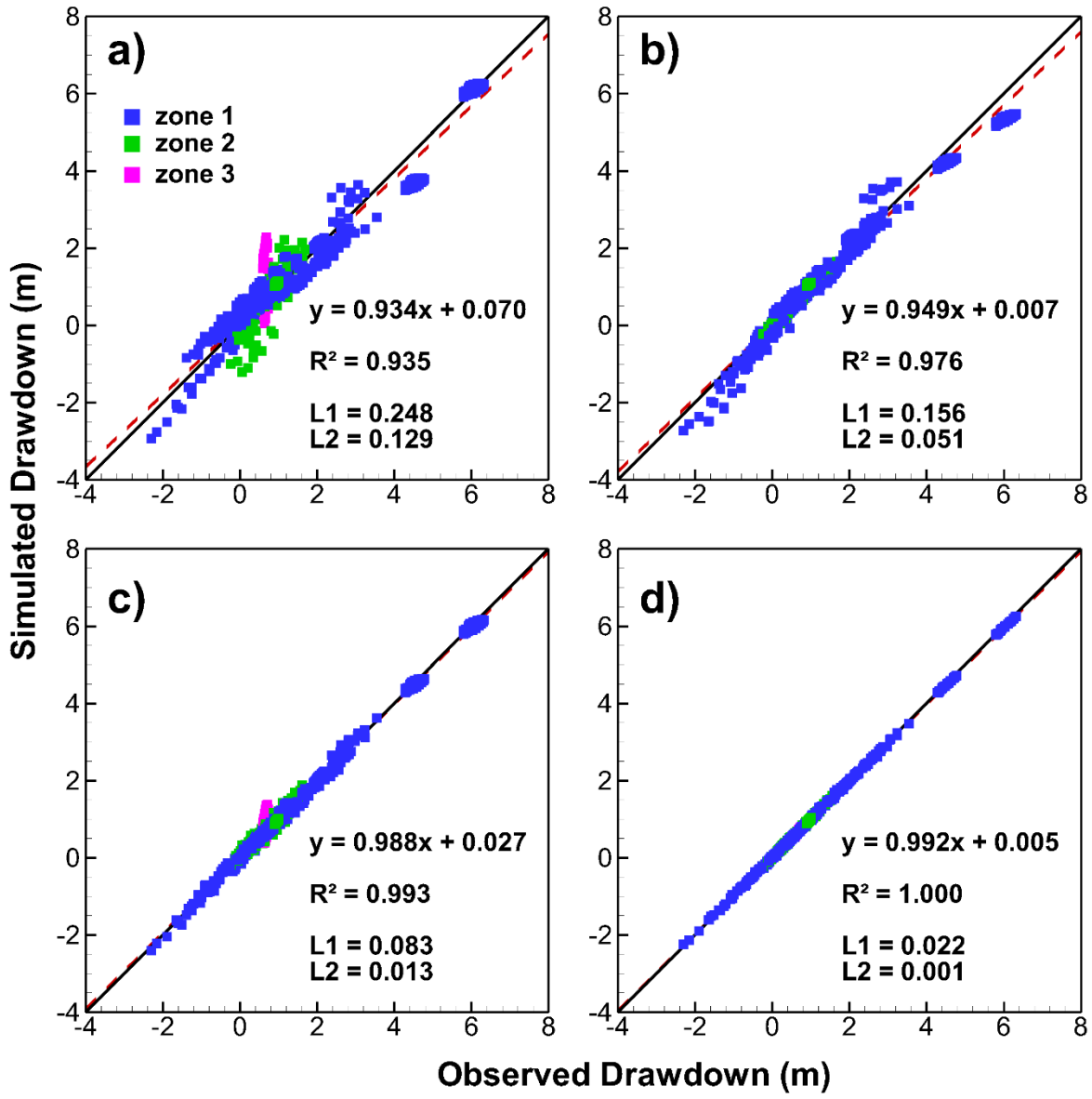

Fig. 4.5: Calibration scatterplots (Dataset A) of simulated versus "observed" drawdowns for four model cases. a) Case 1, b) Case 2, c) Case 3a, and d) Case 3b.

4.4.2 Model Validation

Model validation in this study was performed in two scenarios. For Scenario 1, the municipal well data during the second half year of 2013 were utilized for model validation. Specifically, the obtained K and S_s tomograms were applied to continuously predict head variations using the same well configuration (water-supply and monitoring wells) as that for model calibration. For Scenario 2, the independent pumping test data obtained from additional water-supply wells (AWSWs 1-5) not used in the calibration effort were utilized for model validation. The validation scatterplots of different model cases associated with Dataset A are illustrated in Fig. 4.6 and Fig. 4.7, for Scenarios 1 and 2, respectively.

Examination of Fig. 4.6 reveals that when the municipal well data are utilized for model validation (Scenario 1), the performances of different model cases share the same order as the calibration results. In particular, Case 3d performs the best in continuously predicting drawdown variations for the entire domain, followed by Case 3a, while Case 1 yields the worst prediction results in terms of bias and scatter. Case 2 is found able to adequately predict drawdown variations at monitoring locations in all subdivided zones; however, the lack information of intralayer heterogeneity resulted in relatively large scatter between the simulated and observed head variations. It is of interest to note the K and S_s tomograms obtained from the geostatistical model without geological information (Case 3a) show great loss of heterogeneity details (Fig. 4.3b and Fig. 4.4b for K and S_s, respectively); however, they could still be applied to yield adequate predictions of head variations in all monitoring wells. This is because the data utilized for validation in Scenario 1 share the information of well

configuration as the datasets utilized for model calibration, thus making these validation results biased for the assessment of the performance of different models.

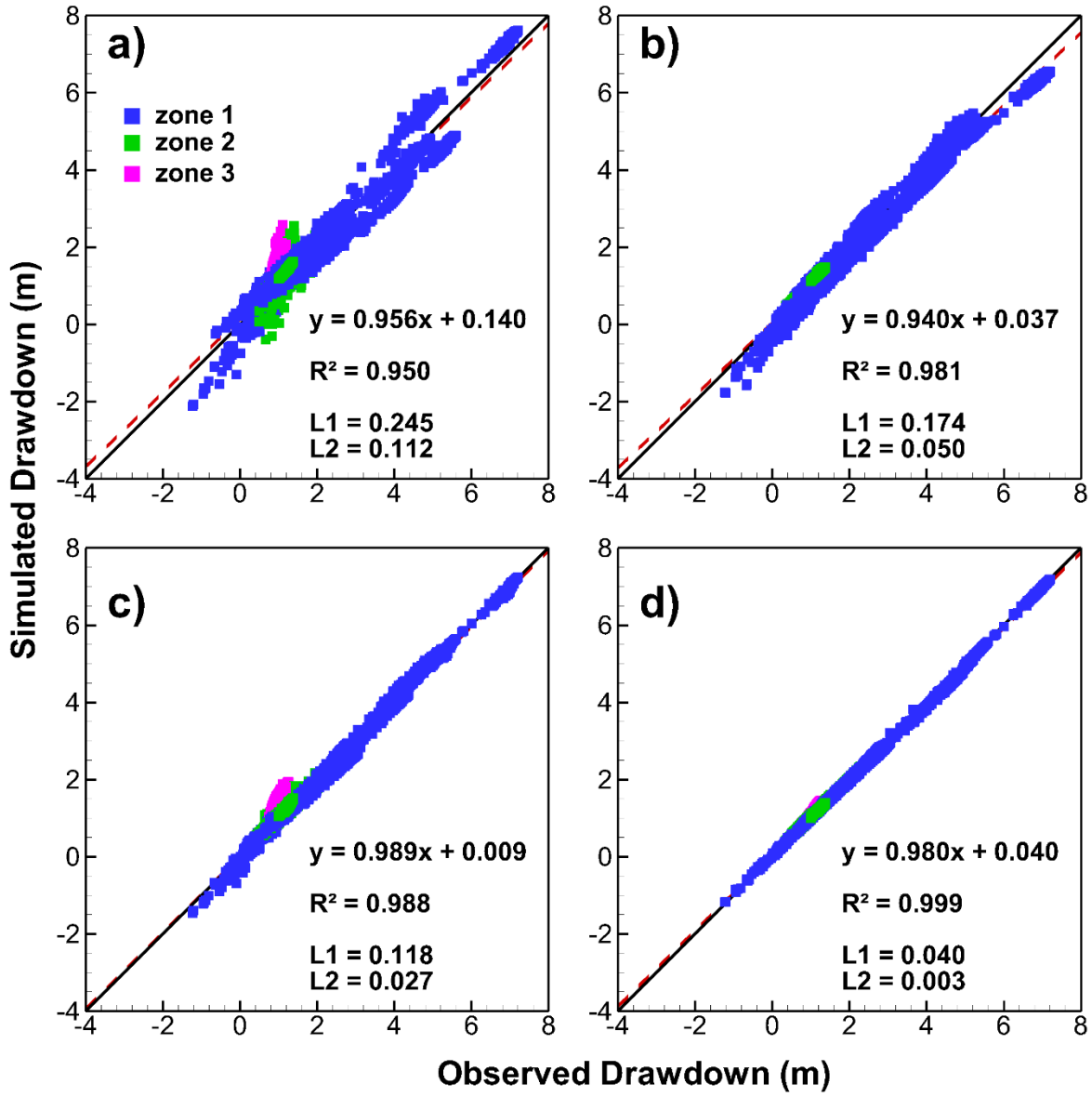

Fig. 4.6: Validation scatterplots (Dataset A) of simulated versus "observed" municipal well data (Scenario 1) for four model cases. a) Case 1, b) Case 2, c) Case 3a, and d) Case 3b.

To ensure a more credible validation of the different models, independent pumping test data that not used in the calibration effort should be utilized for model validation (Illman et al., 2010), which is illustrated as Scenario 2 in this study. Fig. 4.7 shows the validation scatterplots.

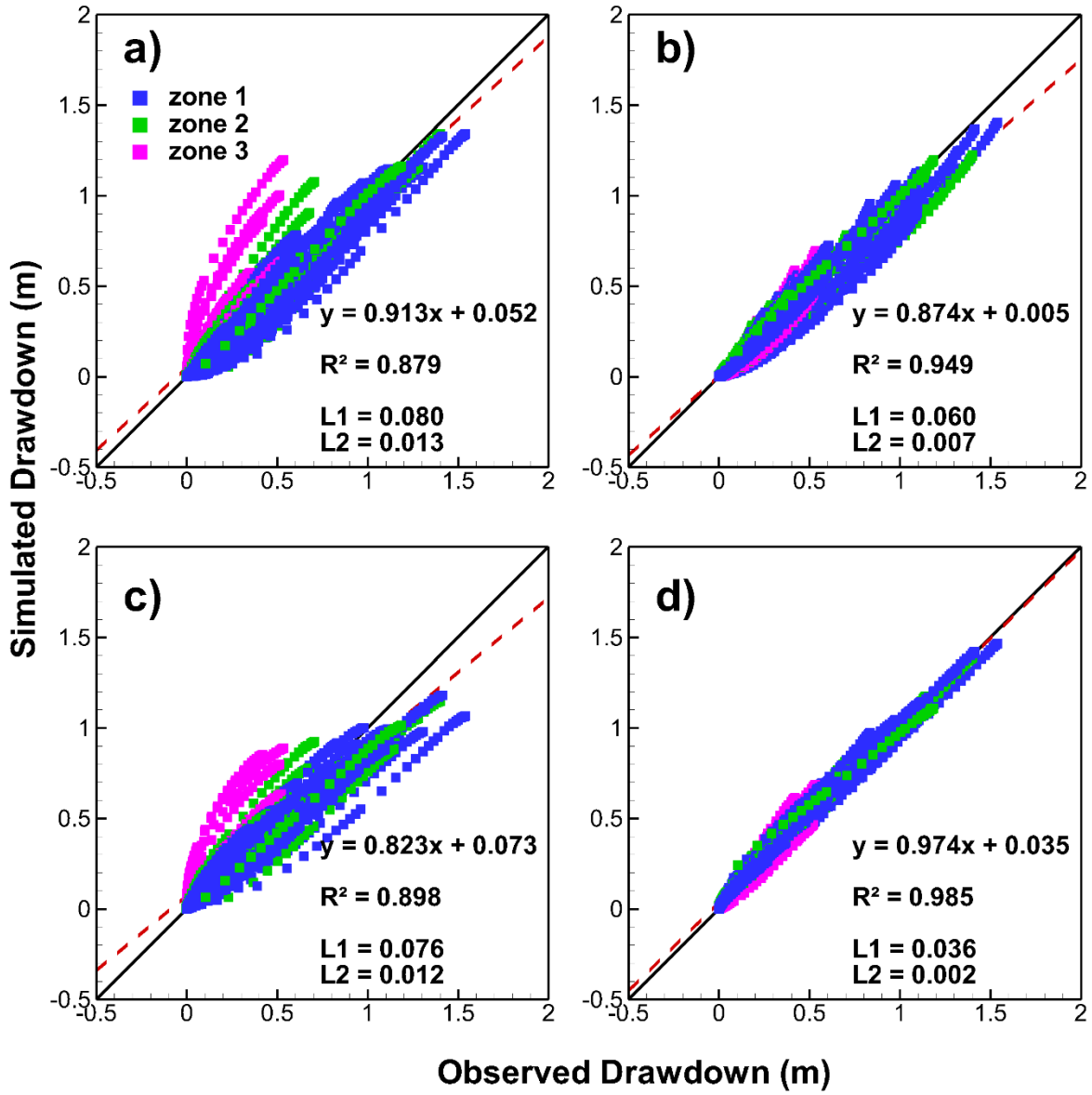

Fig. 4.7: Validation scatterplots (Dataset A) of simulated versus "observed" independent pumping test data (Scenario 2) for four model cases. a) Case 1, b) Case 2, c) Case 3a, and d) Case 3b.

Case 1 still performs the worst in predicting drawdowns from the independent pumping tests. However, it is surprising to find that Case 2 provides much better prediction results in

comparison to Case 3a, especially for monitoring wells screened in ZONEs 2 and 3. The comparison result reveals that the stratigraphic information becomes increasingly important for subsurface heterogeneity characterization when fewer hydraulic head data are available for inverse modeling, which is again in line with the conclusion provided in Chapter 3. After incorporating geological information, Case 3b yields the best prediction results with the highest correlation and smallest discrepancy between simulated and observed drawdowns in comparison to other model cases. The validation results associated with Case 3b reveal that the K and S_s tomograms obtained from the geostatistical model with geological information cannot only be used to predict water-level variations in the existing municipal wells, but also guide the construction of new water-supply wells at locations where the variances of the estimated ln$K$ and lnS_s are small.

The calibration and validation results presented above reveal that stratigraphic information is of critical importance for large-scale site characterization using the municipal well data. The calibrated geological model yields relatively adequate predictions of water-level variations induced by both the existing (Scenario 1) and additional (Scenario 2) water-supply wells, while remarkable improvements in prediction results are observed when accurate geological information was incorporated into geostatistical inversions. However, it should be noted that the stratigraphic information adopted here is extracted from the synthetic model with no error. Following the conclusion provided by Zhao et al. (2016) and the study completed in Chapter 3, close attention should be paid in constructing accurate geological models when utilizing actual municipal well data for site characterization.

4.4.3 *Effect of Data Selection on Inverse Modeling*

To investigate the effect of data selection on inverse modeling, the statistical properties (L_1, L_2, and R^2) of the validation results from all investigated models were computed and plotted in Fig. C41 with all values summarized in Table C3. In general, when different datasets were included for model calibration, the effective parameter model (Case 1) always performed the worst in predicting groundwater flow, while the geostatistical model with geological information (Case 3b) always performed the best. On the other hand, the performance of the geological model (Case 2) and the geostatistical model without geological information (Case 3a) varied from one dataset to another.

When more observation points with longer simulation durations were included for model calibration (from Dataset A to Dataset C), the estimated K and S_s tomograms from Case 2 showed distinct improvement in continuously predicting municipal well data (Scenario 1, as shown in Fig. C41a) in terms of computed L_1, L_2, and R^2 values. Such improvement is not observed for the prediction of independent pumping test data (Scenario 2, as shown in Fig. C41b); however, slightly better prediction results were still obtained when using Dataset C for the geological model calibration. It is interesting to find that Case 3a behaves oppositely to Case 2, in which, worse validation results were obtained for Case 3a after increasing the simulation duration for model calibration. This may be attributed to the fact that with longer simulation durations, pumping/injection influence from the water-supply wells propagates to an area beyond the production area, resulting observation data in monitoring wells affected by the heterogeneity of K and S_s in a greater area without any hydraulic information. When

interpreting municipal well data with long simulation durations, the calibration of geostatistical models using hydraulic head only (Case 3a) is likely solving ill-posed inverse problems, yielding inaccurate estimation of hydraulic parameters.

Dataset D is selected to have the same simulation as Dataset A, but with much smaller magnitude of head variations. Results reveal that the validation results associated with Dataset D were distinctly worse in comparison to those associated with Dataset A for all model cases, implying that the periods with large water-level variations should be included when interpreting the municipal well data for site heterogeneity characterization.

Dataset E shares the same simulation duration as Dataset C, but only the periods with large water-level variations are utilized for model calibration. In comparison to Dataset C, Dataset E yielded slightly worse validation results for all model cases. This may be the case because the analysis presented in this study aims to estimate hydraulic parameters using long-term pumping/injection and water-level records. Instead of using the periods with large head variations only, continuous data points should be included to accurately describe the overall trends of water-level variations in monitoring wells.

Overall, the effects of data selection on inverse modeling are different for different modeling approaches. Through the comparison of validation results from all cases, the geostatistical model with geological information (Case 3b) is found to best interpret continuous records with large head variations for heterogeneity characterization. However, new approaches are needed for big data synthesis and intelligent data selection for inverse modeling.

4.5 Solute Transport Prediction

One remaining question is whether the estimated K and S_s tomograms from the municipal well data can be applied to predict solute transport. This would be another and a stronger form of model validation of HT results with a completely different data set not used in model calibration. To investigate this issue, additional model runs were performed by simulating solute transport using the estimated K and S_s tomograms. Results were then compared to the scenario simulated using the "true" K and S_s fields to evaluate the performances of these K and S_s estimates in predicting solute transport. For this investigation, the estimated K and S_s tomograms from four model cases through the interpretation of Dataset A were utilized.

4.5.1 Solute Transport Simulation

To simulate solute transport, a point source of the conservative solute chloride (Cl) was added into the synthetic system, located in the central area of layer AF1 with coordinates of x, y, and z equal to 2,750 m, 2,750 m, and 175 m, respectively. The source was assigned with a constant Cl concentration of 1,000 mg/L and removed after 50 years of simulation. The dispersivities of the system were assumed to be 20 m, 5 m, and 0.02 m for longitudinal, transverse, and vertical transverse directions, respectively. The porosity was assigned to be 0.4 throughout the simulation domain. Due to the relatively smooth distribution of S_s, the heterogeneity of porosity was not considered in this study for the simulation of solute transport. To mimic the migration of plume under real conditions, regional groundwater flow was accounted for in the solute transport simulation, in which groundwater was considered to flow

from the northwest to southeast with a hydraulic gradient of 0.0014. The influence of municipal water-supply pumping was also accounted for by assigning a constant pumping rate in each water-supply well based on its corresponding rate records. A slightly modified form of conventional advection-dispersion equation was adopted in this study for solute transport simulation, following the work of Burnett and Frind (1987). Specifically, it accounts transverse dispersivities at both horizontal and vertical directions and the governing equations are expressed as Eqs 1.4 and 1.5 in Section 1.1.

In this investigation, solute transport within the domain was simulated with HGS, using the "true" and estimated K and S_s fields. The total simulation duration was set to be 300 years. The performance of the different model cases in predicting solute transport were then assessed by comparing simulation results in terms of plume patterns, Cl concentrations at sampling locations, breakthrough curves and their temporal moments.

4.5.2 Simulation Results

Fig. 4.8 illustrates the contour maps of the Cl plume simulated for the four model cases along the cross-section Northwest-Southeast at four selected times: Year 5 (early time), Year 50 (source removal), Year 100 (peak concentration arrival), and Year 300 (late time). The simulated Cl plumes associated with the "true" K and S_s fields are also included at the bottom for the purpose of comparison. The outer bound of these plumes is set to be 1×10^{-6}.

Examination of Fig. 4.8 reveals that Case 3b provides the best prediction results, yielding Cl plumes quite similar to the observed ones at all time stages. Without incorporating the

stratigraphic information for inverse modeling, Case 1 and Case 3a fail to capture the migration of Cl, especially at the early and late time stages. It is surprising to find that even with known stratigraphic information, Case 2 yields the worst prediction results in comparison to other investigated model cases. This may be attributed to the inaccurate estimation of hydraulic parameters (K and S_s) in the source layer (AF1), where few hydraulic head data are available for model calibration. The simulated Cl plume using the calibrated geological model is found to be distinctly enlarged with the presence of source (Years 5 and 50), but rapidly diluted after the removal of source (Years 100 and 300). These results reveal that solute transport is strongly impacted by the heterogeneity of hydraulic parameters (K and S_s), and the accurate estimation of K and S_s values, as well as their spatial distributions are of critical need for the adequate prediction of solute migration in subsurface conditions.

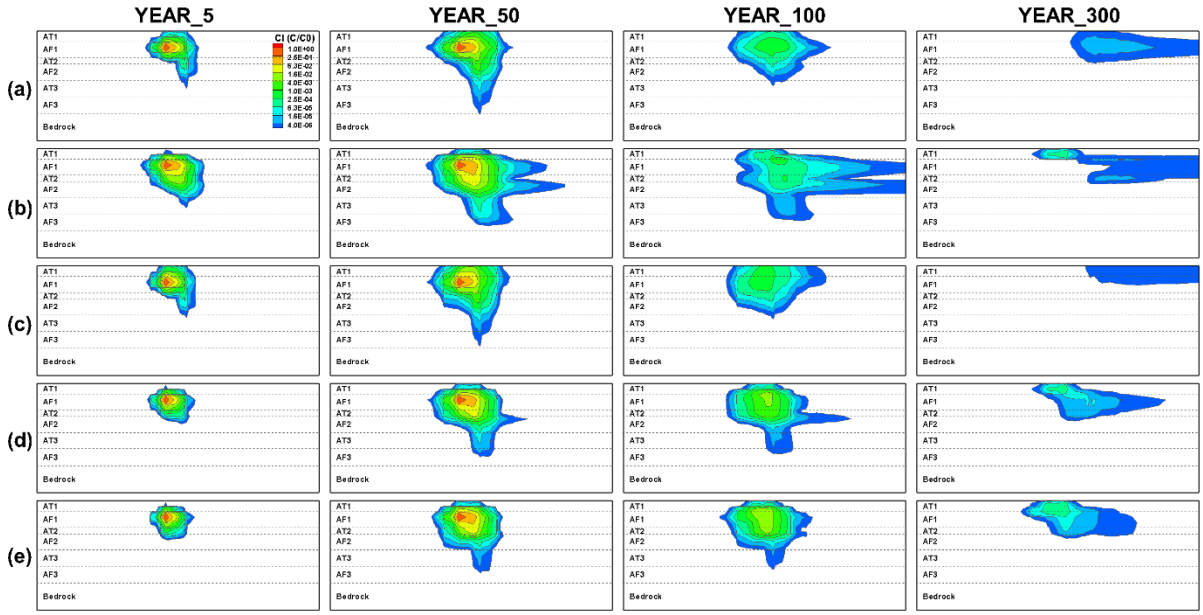

Fig. 4.8: Simulated Cl plumes at four different time stages for four model cases and using the "true" K and S_s fields. a) Case 1, b) Case 2, c) Case 3a, d) Case 3d, and e) "true" K and S_s fields.

The simulation results are then assessed by plotting the scatterplots of simulated versus observed Cl concentrations at water-supply and monitoring wells (sampling points) to visualize the spatial distribution of errors in terms of bias and scatter, as shown in Fig. 4.9. Examination of Fig. 4.9 reveals that prediction results for all model cases are improved from the early time stage to peak concentration arrival (Fig. 4.9a through Fig. 4.9c) with all data points approaching the 45° line. This makes sense since more heterogeneity information is captured when the plume extends to a larger area, and the heterogeneous system behaves more like a homogeneous model with effective hydraulic parameters for solute transport prediction. After the removal of the source, the impact of heterogeneity in hydraulic properties on solute transport is enhanced again, resulting in biased prediction results with enlarged scattering for all model cases at the late time stage (Fig. 4.9d). These results reveal that the heterogeneity of hydraulic parameters (K and S_s) would strongly impact the removal of solute from the subsurface and should be accurately characterized for site contaminant remediation.

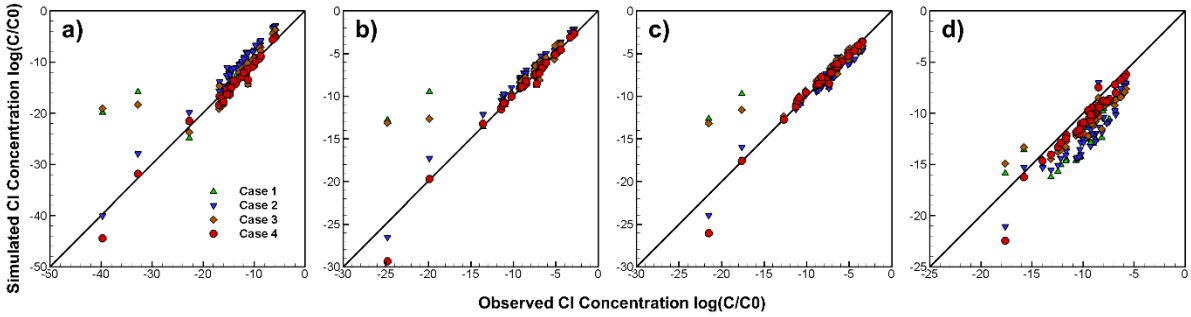

Fig. 4.9: Scatterplots of simulated and observed Cl concentrations at all wells for four model cases at four time stages: a) Year 5 (early time); b) Year 50 (source removal); c) Year 100 (peak concentration arrival); d) Year 300 (late time).

4.5.3 *Breakthrough Curves*

Fig. 4.10 illustrates the breakthrough curves of Cl concentration at three selected sampling points (M8b, M5a, and M4 located in ZONEs 1, 2, and 3, respectively) for four model cases as well as the "true" K and S_s fields. The breakthrough curves of Cl concentration for all sampling points are illustrated as Fig. C42 in Appendix C

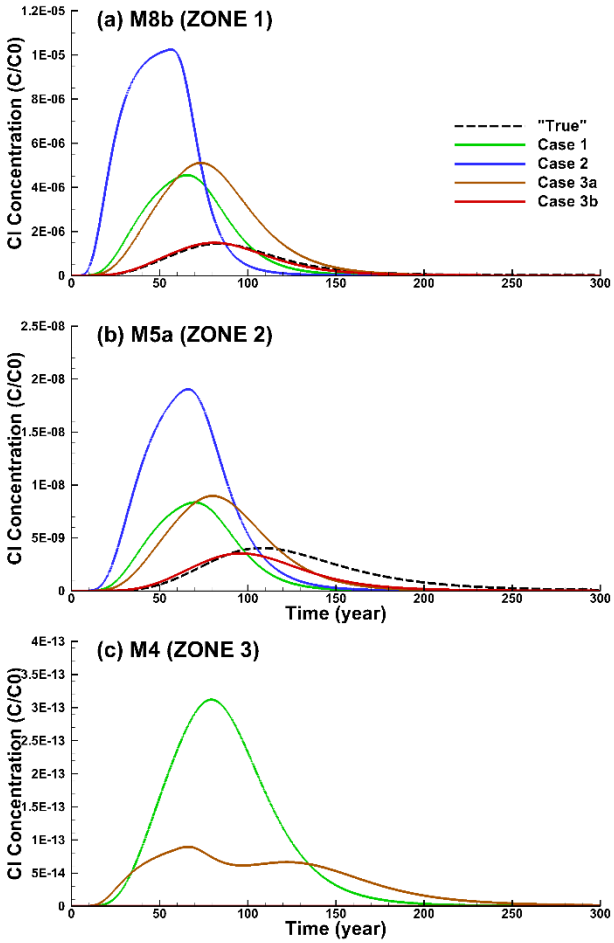

Fig. 4.10: Simulated and observed breakthrough curves of Cl concentration at selected sampling locations (one for each subdivided zones) for four model cases.

In each plot, the "true" breakthrough curve is illustrated as the black dashed line, while the simulated ones from different model cases are illustrated as solid lines with different colors. As shown in Fig. 4.10a, the K and S_s tomograms obtained from the geostatistical model with

geological information (Case 3b) can be utilized to adequately capture the behavior of solute transport, yielding the simulated breakthrough curve at the sampling point M8b be consistent with the “true” one. In contrast, the geology-based zonation model (Case 2) yields quite poor prediction results, with much higher peak concentration, earlier arrival time, and shorter late-time tail in comparison to the “true” breakthrough curve. This is the case for most sampling points located within ZONE 1.

In ZONE 2, where hydraulic head data are lacking for inverse modeling, Case 3b yields slightly biased prediction results at the late-time simulation period, as shown in Fig. 4.10b. Nevertheless, it still performs the best in predicting solute transport in comparison to other model cases. For ZONE 3, the sampling point (M4) located at the bottom layer (Bedrock) is selected and the corresponding breakthrough curves are compared, as shown in Fig. 4.10c. Without incorporating geological information for model calibration, Cases 1 and 3a yield significantly enhanced Cl concentrations at the bottom of the simulation domain. In the following section, temporal moment analyses are presented to quantitatively compare the simulated breakthrough curves to the “true” ones.

4.5.4 Temporal Moment Analysis

Instead of characterizing the breakthrough curves at all wells, two sampling points (M4 and M8a) at the bottom layer (Bedrock) were excluded for temporal moment analysis, since the Cl plume is simulated to be mainly present in the upper layers. The n-th temporal moments (M_n) of Cl concentration at location (x, y, z) at time (t) are given by:

$$M_n = \int_0^{\infty} t^n C(x, y, z, t) dt \qquad \text{(Eq. 4.1)}$$

where t is the time, and C is the Cl concentration. The zeroth (M_0), first (M_1), and second (M_2) moments for all characterized breakthrough curves were then computed through numerical integration of the breakthrough data.

For each breakthrough curve, the calculated M_0 is used to describe the total mass of Cl passing through the corresponding well during the simulation duration. The first normalized moment is used to estimate the mean arrival time of the center of Cl mass (μ):

$$\mu = \frac{M_1}{M_0} \qquad \text{(Eq. 4.2)}$$

The variance σ^2 of breakthrough curves is calculated through:

$$\sigma^2 = \frac{M_2}{M_0} - (\frac{M_1}{M_0})^2. \qquad \text{(Eq. 4.3)}$$

In general, the σ^2 represents the spread of the concentration distribution and is influenced by mechanical dispersion and molecular diffusion. In other words, this parameter can be used to describe the heterogeneity levels of hydraulic parameters within the simulation domain. The calculated M_0, μ, and σ^2 of the simulated and "true" breakthrough curves were compared, with the comparison scatterplots illustrated in Fig. 4.11.

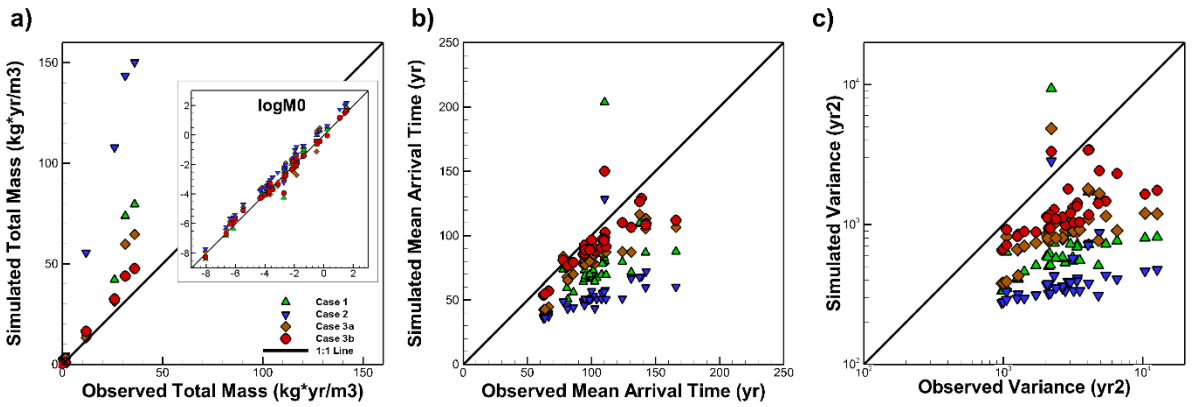

Fig. 4.11: Temporal moment analysis of simulated versus observed breakthrough curves for four model cases. a) total mass (M_0), b) mean arrival time (μ), and c) variance (σ^2).

Fig. 4.11a reveals that at the wells highly impacted by the Cl plume, significantly large M_0 values are estimated from the geological model (Case 2) in comparison to the observed ones. The estimation of M_0 at these wells improves gradually when the effective parameter model (Case 1) and the geostatistical model without geological information (Case 3a) are utilized for prediction, while the geostatistical model with geological information (Case 3b) yields the best estimation of M_0 with smallest discrepancy between the simulated and observed values.

To enlarge the comparison results at the wells with small M_0 values, the logarithm of the simulated and "observed" M_0 estimates are computed and compared, as shown in the subplot of Fig. 4.11a. The comparison results show that the geostatistical model with geological information is able to adequately estimate M_0 values at almost all wells with all data points clustered around the 45° line; however, biased M_0 estimates with relatively large scatters are obtained from other model cases. These results imply that detailed heterogeneity and accurate K and S_s estimates are required to adequately capture the total solute mass.

The comparison of the simulated and observed mean arrival time (μ) for all model cases

is illustrated in Fig. 4.11b. Results show that the estimated mean arrival time at all wells were, on average, shorter in comparison to the observed ones for all model cases. This may be attributed to the poor estimation of K and S_s values in the source layer (AF1), where hydraulic head data are limited for detailed heterogeneity characterization, resulting in biased prediction of solute transport at early time. However, Case 3b still yields the best estimation of the mean arrival time with relatively smaller discrepancy between the simulated and observed values in comparison to other model cases. Based on these results, geostatistics-based HT is suggested to reveal heterogeneity details for more accurate estimation of the mean solute arrival time, which is in line with the conclusion provided by Illman et al. (2012) based on a sandbox study.

Fig. 4.11c illustrates the comparison of the simulated and "observed" variances (σ^2). In general, Case 3b still performs the best in estimating the variances, followed by Cases 3a and 1, while Case 2 yields the worst result. However, the computed variances of breakthrough curves are typically smaller with apparent bias for all model cases in comparison to the observed ones, indicating under predictions of temporal spreading of the plume using the estimated K and S_s tomograms. This may be attributed to the loss of heterogeneity details when using municipal well data for large-scale site characterization. Even with geostatistical inversions, heterogeneity details of hydraulic parameters (K and S_s) can only be revealed where there are sufficient hydraulic head data. These results emphasize again that solute transport is strongly impacted by the heterogeneity of hydraulic parameters (K and S_s). Detailed characterization of subsurface heterogeneity at finer scales is suggested for solute transport investigations.

5 Study III: Large-scale Three-dimensional Hydraulic Tomography Analysis of Long-term Municipal Wellfield Operations

5.1 Site Description

5.1.1 Municipal Water-supply Wellfield

The Regional Municipality of Waterloo (the Region) located in Southern Ontario is one of the largest municipal users of groundwater in Canada consisting of more than 40 wellfields with an excess of 100 operating production wells. This study focuses on the characterization of large-scale subsurface heterogeneity at the Mannheim site, a municipal water-supply wellfield located along the western edge of the City of Kitchener within the Region. Fig. 5.1a shows the location of the Region related to the Province of Ontario and Canada, and Fig. 5.1b indicates the study area. The well data at the Mannheim wellfield utilized for the analyses in this study were obtained from the Region's Water Resources Analysis System (WRAS+) database (Regional Municipality of Waterloo, 2014).

The Mannheim wellfield has been in continuous operation since 1949. In total, 16 production wells have been drilled over the last 70 years and are currently under operation within the wellfield. The distribution of these production wells (P1-P16) is illustrated in Fig. 5.1c as red circles. Based on the previous wellfield characterization reports at the Mannheim site (e.g., Golder Associates Ltd., 2011), the Mannheim wellfield is subdivided into four

smaller well sites, which are identified as Mannheim East and West, Peaking, and Aquifer Storage and Recovery (ASR), with different operation strategies. Specifically, the Mannheim East and West are two original well sites constructed within the wellfield, and the production wells at these two sites (P1-P3 at East and P14-P16 at West) are extracting groundwater with relatively constant flow rates for continuous municipal water supply. The Peaking site is suited to the north of the wellfield, in which, production wells (P4-P7) are designed to extract groundwater only during high municipal water demand periods. The ASR site, located in the south of the wellfield, is designed to inject and store treated water from Grand River (approximate 7 km east of the Mannheim wellfield) during low water demand periods, and the stored water is extracted during high demand periods. In particular, production wells P8 through P11 at the ASR site are used for both pumping and injection, while P12 and P13 are used for pumping only. All of these production wells are screened at the bottom of the water-supply aquifer with elevation ranges from 306 to 328 meters above sea level (masl) and connected electronically to the operation center of the Mannheim wellfield where the flow rates of each production well are monitored and controlled.

To ensure that the operation of production wells is within the capacity of the pumped aquifer, a monitoring network is installed at the Mannheim wellfield with most observation wells concentrated within the production area, while some of them are installed in groundwater recharge areas. In this study, 36 observation wells screened at the water-supply aquifer or screened at the upper and lower hydrogeological layers but within the production area were selected for HT analyses. The distribution of these wells is illustrated in Fig. 5.1c and Fig. 5.1d

for plan and cross-section views, respectively, with green squares indicating the observation wells with pressure transducers, while blue squares indicate wells with manual measurements of groundwater levels.

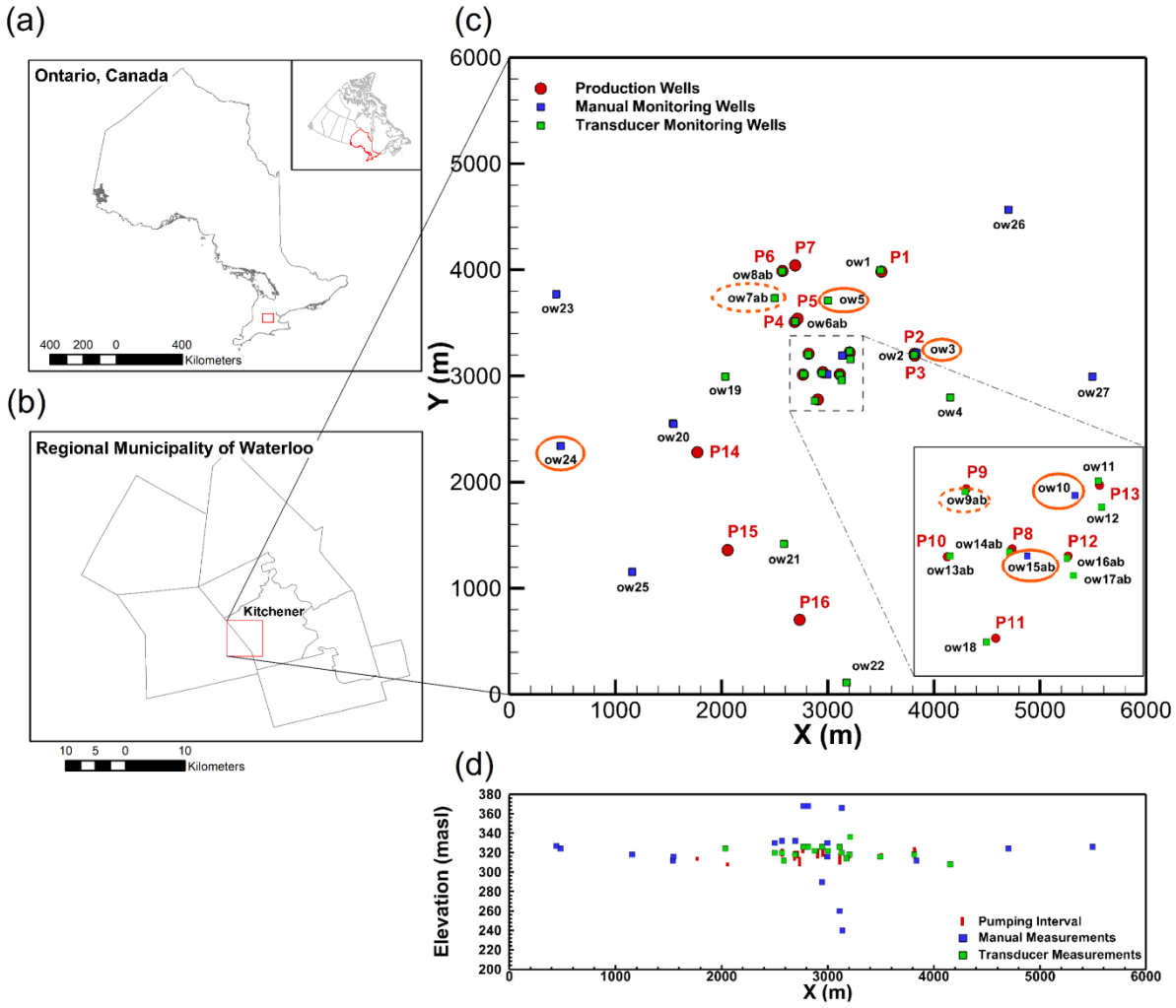

Fig. 5.1: Location of the Mannheim wellfield along with the distribution of production and observation wells within the simulation domain. (a) location of the Regional Municipality of Waterloo (RMOW) in Ontario, Canada, (b) location of the study area within the RMOW, (c) plan view of well locations, and (d) cross-section view showing well screen locations.

5.1.2 Construction of Geological Model

The Mannheim wellfield is located within the core area of the Waterloo Moraine, which is a quaternary kame moraine formed by interlobate glacial activity during the late Wisconsinan glaciation stage (Martin and Frind, 1998). Numerous advances and retreats of ice lobes have resulted in a complex multi-aquifer/aquitard system within the moraine, in which, glacial outwash sands and gravels are separated by silt and clay-rich glacial tills. The quaternary geology and hydrogeology of the Waterloo Moraine have been investigated through previous studies. Initially, Karrow (1993) identified four major glacial tills throughout the Waterloo Moraine (from youngest to oldest, they are Pre-Catfish Creek, Catfish Creek Till, Maryhill Till, and Tavistock/Port Stanley Till), which were considered as aquitards that separate three major aquifers within the moraine. Later, Martin and Frind (1998) constructed a three-dimensional (3-D) hydrogeologic model of the multi-aquifer/aquitard system for better understanding of the hydrogeological condition within the moraine. Most recently, Bajc and Shirota (2007) mapped the surficial deposits of the Region, resulting in a detailed 3-D conceptual hydrogeologic model for the Waterloo Moraine with 19 identified layers (shown as Fig. D1 in Appendix D).

In this study, a 3-D geological model for the Mannheim wellfield was constructed based on the available borehole logs within the study area. The dimensions of the geological model are 6000 m × 6000 m in East (x) and North (y) directions, with the Mannheim wellfield located approximately in the center of the domain, and from 200 masl to the topography in elevation (z). Specifically, the lithology information from 92 wellbores installed within the study area were obtained from the WRAS+ database and interpolated using a commercial software

Leapfrog Geo (Seequent, 2020) for the geological model construction. The distribution of these wellbores as well as their reliabilities on the classified lithologies are illustrated in Fig. D2. The reliability information of borehole logs implies that the constructed geological model may still involve uncertainty regarding the interpolated layer boundaries, especially for the area outside the production zones.

Fig. 5.2 illustrates a perspective view of two cross-sections from the constructed geological model. In total, eight groups of geological units were identified beneath the study area following the work of Bajc and Shirota (2007).

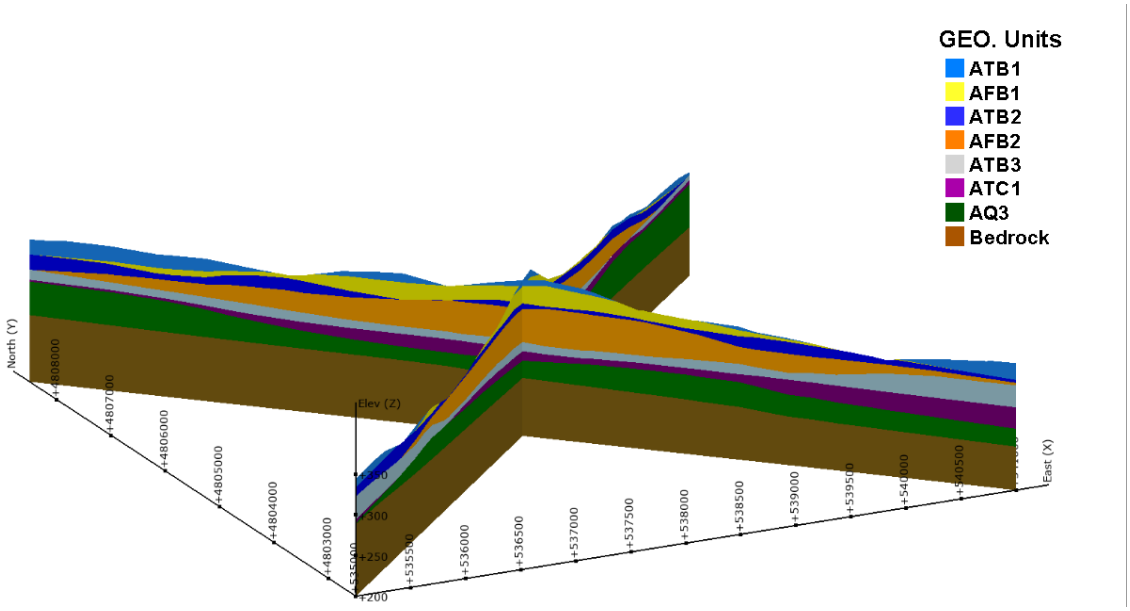

Fig. 5.2: Perspective view of two diagonal cross-sections of the geological model constructed using available borehole logs within the study area.

These units (expect for AQ3) were identified using the nomenclature of Ontario Geological Survey (OGS), in which, AF represents aquifer and AT means aquitard. The letter and number following AT or AF indicate the sequence of groups and units, with “A” as the youngest group followed by “B” and “1” as the youngest unit in group followed by “2”. The geological units

identified in this study are summarized in Table D1 along with the detailed lithology information of each unit. In the Mannheim wellfield, the shallow aquifer AFB2 is the main water-supply aquifer, in which, all production wells and most observation wells are screened.

Based on the constructed geological model (Fig. 5.2), AFB2 is laterally extensive throughout most of the study area with its thickness ranging from 5 to 50 m and appears to pinch out in the southwestern portion of the study area. Such a geometry of AFB2 is consistent with that presented in previous site investigation reports (e.g., Golder Associates Ltd., 2011) . AFB2 is found to be bounded above by aquitard ATB2 and below by aquitard ATB3. ATB2 is characterized as a thin and discontinuous aquitard that separates the water-supply aquifer AFB2 and the upper unconfined aquifer AFB1 in most of the study area. ATB3 is interpreted to be laterally extensive throughout the study area, which physically separates the upper aquifer system (AFB1/ATB2/AFB2) from the lower aquifer system (identified as AQ3 in this study).

In comparison to the conceptual hydrogeological model of the Waterloo Moraine completed in the work of Bajc and Shirota (2007), as well as the geological model of the Mannheim wellfield constructed by Tong et al. (2021), several geological layers previously identified at the lower portion of Waterloo Moraine were merged for the newly constructed geological model in this study (as indicated in Table D1). This is because: (1) these geological layers are relatively thin and spatially discontinuous beneath the Mannheim wellfield (Golder Associates Ltd., 2011) , (2) most borehole logs within the study area are limited to shallow aquifer/aquitard system (AFB1/ATB1/AFB2) investigation, and the limited lithology

information below the Lower Maryhill Till (ATB3) renders the accurate delineation of these layers difficult, and (3) only few observation wells with manual measurements of hydraulic head are available at the lower Waterloo Moraine within the Mannheim wellfield, as illustrated in Fig. 5.1b. In the case of uncertain stratigraphic information and few observation data, the study presented in Chapter 3 suggested the utilization of a simplified geological model as prior structural information for geostatistical analysis of HT data for subsurface heterogeneity characterization.

5.2 Experimental Setup

5.2.1 Model Setup

For groundwater flow modeling, the simulation domain, which has the same dimension as the constructed geological model, was discretized into 37,780 triangular prism elements with 20,626 nodes, as shown in Fig. D3. On the horizontal plane (2-D), the mesh was refined at locations where there were production and observation wells. In the vertical direction, the mesh was gradually refined from the bottom (20 m) to the top (3 m) due to fact that all production wells and most observation wells are screened in the upper located water-supply aquifer (AFB2). Topography was also considered in this study for large-scale site characterization. Since tetrahedron elements have not been implemented in the current version of the code (SimSLE) adopted here for geostatistical inversions, the topography is expressed as steps, as shown in Fig. D3b. Without accounting for surface water flow (e.g., runoff) in this study, such step-shaped topography is considered to have minimal impacts on estimation results.

For the presented analysis, transient groundwater flow was considered to simulate long-term water-level variations induced by water-supply operations at different locations with variable flow rates in a municipal wellfield. To mimic the real hydraulic conditions at the Mannheim wellfield, regional groundwater flow as well as recharge from the top boundary of the model were included for the analyses presented in this study. Specifically, specified head values were assigned at lateral boundaries while specific flux was assigned at the top boundary. The determination of initial and boundary conditions as well as their uncertainties regarding inverse modeling are discussed in the following section.

5.2.2 Initial and Boundary Conditions

In this study, initial and lateral boundary conditions for groundwater flow modeling were approximated based on static-water level measurements taken from the water-supply aquifer (AFB2). In detail, the method of kriging was applied to interpret 31 static water-level measurements collected at different locations within the study area during September and October in 2006 and 2008 (Golder Associates Ltd., 2011) . Specifically, a spherical model with a range of 0.3 and a nugget value of 0 was fit to yield a modeled semi-variogram for kriging. The distribution of these measurements as well as the interpolated 2-D initial head map are illustrated in Fig. D4a. The interpolated head values were then assigned to the entire simulation domain to yield the initial head distribution for inverse modeling, as shown in Fig. D4b. Since the regional groundwater flow is approximately horizontal within the study area and no static water-level measurements were available at lower layers, no natural hydraulic gradient in the vertical direction was initially considered. It should be noted that kriging of static water-level

measurements does not consider the heterogeneous nature of the subsurface, rendering the interpolated head distribution to involve large uncertainty in representing the true hydrostatic condition within the study area.

On the other hand, four lateral boundaries of the model were set as specified head boundaries equal to the interpreted initial heads. Such lateral boundary conditions result in the simulated regional groundwater flow similar to that observed in field (Golder Associates Ltd., 2011) , where groundwater flow takes place primarily from the northwest to southeast with a natural hydraulic gradient less than 0.002 in the horizontal direction. However, the identified lateral boundary conditions may still involve great uncertainty regarding the assigned head values, which could result in a biased simulation of long-term hydraulic head data at monitoring wells. To minimize the effect of uncertain initial and lateral boundary conditions on inverse modeling, a certain simulation period before data assimilation was incorporated for model spin-up (uncertain initial condition), as suggested by Yu et al. (2019), and water-level variations based on a datum point in each observation well were utilized as observation data for model calibration (uncertain boundary condition). Detailed descriptions of these procedures are provided in subsequent sections. The effectiveness of these procedures was further assessed by calibrating the geology-based zonation model with linearly interpolated initial and boundary conditions (not shown here), with results showing that the obtained K and S_s estimates for each geological unit are almost the same in comparison to the case with kriged initial and boundary conditions.

For groundwater recharge, the top boundary of the model was set as a specified flux boundary with nodal flux estimated from precipitation data. Specifically, precipitation data during the simulation period (2012-2013) were obtained from the University of Waterloo weather station, which is located approximately 9 km north of the Mannheim wellfield. Excluding the partitions of evapotranspiration, surface runoff and streamflow, the net groundwater recharge in this study was estimated to be 25% of the total precipitation, yielding the estimated annual groundwater recharge within the study area (195 mm for 2012 and 301 mm for 2013), which agrees with the regional groundwater recharge map provided in a previous wellfield characterization report (Golder Associates Ltd., 2011) . Due to the fact that most observation data were collected from water-supply aquifer AFB2, which located 20-50 m beneath the ground surface and overlain by two aquitards ATB1 and ATB2 (as shown in Fig. 5.2), groundwater recharge in this study was simplified as constant and continuous flux from the top boundary as variations should be muted. In a different case, variable recharge rates based on daily precipitation data was utilized for the geology-based zonation model calibration (not shown here), with results showing negligible differences in comparison to the use of a constant averaged recharge rate. This suggests that ATB1 and ATB2 mutes the high frequency recharge responses into AFB2 for this study area.

5.2.3 Data Utilized for Inverse Modeling

Hydraulic head data collected from 36 monitoring wells during the year of 2013 (Fig. D5) and pumping/injection flow rate records at all production wells during the years of 2012 and 2013 (Fig. 5.3) were obtained from the WRAS+ databased and subjected to HT analyses for

subsurface heterogeneity characterization. Due to the continuous operation of production wells within the wellfield, we incorporated pumping/injection rate records for inverse modeling go back one year prior to the hydraulic head data to eliminate the influence from previous pumping/injection operations (Harp and Vesselinov, 2011; Luo and Illman, 2016).

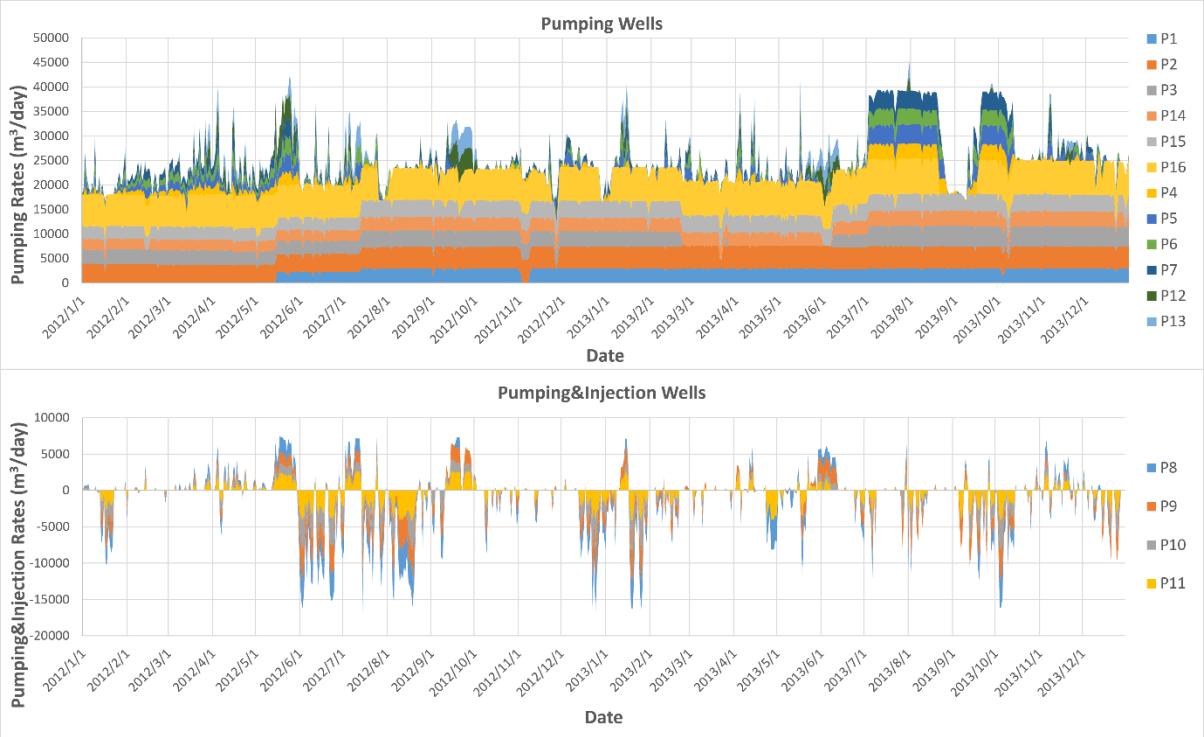

Fig. 5.3: Cumulative pumping (positive) and injection (negative) flow rates at all 16 production wells during the years of 2012 and 2013 from the WRAS+ database.

Instead of real-time flow rates, pumping/injection rates in production wells were recorded as daily pumped/injected volume of water in the WRAS+ database. As shown in Fig. 5.3, the flow rate at the production wells with seasonal operation strategy (P4 through P13) appears to be highly variable, which has resulted in an intricate hydraulic condition within the production area. Hydraulic head data at monitoring wells were recorded as elevation in the WRAS+ database, and these data were processed to remove barometric effects prior to its inclusion into the database. For the monitoring wells with a pressure transducer, hydraulic heads were

continuously recorded every hour during the selected period, while the head data at the beginning of each day (12:00 am) were extracted for analyses in this study. On the other hand, monitoring wells with manual measurements recorded hydraulic head data at a frequency varying from monthly to bi-monthly. For these wells, all hydraulic head records were utilized.

To minimize the effect of uncertain boundary conditions on inverse modeling, water-level variations at each monitoring well based on the corresponding datum point were computed and utilized as observation data for model calibration and validation, as shown in Fig. 5.4. For the presented analyses, hydraulic head data recorded at the date of 2013/01/10 were utilized as datum points because (1) these hydraulic head data have minor influence from water-supply operations and (2) most observation wells have hydraulic head measurement at this date. For those where head measurements at 2013/01/10 not being available, a different date was selected with corresponding head measurements utilized as their datum points (ow9b and ow13b: 2012/07/13; ow6b, ow7b, and ow8b: 2013/01/11).

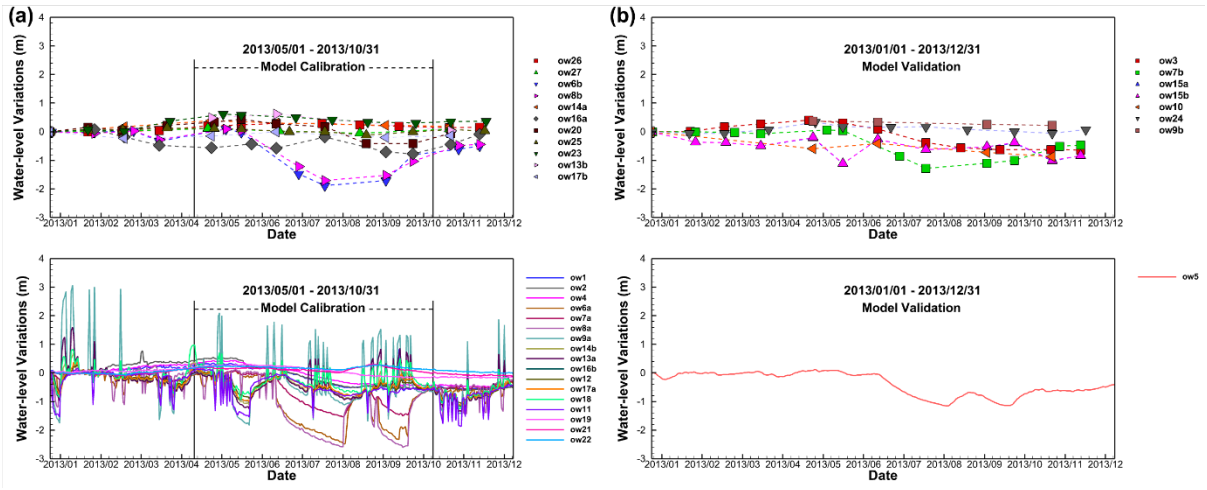

Fig. 5.4: Computed water-level variations based on the measured hydraulic head at datum points. (a) shows the 28 monitoring wells selected for model calibration along with indicated observation period, while (b) shows the eight monitoring wells left for model validation.

The temporal sampling strategy for transient HT analysis of long-term municipal well data has been investigated through a synthetic study on HT analysis of municipal well records described in Chapter 4. Through the utilization of five datasets consisting of different time durations and periods for model calibration, we concluded that continuous data points with large water-level variations should be selected and incorporated for large-scale heterogeneity characterization. Following the previous synthetic study, water-level variations at 28 monitoring wells from May to October were utilized for model calibration, as shown in Fig. 5.4a. The significant change of groundwater levels during this period was mainly attributed to the seasonal operation of production wells at the subdivided ASR (early time) and Peaking (intermediate and late times) sites, as indicated in Fig. 5.3.

Model validation in this study was performed under two scenarios. For Scenario 1, the water-level variations at eight monitoring wells (municipal well records) that have not been used in calibration efforts were utilized for model validation (Fig. 5.4b). The distribution of these wells is illustrated in Fig. 5.1 and indicated by orange circles. These wells were selected to assess the performance of HT analysis of municipal well data in revealing heterogeneity details at different locations: 1) close to the production well (ow3); 2) located at different subdivided well sites and screened at the water-supply aquifer (ow5, ow7b, and ow15ab); 3) within the production area, but screened at the upper of lower geological units (ow9b and ow10); and 4) away from the production area, but screened at the water-supply aquifer (ow24).

For Scenario 2, the dedicated pumping test data from wells P1, P2, and ow20 were obtained

from previous site investigation reports (CH2MHILL and Papadopulo&Associates, 2003) and utilized for model validation in this study. Due to the lack of raw data, observation data for these dedicated pumping tests were extracted from the plotted drawdown curves using the digitization tool WebPlotDigitizer (Rohatgi, 2021). The drawdown curves and extracted data points are illustrated in Fig. D6, and the details of each test are summarized in Table D2.

5.3 Inverse Modeling Approaches

5.3.1 Geology-based Zonation Modeling

In this study, the Mannheim wellfield was firstly characterized using a geology-based zonation model, in which, a uniform set of K and S_s was estimated and assigned to each geological unit to describe its hydraulic properties. Although the site has been previously characterized using different geology-based zonation models through transient HT analysis of the municipal well records (Tong et al., 2021), differences between these models and the one adopted here are apparent in terms of model conceptualization, grid generation, data quantity, and data processing for inverse modeling. Specifically, in comparison to those adopted in the study of Tong et al. (2021), the geology-based zonation model utilized in this study has a larger simulation domain with merged lower geological units, as described above. Also, the grid of the geology-based zonation model is customized so that it can be used for geostatistics-based THT analysis. Moreover, three additional production wells (P14 through P16) with more observation wells were included for the presented THT analyses. Most importantly, novel data processing and analyses strategies were proposed in this study regarding the uncertainties of

identified initial and boundary conditions.

The geology-based zonation model was calibrated using the groundwater flow model HGS (Aquanty, 2019) coupled with the parameter estimation code PEST (Doherty, 2015) by simultaneously matching all data points. For this study, an *R* script was created to compute water-level variations based on HGS output files and to create input files for PEST. Fig. 5.5 illustrates the flowchart designed for the geology-based zonation model calibration.

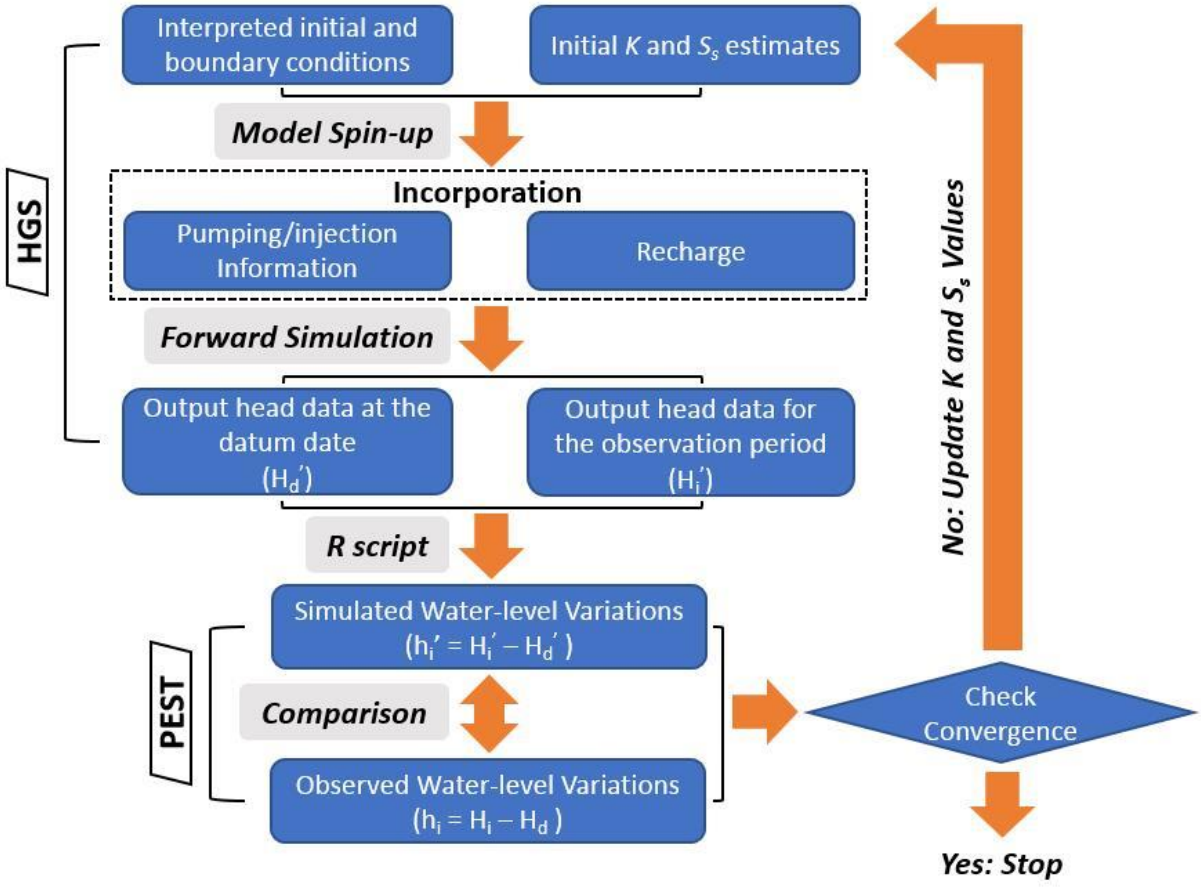

Fig. 5.5: Flowchart designed for geology-based zonation model calibration through coupling of HGS and PEST.

In summary, the calibration started with a certain simulation period with interpreted initial and boundary conditions as well as initial K and S_s estimates for model spin-up. Once the

system reached a hydrostatic condition, pumping/injection and recharge information start from the beginning of 2012 were incorporated to simulate the hydraulic head data at selected datum points and during the observation period, which were then used to compute simulated water-level variations in each observation well. In the end, PEST was employed to compare the simulated and observed water-level variations and update K and S_s values for the next iteration. The iteration process continued until model convergence was achieved by the stabilization of the differences between simulated and observed water-level variations.

To avoid unrealistic K and S_s estimates in geological units where no hydraulic head data are available, initial K and S_s guesses for geology-based zonation model were obtained based on the predominant materials for each geological unit. Table D1 summarizes these initial guesses. For aquifers, the lower and upper bounds in PEST were set as 8.64×10^{-1} and 8.64×10^{2} m/day for K, and 4.9×10^{-5} and 1.0×10^{-3} /m for S_s. For aquitards, the lower and upper bounds were set as 8.64×10^{-4} and 8.64×10^{-1} m/day for K, and 9.2×10^{-4} and 2.0×10^{-2} /m for S_s.

5.3.2 Geostatistical Inverse Modeling

The subsurface heterogeneity within the Mannheim wellfield was then characterized through THT analysis of municipal well records with a highly parameterized geostatistical model. Geostatistical inversions were carried out using the Simultaneous Successive Linear Estimator (SimSLE), originally developed by Xiang et al. (2009) and modified later to account for temporally variable pumping/injection flow rates. Also, a reduced covariance matrix

storage technique was employed to save on memory required for geostatistical inversion with a great number of nodes and data. Specifically, the covariance matrix was banded by considering the cross-covariance between head and hydraulic parameters to be zero when the value is smaller than a user-defined percentage of the variance. The fundamental concepts behind SimSLE and its mathematical algorithms used for parameter estimation can be found in Section 1.2 and Appendix A, respectively. It should be aware that other than the estimated K and S_s tomograms, SimSLE also provides the residual variance maps of ln$K$ and lnS_s by revising the covariance matrix of the parameters in each iteration to describe the uncertainty associated with the estimated values.

The efficiency of SimSLE in interpreting dedicated HT data for subsurface heterogeneity characterization has been evaluated through many studies (e.g., Illman et al., 2015; Luo et al., 2017; Xiang et al., 2009; Zhao and Illman, 2018). However, it may still be restricted in interpreting long-term municipal well records with uncertain initial and boundary conditions due to the fact that: (1) the computationally intensive nature of highly parameterized geostatistical model makes it impractical to involve a long simulation period for model spin-up and (2) it is unable to compute water-level variations based on simulated hydraulic head data at datum points and during observation period for model calibration.

To overcome these limitations, a flowchart that is similar to, but more complicated than the one for geology-based zonation model was proposed for geostatistical inversions, as shown in Fig. 5.6.

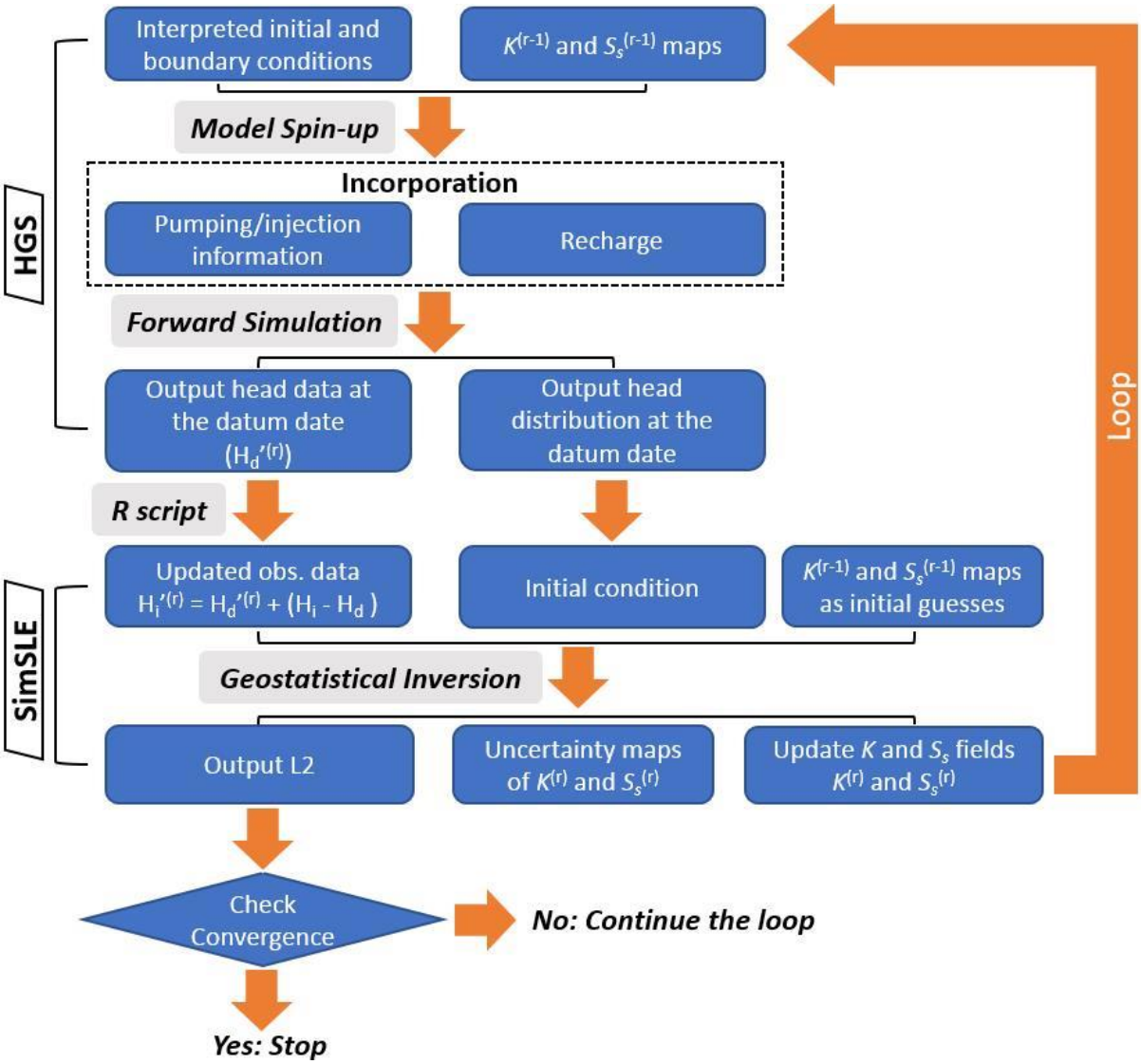

Fig. 5.6: Flowchart designed for geostatistical model calibration, with the joint utilization of HGS and SimSLE. The utilization of HGS reduces time costs for model spin-up and updating observation data in comparison to that using the forward simulator (VSAFT3) implemented in SimSLE for model spin-up.

In each iteration, an HGS model was initially adopted for model spin-up and used to generate hydraulic head data at the selected date (2013/01/10) based on the K and S_s tomograms from the previous iteration. The simulated datum head data in each observation well was then used to update the observation data (in the form of pressure head) for model calibration using SimSLE, while keeping the water-level variations in this well same as the observed ones. On the other hand, the simulated head distribution throughout the simulation domain at the date of

2013/01/10 was applied as an initial condition for geostatistical inversion. In the end, SimSLE was employed to update K and S_s fields for the next iteration. The iteration process continued until convergence was achieved by stabilization of the mean square error (L_2) of simulated and modified hydraulic head data.

In this study, the knowledge of stratigraphic data based on the constructed geological model was incorporated as prior information for subsurface heterogeneity characterization using the highly parameterized geostatistical model. Specifically, the calibrated geology-based zonation model was utilized as initial K and S_s guesses for the first iteration of geostatistical inversion. These initial guesses were then updated iteratively based on the estimated K and S_s fields from the previous geostatistical inversion iteration. The variances of K and S_s ($\sigma^2_{\ln K}$, $\sigma^2_{\ln S_s}$) were initially set to be 4.0 and 2.0, respectively, while the correlation lengths were set to be $\lambda_x = 200\ \mathrm{m}$, $\lambda_y = 200\ \mathrm{m}$, and $\lambda_z = 10\ \mathrm{m}$ for both K and S_s. The variance values were approximated based on the T and S estimates at the Mannheim East site obtained from the work of Luo and Illman (2016), while the correlation lengths were approximated based on the size of subdivided well sites, as well as the thickness of identified geological units. Due to the availability of sufficient hydraulic data from municipal well records, such statistical properties input for geostatistical inversions were considered to have negligible effects on characterization results (Yeh and Liu, 2000).

5.4 Results

5.4.1 Estimated K and S_s Tomograms

5.4.1.1 Geology-based Zonation Model

The calibration of the geology-based zonation model treated each geological unit as an isotropic and homogeneous medium, and a uniform set of K and S_s were estimated and assigned to each geological unit. In total, eight sets of K and S_s were estimated from the geology-based zonation model for subsurface heterogeneity characterization within the study area. Fig. 5.7 illustrates the estimated K and S_s tomograms. The corresponding ln$K$ and lnS_s values as well as their 95% confidence intervals are plotted as Fig. D7.

With sufficient hydraulic response data, a reliable K value is estimated for the water-supply aquifer (AFB2) with the narrowest confidence interval ($\ln K_{AFB2}$ = 4.15 ± 0.05 m/day). This value is found to be comparable to that estimated from the work of Tong et al. (2021) ($\ln K_{AFB2}$ = 4.25 ± 0.03 m/day for the case of 11-layer geological model), revealing that the estimation of K is less likely to be affected by the difference of model conceptualization when sufficient observation data are available for geology-based zonation model calibration. It is of interesting to note that relatively narrow confidence intervals are also observed for the K estimates in ATB3 and ATC1, which are two aquitards underlain AFB2, but with no hydraulic response data. This may be attributed to the inclusion of a long-term simulation period (2012/01/01-2013/04/30: prior pumping/injection information; 2013/05/01-2013/10/31: observation data) for model calibration, which results in a large influence area of municipal water-supply

pumping that modifies the regional groundwater flow field in ATB3 and ATC1. In contrast, the estimated K values for the upper and lower geological units (ATB1, AFB1, ATB2, AQ3, and Bedrock) are found to be close to the provided initial values, and the corresponding large confidence intervals indicate that these K estimates are insensitive to the observation data utilized for model calibration.

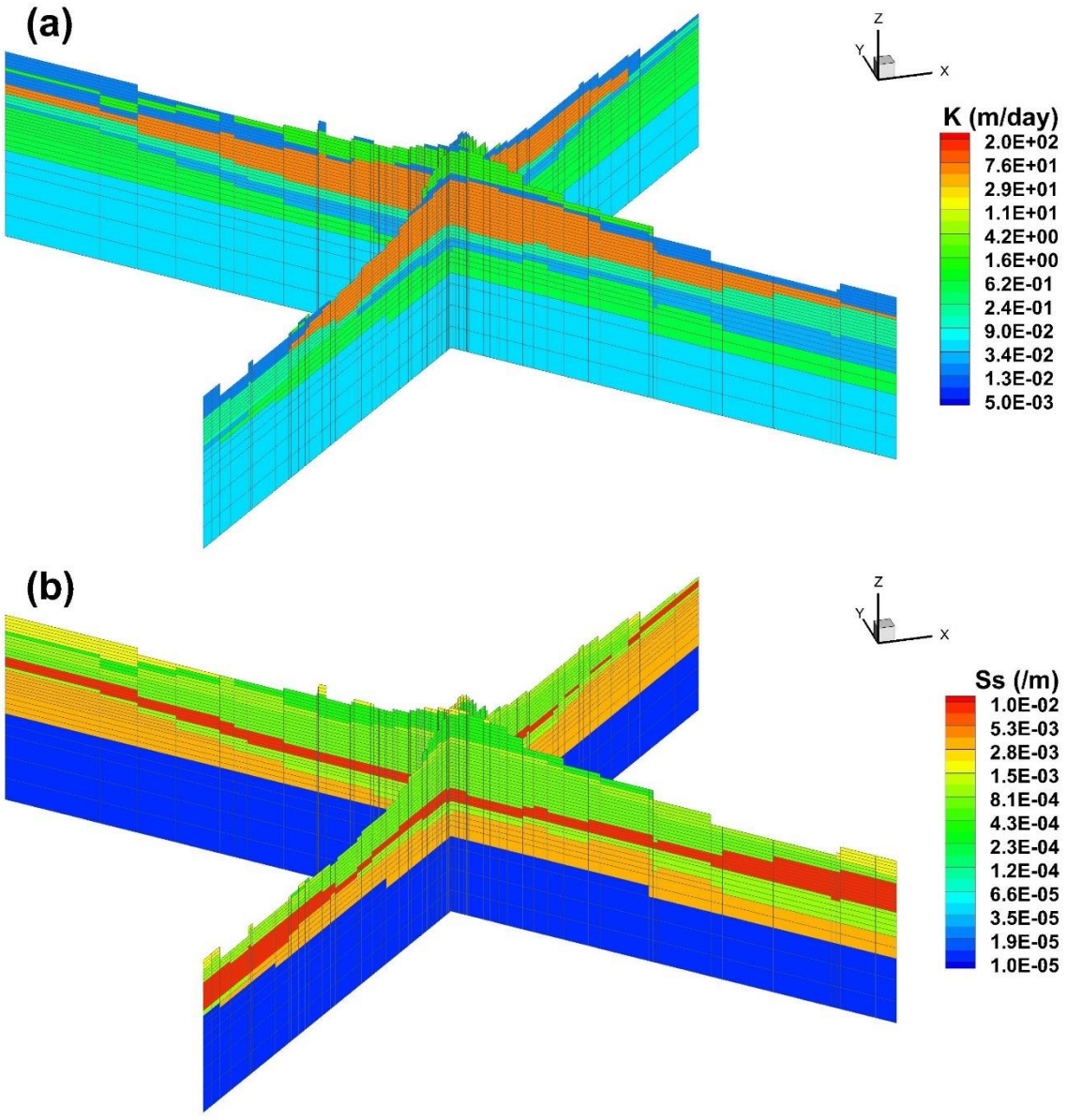

Fig. 5.7: Estimated K and S_s tomograms from the geology-based zonation model. (a) shows the estimated K tomograms, while (b) indicates the S_s tomogram.

On the other hand, S_s estimates obtained from the geology-based zonation model share similar characteristics as the K estimates. In particular, S_s estimates with high confidence are obtained in geological units with sufficient hydraulic head data (AFB2) or when the corresponding hydraulic properties are sensitive to the simulation of observation data (ATB3 and AQ3), while S_s estimates in the rest geological units are found to be close to initial values and the evaluated large confidence intervals indicate high uncertainty associated with these estimates.

5.4.1.2 Geostatistical Inverse Model

Geostatistical inverse modeling of long-term municipal well records updated K and S_s estimates iteratively, and Fig. D8 illustrates the change of L_2 of the simulated and modified hydraulic head data for all completed iterations. In this study, the estimated K and S_s tomograms at the iteration of 16 were selected and plotted as Fig. 5.8 for subsurface heterogeneity characterization within the Mannheim wellfield.

As shown in Fig. 5.8a, geostatistical inversion of the long-term municipal well records with prior geological information reveals great details of intralayer K heterogeneity, similar to the work of Zhao and Illman (2018) who conducted a THT study at a small-scale field site. However, such intralayer heterogeneity in K is not visible in the upper and lower geological units due to the lack of hydraulic response data in these units.

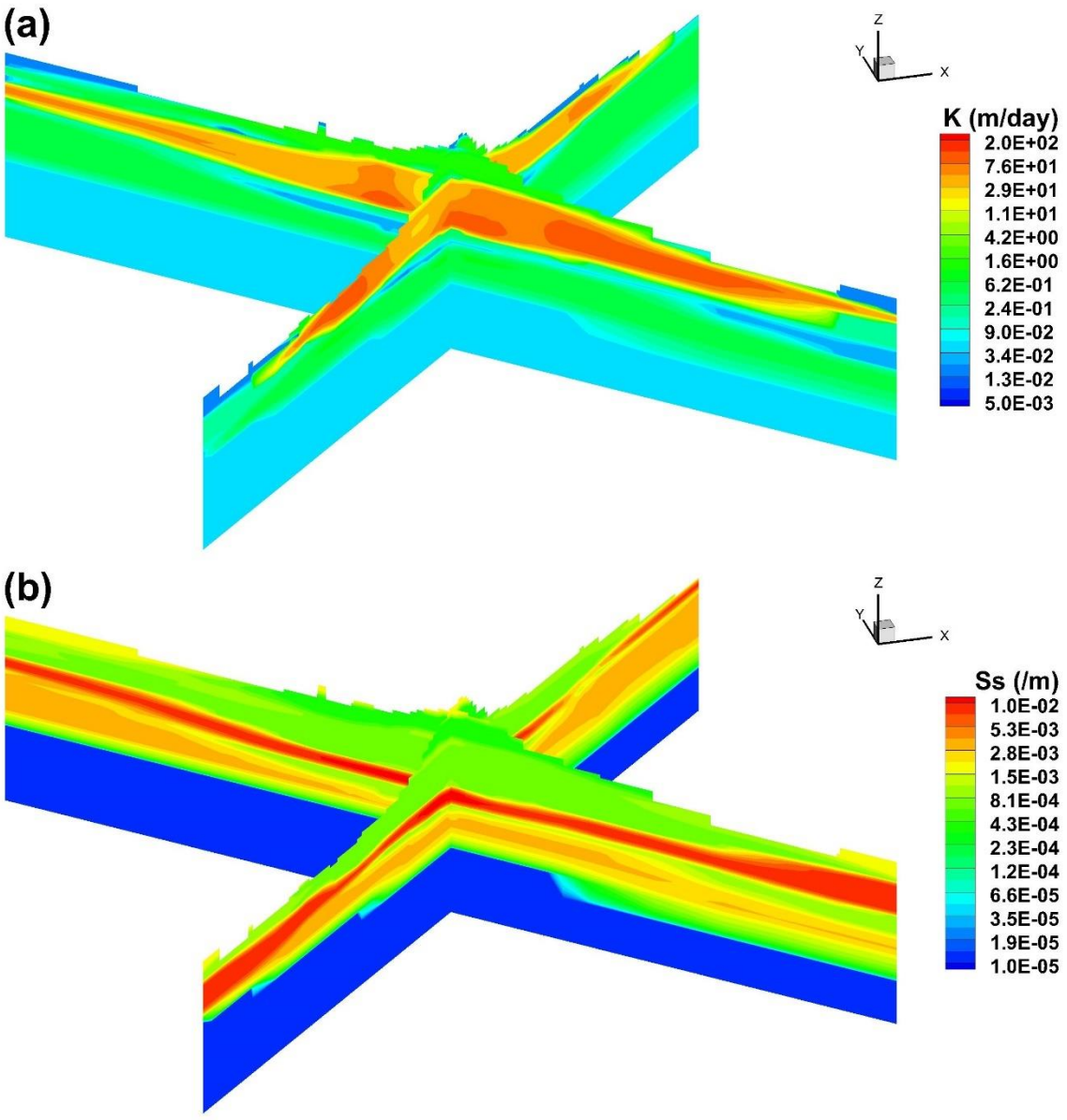

Fig. 5.8: Estimated K and S_s tomograms from the geostatistical model. (a) shows the estimated K tomogram, while (b) indicates the estimated S_s tomogram.

In comparison to the K tomogram, the estimated S_s tomogram from the geostatistical inverse model does not show any distinct intralayer heterogeneity pattern, even for the water-supply aquifer (AFB2) with sufficient hydraulic response data (as shown in Fig. 5.8b). This is in line with the results from the synthetic study described in Chapter 4, in which, smooth S_s tomograms were obtained when characterizing a synthetic multi-aquifer/aquitard system through THT analysis of simulated municipal well data. The failure in revealing heterogeneity

details of S_s may mainly be attributed to the utilization of daily observation data extracted at the beginning of each day (12:00 am) for model calibration in both studies. Such a dataset ignores the early-time water-level variations regarding daily varying pumping/injection flow rates, which have been evaluated to be of critical importance in obtaining reliable S_s estimates (Sun et al., 2013).

The estimated K and S_s values from the geostatistical inverse model were then evaluated regarding their uncertainties based on corresponding ln$K$ and lnS_s residual variance maps (shown as Fig. D9). In general, larger residual variances of hydraulic parameters indicate higher uncertainties of these estimates, and vice versa. Examination of Fig. D9 reveals that the estimated ln$K$ and lnS_s variances are relatively small within the production area of the water-supply aquifer (AFB2), where sufficient hydraulic response data are available for reliable K and S_s estimates. The variances become larger when moving to the upper and lower geological units and towards the lateral boundaries of the simulation domain, indicating higher uncertainty of K and S_s estimates due to the lack of hydraulic response data. On the other hand, estimated lnS_s residual variances are larger than those of ln$K$, suggesting the higher uncertainty of S_s estimates in comparison to K estimates. This again emphasizes that the temporal resolution of observation data selected for the analyses may be restricted for reliable S_s estimates, as discussed above. Furthermore, such uncertainty analyses point out that automatically recording water-level variations using transducers and additional observations wells completed in the upper and lower geological units are highly recommended to improve the characterization results and to yield reliable estimates of K and S_s throughout the simulation domain.

5.4.1.3 *Difference between Two Models*

In order to better illustrate the intralayer and interlayer heterogeneity features of hydraulic properties revealed through the geostatistical interpretation of municipal well records, differences of ln$K$ and lnS_s values obtained from the geology-based zonation and geostatistical models were computed and plotted, as shown in Fig. 5.9.

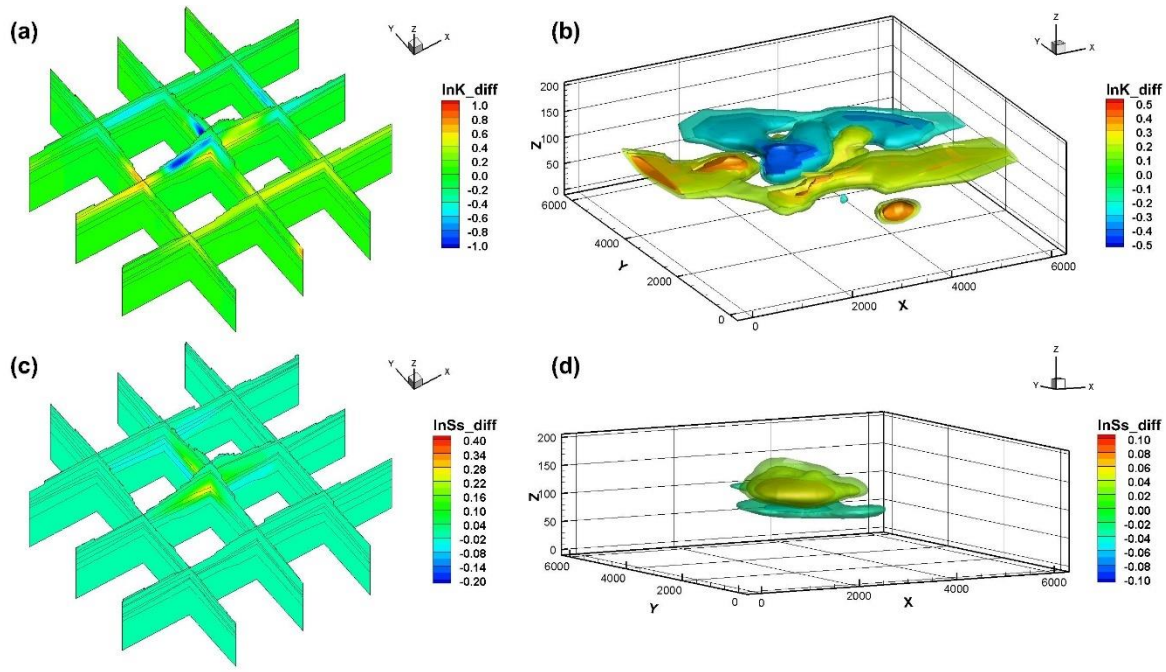

Fig. 5.9: Computed differences of lnK (m/day) and lnS_s (/m) values obtained from the geology-based zonation and geostatistical models. Positive value means higher lnK or lnS_s values associated with the geostatistical model, and vice versa. (a) and (c) show the distribution of lnK and lnS_s differences, respectively, while (b) and (d) illustrate the revealed large-scale heterogeneity patterns for K and S_s, respectively.

In particular, Fig. 5.9a and Fig. 5.9c illustrate the spatial distribution of ln$K$ and lnS_s differences, respectively, along with the identified layer boundaries, while Fig. 5.9b and Fig. 5.9d highlight the revealed large-scale heterogeneity patterns for K and S_s, respectively, using 3-D perspective maps with the outer bounds set to be ±0.2 for K and ±0.03 for S_s. In all of these figures, positive values indicate that the K and S_s estimates obtained from the geostatistical

model are higher than those from the geology-based zonation model, while negative values indicate that smaller K and S_s estimates are obtained from the geostatistical model.

As shown in Fig. 5.9a, large-scale intralayer heterogeneity in K is mainly revealed within the water-supply aquifer (AFB2), where most wells are screened with sufficient hydraulic response data. However, such large-scale intralayer heterogeneity is not observed for S_s (Fig. 5.9c), resulting in estimated S_s values within AFB2 to be similar to those obtained from the geology-based zonation model. This indicates that the selected observation data for model calibration are not sensitive to the spatial distribution of S_s within the study area. Examination of Fig. 5.9a and Fig. 5.9c also reveals that beyond the different K and S_s estimates obtained from the zonation model (initial K and S_s distributions for geostatistical inversions), interlayer heterogeneities in both K and S_s are refined within the production area regarding the upper and lower geological units of AFB2. As mentioned previously, the long-term simulation period incorporated for model calibration results in a large influence area of municipal water-supply pumping, making it possible to estimate K and S_s values in the adjacent aquitards (ATB2 and ATB3) where no observation data are available.

As shown in Fig. 5.9b, the delineated large-scale heterogeneity patterns of K within AFB2 illustrate that a relatively high K zone is identified in the southwest of the domain, while a relatively low K zone is identified in the northeast of the domain. Such a distribution of high and low K zones within the water-supply aquifer (AFB2) is consistent with the K estimates obtained from traditional pumping tests that have been previously performed in the Mannheim

wellfield. In particular, relatively high K values were obtained from the production wells located at the west (Mannheim West site) and south (ASR site) of the wellfield, while relatively low K values were estimated from the dedicated pumping tests conducted at the east (Mannheim East site) and north (Peaking site) of the wellfield. Table D3 summarizes the estimated K values from previously conducted pumping tests using traditional analytical solutions (e.g., Cooper and Jacob, 1946; Theis, 1935). On the other hand, estimated K values in this study are found to be similar compared to those estimated from previous studies in terms of magnitudes, showing that THT analysis of long-term municipal well records is able to yield reliable K estimates. In comparison to K estimates, geostatistical inversion of municipal well records yields slight modification of S_s estimates mainly within the production area (Fig. 5.9d).

5.4.2 Model Calibration

The calibration results were then assessed qualitatively and quantitatively by plotting the scatterplots of simulated versus observed water-level variations and computing the corresponding mean absolute error (L_1), mean square error (L_2), and coefficient of determination (R^2) using Eqs. 3.1-3.3 presented in Section 3.3.5. Fig. 5.10a and Fig. 5.10b illustrate the calibration scatterplots for the geology-based zonation and geostatistical models, respectively. In each scatterplot, the 45° line that represents the best match of simulated and observed water-level variations is plotted as a black solid line, and a linear model that fits all data points is provided and illustrated as a red dash line along with the corresponding coefficient of determination (R^2) value, as well as calculated L_1 and L_2 norms.

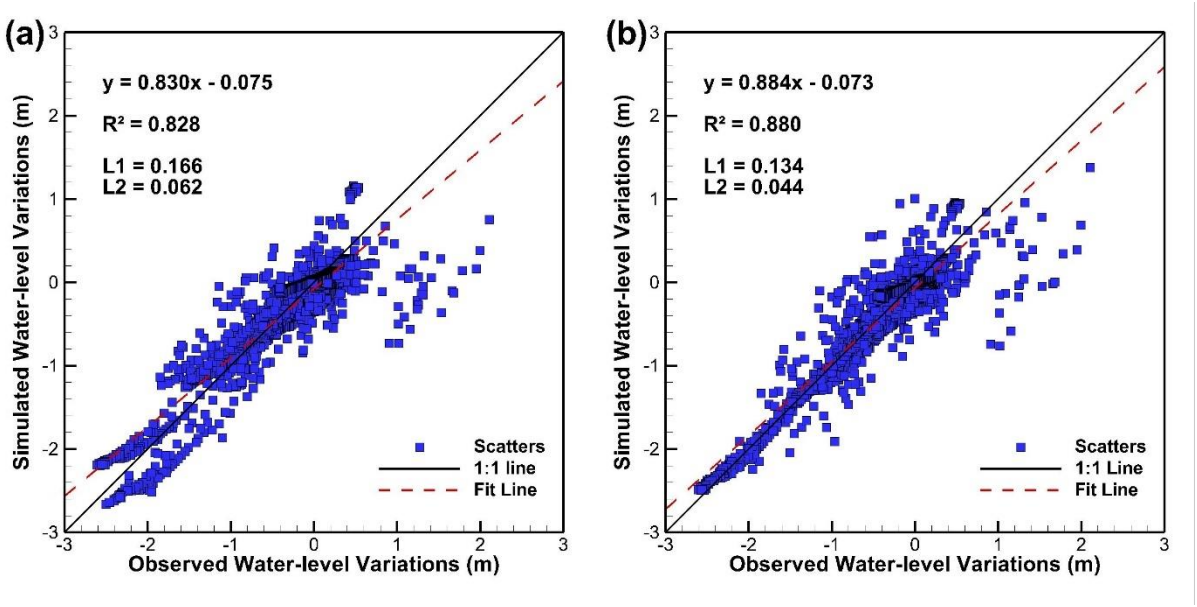

Fig. 5.10: Calibration scatterplots of the simulated versus observed water-level variations. (a) geology-based zonation model, (b) geostatistical model.

These scatterplots reveal that the geostatistical interpretation of the municipal well records yields improved calibration results in comparison to the geology-based zonation model in terms of the computed statistical properties (L_1, L_2, and R^2) as well as the linear model fit to all data points. This is reasonable since only eight sets of K and S_s are estimated for subsurface heterogeneity characterization using the zonation model, while the highly parameterized geostatistical model adjusts K and S_s estimates in each element to fit the observation data. However, several scattered points with significant bias (simulated water-level variations are much smaller than the observed ones) are observed for both the geology-based zonation and geostatistical models. This may be attributed to the failure in capturing rapid changes of water levels using both models, which will be discussed in the following section.

To further assess the calibration results, transient records of simulated and observed water-level variations in each observation wells are plotted and illustrated in Fig. D10 and Fig. D11

for the zonation and geostatistical models, respectively. In general, both models are found to be able to capture the overall behavior of water-level variations at all monitoring wells, especially for the wells located within the production area screened in the water-supply aquifer (AFB2). In comparison to the geology-based zonation model, the geostatistical model yields significantly improved matching of simulated versus observed water-level variations at monitoring wells ow6a through ow13a, ow16b, and ow19. All of these wells are screened in AFB2, and such improvement emphasizes the inherent heterogeneous nature of this water-supply aquifer. On the other hand, both models fail to capture the instantaneous changes of water-level variations with large magnitudes at some monitoring wells (i.e., ow8a, ow9a, and ow11). This may be attributed to the following reasons: (1) daily pumped/injected volume of water was treated as daily flow rates and utilized for model calibration, which may smooth out some short-term (i.e., several hours) operations of production wells with large flow rates, and (2) there is a high K pathway between the observation well and adjacent production wells, which might not be revealed through the analyses presented in this study due to the lack of available data (i.e., small scale features) between production and observation wells

5.4.3 Model Validation

The estimated K and S_s tomograms from the geology-based zonation and geostatistical models were then used to predict municipal well records that have not been utilized for the calibration effort (Scenario 1) and drawdown data from dedicated pumping tests conducted within the wellfield (Scenario 2) for model validation.

For Scenario 1, water-level variations at eight monitoring wells during the year of 2013 were simulated and compared to the observed ones. As mentioned previously, these wells were selected to validate the reliability of estimated K and S_s values at different locations within the simulation domain (i.e., within the production area, in the upper and lower geological units, and away from the production area). Fig. 5.11a and Fig. 5.11b illustrate the scatterplots of simulated versus observed water-level variations for the geology-based zonation and geostatistical models, respectively. In each scatterplot, a linear model that fits all data points was provided along with the corresponding R^2 value as well as the computed L_1 and L_2 norms. In comparison to the geology-based zonation model, the geostatistical inverse model yields significantly improved validation results with less overall scatter, less bias (i.e., slope of the linear model), higher correlation (R^2), and smaller discrepancy (L_1 and L_2) between simulated and observed water-level variations.

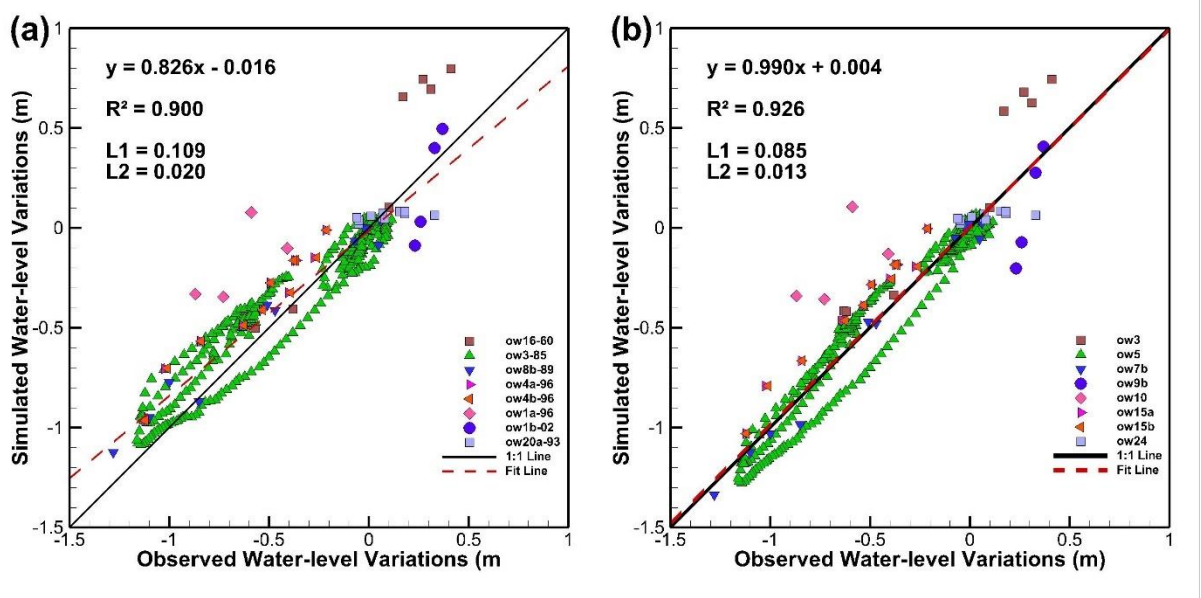

Fig. 5.11: Validation scatterplots of the simulated versus observed water-level variations for Scenario 1. (a) geology-based zonation model, (b) geostatistical model.

However, it should be noted that the improvement is different from one observation well to another depending on their locations within the domain. Fig. D12 illustrates the simulated and observed temporal water-level variations at individual monitoring wells, while Table D4 summarizes the statistical properties (L_1 and L_2 norms as well as slope, intercept and R^2 of the linear model) of the comparison results. Examination of Fig. D12 and Table D4 reveals that, with revealed heterogeneity details of K and S_s, the geostatistical inverse model yields improved prediction of water-level variations at monitoring wells located within the production area and screened in AFB2 (ow3, ow5, ow7b, and ow15ab). However, such improvement is not observed for the wells screened in the upper and lower geological units (ow9b in AFB1 and ow10 in Bedrock), as well as the one located away from the production area (ow24). Due to the lack of hydraulic response data, K and S_s estimates in these areas obtained from the geostatistical inverse model remain almost the same as their initial guesses (the calibrated geology-based zonation model, shown as Fig. 5.9), resulting in similar prediction results at these well locations for the zonation and geostatistical inverse models.

For Scenario 2, drawdown data obtained from three dedicated pumping tests conducted within the Mannheim wellfield were utilized as independent pumping test data for model validation, and Fig. 5.12 illustrate the validation results. Specifically, Fig. 5.12a shows the scatterplots of simulated versus observed drawdowns for the geology-based zonation and geostatistical inverse models, while Fig. 5.12b illustrates the corresponding transient drawdown curves. Comparison results reveal that the geostatistical inverse model also yields better validation results regarding independent pumping test data than the geology-based

zonation inverse model. Specifically, with revealed heterogeneity details within the production area, the geostatistical inverse model provides better prediction of drawdown data obtained from pumping tests conducted at P1 and P2 with smaller differences (L_1 and L_2 norms) and less bias in comparison to the zonation-based inverse model, while the predicted drawdown curves relating to the pumping test conducted at ow20 are almost the same for two investigated models.

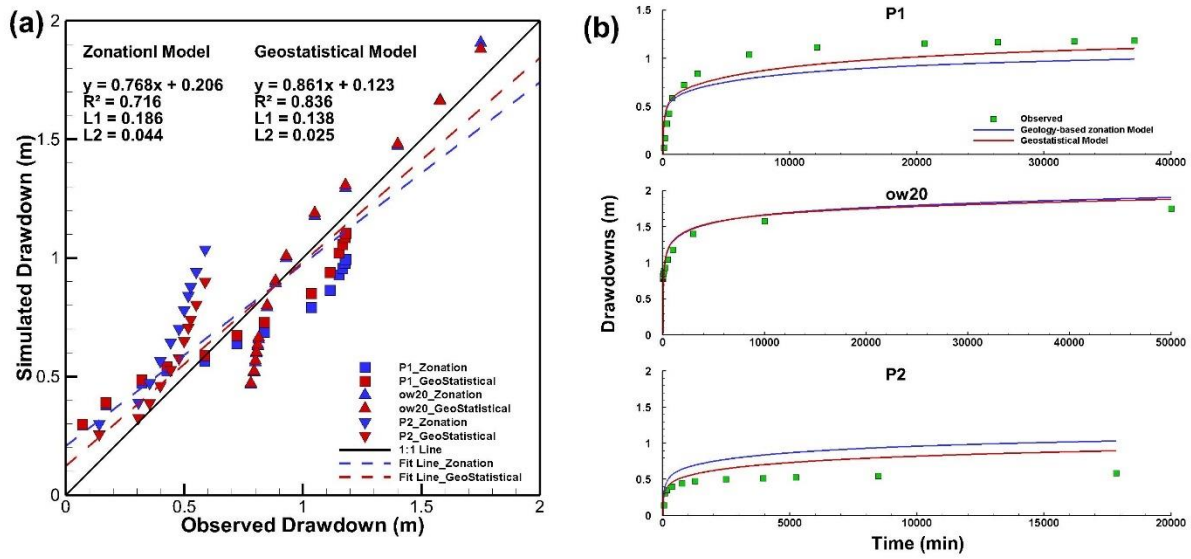

Fig. 5.12: Validation results using dedicated pumping test data (Scenario 2). (a) shows the scatterplots of simulated versus observed drawdowns for the geology-based zonation and geostatistical models, while (b) shows the comparison of simulated and observed drawdown curves for both models.

Although the K and S_s tomograms obtained from the geostatistical inverse model can be used to adequately predict drawdown curves from dedicated pumping tests, biased estimates of drawdown data are also observed at early, intermediate, and late times for ow20, P1, and P2, respectively, as shown in Fig. 5.12a. This may be attributed to the following reasons: (1) these observed drawdown data were obtained through the digitization of existing drawdown curves (as shown in Fig. D6), which might introduce errors to early-time data; (2) the extracted

observation data for model calibration are insensitive to S_s, resulting in estimated S_s values that still involve larger uncertainty; and (3) the distance between observation and corresponding pumping wells ranges from 9 to 21 m for these dedicated pumping tests (as summarized in Table D2), while heterogeneity details in such small scales might not be well characterized due to the large-scale heterogeneity patterns revealed through the analyses presented in this study.

5.5 Discussion

5.5.1 Utilization of Municipal Well Records for Site Heterogeneity Characterization

This study illustrates that the existing hydrographs in a municipal wellfield that have been affected by multiple production wells at variable flow rates can be utilized for subsurface heterogeneity characterization using the approach of THT. Such municipal well records behave similarly to the tomographic snapshots obtained through traditional HT surveys. Specifically, the temporally variable pumping/injection flow rates at multiple municipal water-supply wells alter groundwater flow field frequently, yielding highly fluctuated water levels at monitoring wells that are considered to carry non-redundant information about the heterogeneity in hydraulic properties (e.g., K and S_s). Superior to traditional HT surveys for field applications (e.g., Zhao and Illman, 2017; 2018), the analysis of the existing municipal well records provides a cost-effective characterization of subsurface heterogeneity without the requirement of additional dedicated cross-hole pumping tests. Furthermore, the long-term operation of multiple production wells stresses the aquifer more intensively and propagates the pressure

signal to a greater area compared to dedicated pumping tests, making it possible to reveal large-scale heterogeneity patterns through geostatistical interpretation of municipal well records.

On the other hand, municipal well records are typically abundant due to the continuous operation of water-supply wells and the continuous recording of water-level variations at monitoring wells. Since the joint interpretation of all data points is impractical for most inversion approaches, especially for highly parameterized geostatistical inversions, a question remains on how to select certain observation periods for model calibration. The effect of data selection on inverse modeling when adopting municipal well records for subsurface heterogeneity characterization has been investigated through the synthetic study (Chapter 4) by manually selecting datasets with different durations and periods for model calibration. Following the work presented in Chapter 4, continuous water-level records with a large magnitude of water-level variations is selected and utilized for model calibration in this study (as shown in Fig. 5.4a). The utilization of such datasets is found to yield reliable K estimates, where hydraulic response data are concentrated; however, estimated S_s values still involve great uncertainty. Furthermore, the characteristics of existing hydrographs might be quite different from one wellfield to another depending on the designed pumping/injection regime, rendering the previously proposed sampling strategy still to be restrictive for general application. As a result, new approaches need to be developed for intelligent data selection for inverse modeling, perhaps on the basis of the sensitivity of hydraulic parameters to head response data.

5.5.2 *Performance of two Modeling Approaches and Its Implications*

The geology-based zonation model is commonly adopted for groundwater flow modeling at large scales (e.g., regional or basin scales). This is because lithology information is typically collected at various locations during site characterization efforts, which can be used to construct a conceptual hydrostratigraphic model of the investigated site that shows main heterogeneity features with contrasting hydraulic properties (aquifers versus aquitards). Also, the calibration of a geology-based zonation model is more computationally efficient with a reduced number of parameters to be estimated via inverse modeling for subsurface heterogeneity characterization. However, attention should be paid in obtaining accurate maps of geological structure since the zonation model based on poorly identified geological structures may worsen inverse modeling results.

As mentioned previously, the constructed geological model for this study may still involve uncertainties regarding the interpolated layer boundaries due to the medium to low reliability of most borehole logs. To overcome this issue, additional surveys (i.e., permeameter analysis of core samples, geophysical survey, and so on) are suggested to verify, refine, or update the layer information interpolated from borehole logs. Furthermore, mathematical approaches are also available to reduce the uncertainty of interpolated geological models by jointly updating layer information and hydraulic parameters (zonation model) on the basis of residuals between simulated and observed hydraulic heads (e.g., Lu et al., 2018).

In this study, the geology-based zonation model treats each identified geological unit as a

homogeneous/isotropic medium, and the estimated K and S_s values are representative of averaged hydraulic properties in each unit. With sufficient hydraulic response data, reliable K and S_s estimates are obtained for the water-supply aquifer (AFB2); however, the estimated K and S_s for the layers with few observation data are close to the initial guesses with relative high uncertainties. The calibrated geology-based zonation model is found to be capable in capturing the overall behavior of water-level variations at all monitoring wells screened in AFB2, but yields biased calibration and validation results regarding individual wells due to the neglect of inherent intralayer heterogeneous of hydraulic parameters in each geological unit. On the other hand, the neglect of intralayer heterogeneity could lead to significantly biased predictions of solute transport in subsurface condition, which have been evaluated to be more sensitive to the spatial distribution of hydraulic parameters in comparison to groundwater flow (Luo et al., 2020). Such drawbacks render the geology-based zonation inverse modeling approach to be still restrictive in building accurate groundwater models, even for the case when sufficient hydraulic response data are available to yield reliable estimates of averaged hydraulic parameters for each identified geological units.

Nevertheless, the calibrated geology-based zonation model can be used as initial guesses of hydraulic parameters for highly parameterized geostatistical inversion of head data for subsurface heterogeneity characterization. Such prior geological information is also of critical importance for geostatistical inversions with limited number of pumping tests and sparsely distributed observation wells, which is normally the case for field applications of HT. With prior geological information, the geostatistical inversion of municipal well records in this study

reveals great details of intralayer heterogeneity in the water-supply aquifer (AFB2) where sufficient hydraulic response data are available for model calibration, yielding significantly improved calibration and validation results in comparison to at geology-based zonation model. A case of geostatistical inversion without prior geological information is not performed in this study; however, relevant comparison results that show the importance of using geological data as prior information for geostatistical interpretation of municipal well records can be found in the synthetic study presented in Chapter 4. Furthermore, the geostatistical inverse modeling approach adopted in this study provides an estimate of the uncertainty associated with the estimated K and S_s values on the basis of the computed ln$K$ and lnS_s variance maps. Other than confirming the reliable K and S_s estimates in the area with sufficient hydraulic response data, such variance maps can also be used to guide the collection of additional head data where needed to improve inverse modeling results.

Overall, the characterization results from two modeling approaches suggest that the geology-based zonation model with well identified geological structures can be used as the first step for subsurface heterogeneity characterization, while geostatistical inverse modeling is advocated as the second step to reveal intralayer heterogeneity details of hydraulic properties to achieve more robust site characterization results.

5.5.3 Uncertainties

In this study, the effect of uncertain initial and lateral boundary conditions on inverse modeling is minimized based on the proposed data processing and analysis strategies. However,

uncertainties associated with model conceptualization are still observed on the basis of simulated and observed water-level variations from model calibration and validation for both modeling approaches (as shown in Fig. D10 through Fig. D12).

First, surface water features (e.g., ponds and streams) were not accounted for the presented analyses, and the entire simulation domain was assumed to be fully saturated throughout the simulation period. This simplification was considered to have a minor effect on our inverse modeling results as most observation data were obtained from the water-supply aquifer (AFB2), which is located 20-50 m below ground surface. However, it leads to relatively poor assimilation and prediction of head response data at monitoring wells (ow9b, ow13b, and ow17b) screened in AFB1, which is actually variably saturated with water levels affected by recharge from surface water features or directly from precipitation. This implies that an integrated hydrologic model that accounts for both surface and ground water flow as well as their interactions is required when analyzing long-term water-level variations from shallow aquifers for subsurface heterogeneity characterization. However, the use of available integrated hydrology models such as HGS for HT analysis is still computationally intractable for large scale problems such as the one that is under consideration in this study.

On the other hand, the poor fits of simulated water-level variations to observation data regarding model calibration and validation are also observed for the monitoring wells screened in lower geological units (ow14a, ow16a, and ow10) and for those located away from the production area (ow20, ow23, and ow24). One potential reason is that these observation wells

were subjected to manual measurement of water levels, which resulted in few observation data that fail to represent the overall water-level variation patterns and were insufficient for revealing heterogeneity details through geostatistical inverse modeling. Another more likely reason is that water-level variations in these observation wells might be mainly affected by some unidentified stimulus other than the pumping/injection operations of municipal water-supply wells, leading to poor fitting of simulated head data to observed data even when highly parameterized geostatistical inverse model was adopted for model calibration. As a result, continuous recording of water-level variations in all existing observation wells along with fully identified stimuli within the simulation domain are suggested to yield improved heterogeneity characterization results for the entire simulation domain.

6 Study IV: Three-dimensional Steady-state Hydraulic Tomography Analysis with Integration of Cross-hole Flowmeter Data at a Highly Heterogeneous Site

6.1 Site and Data Description

6.1.1 Site Geology and Instrumentation

Data for this study were obtained from the NCRS, located at the University of Waterloo (U.W.), in Waterloo, Ontario, Canada. The NCRS site has been developed and utilized for HT-related studies since 2010 (Alexander et al., 2011), and abundant types of data (i.e., laboratory core sample analyses, geophysical data, steady-state/transient head data, flowmeter measurements, and HPT data) were collected at the site for subsurface heterogeneity characterization. Here, we focus on estimating spatial K distribution at the NCRS through geostatistical inverse analyses of head and cross-hole flowmeter data collected under steady-state conditions with/without prior geological information. The characterization domain is 45 m × 45 m ×17 m in X, Y, and Z directions, respectively, with a 15 m by 15 m square well site located in the center of the domain, as shown in Fig. 6.1a.

Quaternary geology investigations showed that the NCRS is located on the Waterloo Moraine (Karrow, 1979; 1993), formed by interlobate glacial activity during the late Wisconsinan glaciation stage (Martin and Frind, 1998). Numerous advances and retreats of ice lobes had resulted in highly heterogeneous glaciofluvial deposits beneath the study area, which consists of multiple glacial till layers mixed with glacial outwash sands and gravels (Alexander et al., 2011; Karrow, 1979). The K heterogeneities in both horizontal and vertical directions are apparent through laboratory permeameter tests of core samples (Alexander et al., 2011; Zhao

and Illman, 2017), which yielded K values ranging from 10^{-10} to 10^{-3} m/s with a ln$K$ variance as high as 8.2.

Based on the lithology information at 18 well locations and permeameter test results from nine boreholes, Zhao and Illman (2017) developed a 19-layer geological model, describing the stratigraphic information beneath the study area (Fig. E1 in Appendix E). The dimensions of the geological model are 70 m, 70 m, and 17 m in X, Y, and Z directions, respectively, with the square well site located in the center of the model. The constructed geological model shows two highly permeable zones composed of sand to sandy gravel that behave as major pathways for groundwater flow at the site. Examination of Fig. E1 also reveals that the interpolated geological layers exhibit a kame feature that is discontinuous within the domain. This model is consistent with previous geological investigations conducted within the Waterloo Moraine, in which significantly eroded glacial tills were reported to be well mixed with glaciofluvial deposits (i.e., silt, sand, and gravel) (Karrow, 1993). This study incorporated 19 geological layers beneath the study area as prior structural information for geostatistical inversions of cross-hole flowmeter and/or steady-state head response data.

As shown in Fig. 6.1b, the 15 m × 15 m wellfield includes nine wells with different designs. Pumping wells PW1, PW3, and PW5 are three 4-inch multi-screen wells; PW1 is screened at seven different depths, and PW3 and PW5 at five different depths. In each well, the distance between 1-m long screened sections is 1 m, and a sand pack surrounds each screened interval. Pumping wells, PW2s and PW4s, are two well nests, each consisting of three 2-inch wells drilled to different depths with 1-m long screen at the bottom in direct contact with the native formation. In addition to these pumping wells, four Continuous Multichannel Tubing (CMT) wells are located at the mid-point of square site edges and utilized as water-level monitoring

wells (CMT1 through CMT4). Each CMT well contains seven monitoring channels with 0.2-m long screens spaced approximately 2-m apart. Details about the construction of these wells are available in Alexander et al. (2011) and Berg and Illman (2011b). Additional single-screen wells have also been drilled outside the square well site for future HT-related studies; however, they are excluded from the presented analysis since neither head nor flowmeter data is available at newly constructed wells.

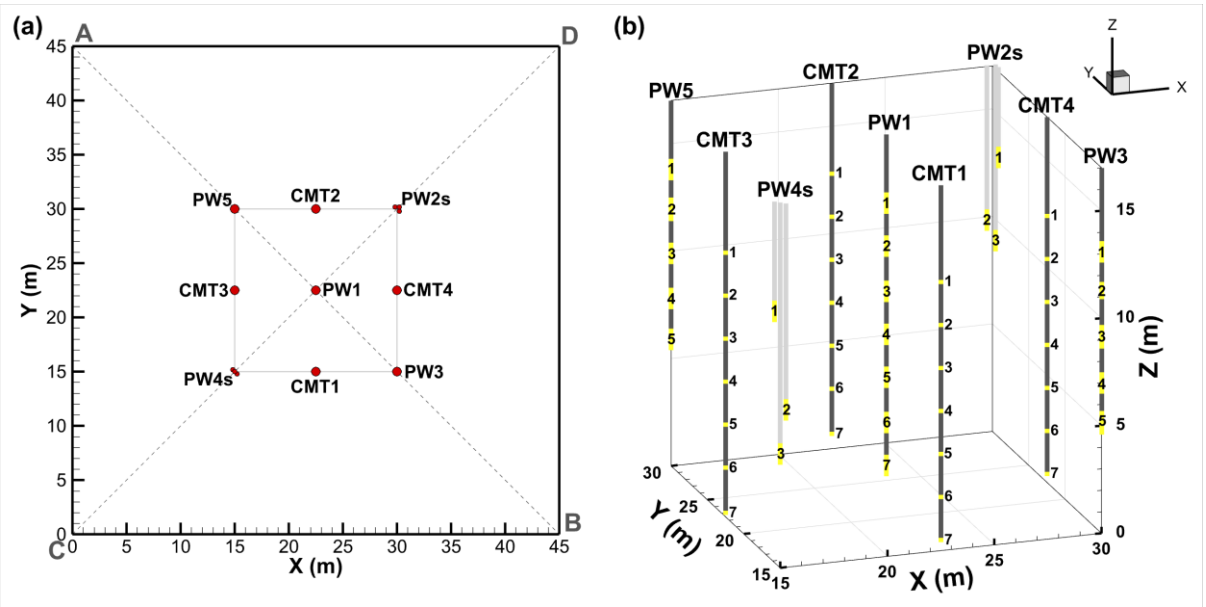

Fig. 6.1: Instrumentation for pumping/injection tests and flowmeter surveys at the NCRS: (a) plan view of simulation domain and well distribution; (b) three-dimensional perspective view of well configurations. Diagonal dashed lines on Figure 1a indicate locations of cross-sections A-B and C-D.

6.1.2 Cross-hole Flowmeter Measurements

At the NCRS, two cross-hole flowmeter surveys were performed at five pumping wells (PW1, PW2s, PW3, PW4s, and PW5). Specifically, we conducted steady-state pumping at PW1-3 (PW1 interval #3) or PW5-3 (PW5 interval #3), while measuring vertical flow inside other wells. Prior to the pumping tests, vertical flow in individual wells under ambient condition were measured. As pointed out by Molz et al. (1994), the measured total flow needs

to be subtracted from the ambient flow to yield pump-induced flow logs for *K* estimation. These flowmeter data were scrutinized prior to their inclusion in geostatistical inversions. Specifically, an initial analysis of flowmeter data was performed to examine flow logs under ambient and steady-state pumping conditions and compare them directly to the lithology at each well location.

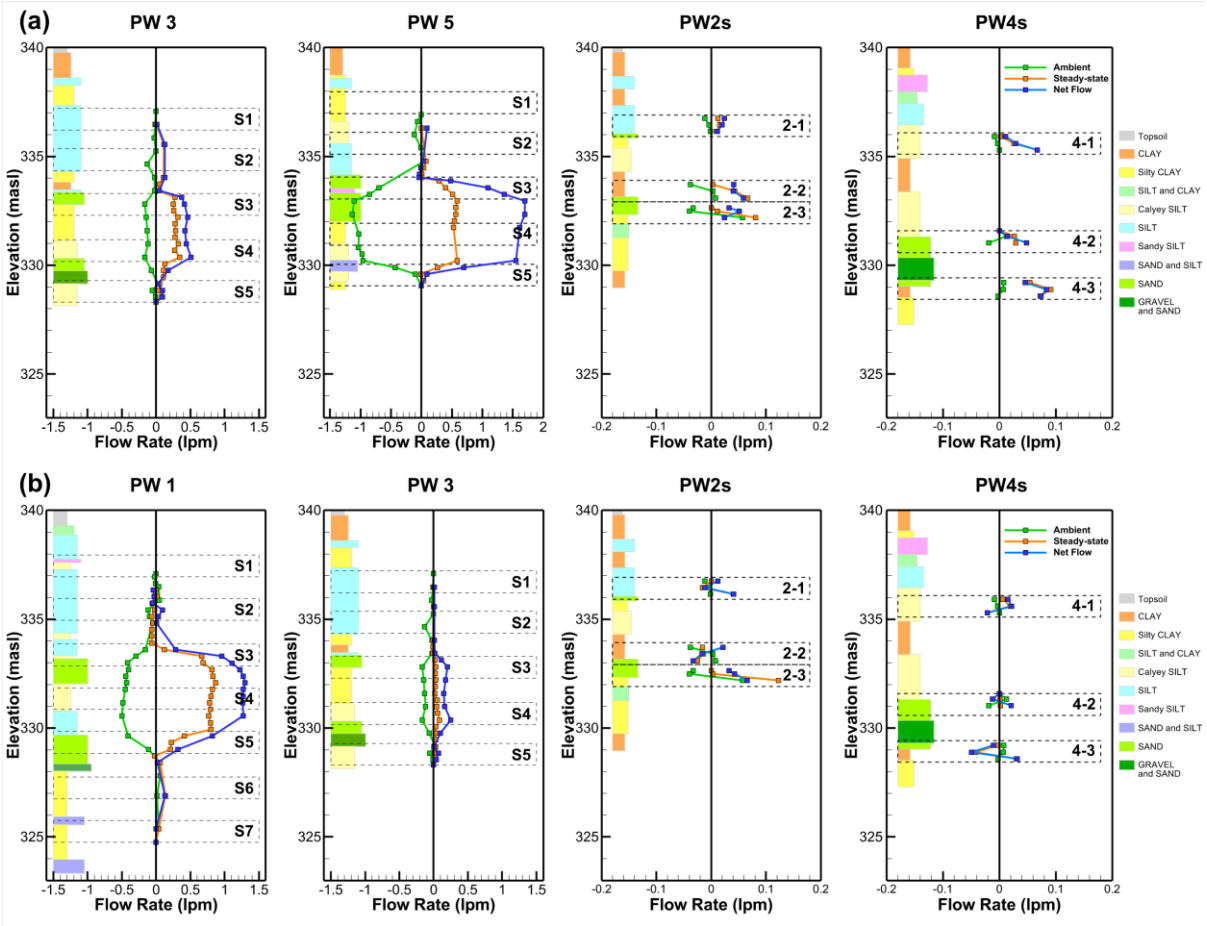

Fig. 6.2: Cross-hole flowmeter measurements at observation wells when pumping at (a) PW1-3 and (b) PW5-3. Negative and positive values indicate groundwater flow downward and upward, respectively. Green curves show the measured flow logs under the ambient condition, orange curves show the measured flow logs under steady-state pumping conditions, and blue curves show the computed net flow indicating the vertical flow induced by pumping only.

Fig. 6.2 summarizes cross-hole flowmeter measurements at the site. Fig. 6.2a illustrates the measured flow logs along PW3, PW5, PW2s, and PW4s when pumping at PW1-3 reached a steady-state condition, while Fig. 6.2b shows the measured flow logs along PW1, PW3, PW2s, and PW4s with steady-state pumping at PW5-3. In each sub-figure, vertical flow logs within

the borehole under ambient (green curve) and steady-state pumping (orange curve) conditions are plotted along with the net flow (blue curve) that represents the vertical flow induced by pumping only. Net flow rates were only computed at depths where both ambient and induced flow measurements are available. The well screen intervals are dashed rectangles and the lithology along each well are illustrated as color bars on the left. These figures reveal that the measured ambient flow is consistent with the lithology along each well. The natural hydraulic gradient in the vertical direction drives groundwater into the borehole through the upper high *K* zone, then moves it down and forces it to leave the borehole in the lower high *K* zone. However, the magnitudes of the measured ambient flow differ from one well to another, manifesting the highly heterogeneous nature of the NCRS.

Fig. 6.2 reveals that the computed net flow (N.F.) logs at multilevel-screened wells (PW1, PW3, and PW5) also correspond well to the lithology. When conducting steady-state pumping at PW1-3/PW5-3, we found that groundwater mainly flowed into multilevel-screened wells via the lower high *K* zone, then moved upward and leaved the borehole in the upper high *K* zone. On the other hand, cross-hole flowmeter measurements at single-screen wells in two well nests (PW2s and PW4s) are very small (less than 0.1 LPM) compared to those taken from multilevel-screened wells during the two steady-state pumping tests. This implies that the pumping-induced groundwater flow at these single-screen wells may not be dominant in the vertical direction, causing difficulties to inverse modeling of cross-hole flowmeter data (details are discussed later). Furthermore, such small values may be within measurement errors of flowmeter used at the NCRS and may involve great uncertainty. Consequently, cross-hole flowmeter data from these two well nests' single-screen wells were excluded from the analysis.

As a result, cross-hole flowmeter data from multilevel-screened wells (PW1, PW3, and PW5) under the two steady-state pumping conditions (PW1-3 and PW5-3) were selected and interpreted for 3-D characterization of *K* heterogeneity. The measured vertical flow rates were

converted to fluxes based on the cross-section area of the 4-inch wellbore, and one flux data was extracted from each screen/casing interval and included for inverse modeling. The resulting flux profiles along multilevel-screened wells during two steady-state pumping tests are illustrated in Fig. 6.4. In total, 35 flux data were selected and subjected to inverse modeling for the presented analysis. Albeit cross-hole flowmeter data have been utilized to define hydraulic connections in fractured rocks (e.g., Williams and Paillet, 2002), this study is the first that uses such datasets and an inverse approach (SimSLE) (Xiang et al., 2009) to characterize the heterogeneity in porous media.

6.1.3 Steady-state Head Data

Previous studies (Berg and Illman, 2011b, 2013; Zhao and Illman, 2017, 2018) conducted 15 pumping/injection tests at various locations at the site and utilized them for characterizing subsurface heterogeneity. Different from cross-hole flowmeter tests in which multilevel-screened wells not used for pumping activities acted as vertical conduits for groundwater flow, individual screens in multilevel-screened wells were isolated using FLUTe systems (FLUTe Ltd.) and head responses were collected at many depths at the site during each test. This study takes advantage of their existing data sets without conducting new pumping/injection tests. For the presented analyses, we used steady-state head data from 14 pumping/injection tests, as indicated in Table 6.1. Specifically, the results of their seven tests (PW1-1, PW1-4, PW1-6, PW2-3, PW3-3, PW4-3, and PW5-3) were utilized for inverse modeling in the following sections. The remaining seven tests (PW1-3, PW1-5, PW1-7, PW3-1, PW3-4, PW5-4. and PW5-5) were treated as independent pumping/injection tests and utilized for validating the estimated K fields from this study. The injection test performed at PW5-1 is excluded since quasi-steady state was not achieved due to the low flow rate and short survey period.

We followed the work of Zhao and Illman (2017) to select only late time head data from

ports indicating steady or quasi-steady state for the presented analysis. In total, 195 head data were selected for inverse modeling and 199 head data for model validation. Different from Zhao and Illman (2017), this study integrated the newly collected cross-hole flowmeter data as additional datasets with HT analysis for *K* heterogeneity at the NCRS.

Table 6.1: Summary of pumping/injection tests conducted at the NCRS.

Operation Well	Type	Flow Rate (lpm)	Duration (h)
PW1-1*	Injection	1.89	4.5
PW1-3*	Pumping	10.50	6.0
PW1-4*	Pumping	6.30	8.5
PW1-5*	Pumping	4.40	22.5
PW1-6*	Pumping	0.95	6.5
PW1-7*	Pumping	1.05	26.5
PW2-3*	Pumping	1.91	7.0
PW3-1*	Injection	0.94	4.4
PW3-3*	Pumping	2.10	22.0
PW3-4*	Pumping	1.50	22.0
PW4-3*	Pumping	30.20	22.5
PW5-1	Injection	0.85	4.5
PW5-3*	Pumping	7.80	22.0
PW5-4*	Pumping	7.80	8.5
PW5-5*	Pumping	8.10	22.0

* Indicates the test utilized for analyses presented in this study.

6.2 Data Analysis

6.2.1 Numerical Model

The entire domain for groundwater flow was discretized into 40,425 cuboid elements with 44,064 nodes (Fig. E2). The mesh was refined at well locations, yielding element size ranges from 0.2 m (X) x 0.2 m (Y) x 0.25 m (Z) to 5 m (X) x 5 m (Y) x 0.7 m (Z). Following the work of Zhao and Illman (2017), the top and bottom boundaries were set as no-flow, while the four lateral boundaries were kept as constant heads during the simulation. Steady-state groundwater

flow in a three-dimensional, saturated, porous medium is described using Eqs 1.1 and 1.3 in Section 1.1. This study solves the steady-state groundwater flow equation using a 3-D finite element model VSAFT3 (Yeh et al., 1993) to simulate head data and specific discharge under steady-state conditions.

Using VSAFT3, this study constructed numerical boreholes to simulate cross-hole flowmeter tests. Specifically, a series of elements corresponding to multilevel-screened wells (PW1, PW3, and PW5) were selected, and specific K values were assigned to these elements to represent vertical conduits for groundwater flow or impermeable cased intervals between screens. Fig. E3 illustrates the conceptualization of boreholes in the numerical model. A previous HT study at the NCRS (Zhao and Illman, 2017) yielded K estimates, ranging from 1×10^{-8} m/s to 1×10^{-4} m/s for the geologic material at the site. As a result, we assigned the K value of 1×10^{-2} m/s to the elements that represent vertical conduits for groundwater flow and the K value of 1×10^{-10} m/s to elements representing cased intervals. These values avoid possible computational issues arising from significant K-value contrasts during inverse modeling. These K values for the numerical boreholes were fixed during inversions of cross-hole flowmeter data.

The representativeness of the numerical boreholes was evaluated by the forward simulation of cross-hole flowmeter tests using the best K tomogram estimated from previous SSHT study at the NCRS (Case 3d in the work of Zhao and Illman (2017)). Simulated flow logs were compared to the observed ones, and Fig. E4 shows the comparison results. Despite the simulated vertical fluxes being smaller than those measured, the simulated and observed flow logs are comparable in terms of their variation patterns along the borehole. These results suggest that this borehole conceptualization is adequate for cross-hole flowmeter data simulation at the NCRS. On the other hand, the developed flux inversion algorithms (Zha et

al., 2014) account only flux magnitudes or all flux vectors in three dimensions (X, Y, and Z) for *K* heterogeneity characterization. The second option is not valid in this study since cross-hole flowmeter measurements only represent the vertical flow rates, and we were unable to derive horizontal flow components from those measurements. As a result, we treated the simulated flux magnitudes along numerical boreholes approximately equal to the measured vertical fluxes during cross-hole flowmeter surveys. When adopting numerical boreholes, such approximation was found to be valid at multilevel-screened wells where vertical flow is dominant, but not for the single-screen wells (PW2s and PW4s) where vertical and horizontal fluxes were simulated to have the same orders of magnitude. Thus, cross-hole flowmeter measurements at single-screen wells were excluded for this study.

6.2.2 *Inverse Modeling*

Characterizing the 3-D *K* distribution at the NCRS, using head and/or cross-hole flowmeter data, was carried out using the Simultaneous Successive Linear Estimator (SimSLE). SimSLE was originally developed by Xiang et al. (2009) and modified later by Zha et al. (2014) to account for the inversion of flux data for subsurface heterogeneity characterization. When conducting SSHT analysis, SimSLE treats the natural logarithm of *K* (ln*K*) as a multi-Gaussian, stationary, stochastic process. Briefly, SimSLE starts with the estimation of conditional ln*K* field through cokriging analysis of initial *K* estimates and all observation data with prior knowledge of unconditional statistical properties of ln*K* (mean, correlation length λ_x, λ_y, λ_z, and the variance of ln*K*: $\sigma^2_{\ln K}$). The initial ln*K* field is then iteratively updated using the implemented Successive Linear Estimator (SLE) (Yeh et al., 1996) based on the difference between the observed and simulated head and/or flux data as well as the updated error covariance of ln*K* during each iteration. Other than the spatial distribution of *K*, SimSLE also provides uncertainty estimates (i.e., residual variances of ln*K*). Details of the algorithm and the

inversion of flux data are available in Appendix A and Zha et al. (2014), respectively.

For all inversions performed in this study, $\sigma^2_{\ln K}$ was set to be 5.0, while the correlation lengths were $\lambda_x = \lambda_y = 4$ m, and $\lambda_z = 2$ m, the same as those used by previous HT studies (e.g., Berg and Illman, 2011b; Zhao and Illman, 2017). Due to the availability of a great number of head and flowmeter measurements at the NCRS, errors in these prior statistical parameters for HT analysis likely have negligible impacts on characterization results (Yeh and Liu, 2000).

6.2.3 *Prior Information*

Prior geological information was accounted for characterizing spatial *K* distribution at the NCRS. Specifically, a zonation model based on the 19-layer geological model (Fig. E1) was constructed with the mean *K* values in individual layers obtained from permeameter test results (Alexander et al., 2011; Zhao and Illman, 2017). The distributed mean *K* field (Fig. E5) was utilized as the initial *K* field for geostatistical inversions of head and/or flux data. This is different from classical prior geostatistical information that accounts for both prior mean and covariance of hydraulic parameters. In contrast, we assume that large-scale heterogeneity (geological layers) is well characterized at the NCRS with mean *K* values for each layer estimated from permeameter test results, and the inverse modeling focuses on estimating inter-layer *K* heterogeneity. When using SimSLE for HT analysis, such prior geological information has been found to be useful in characterizing spatial distribution of hydraulic parameters with significantly improved heterogeneity details (Luo et al., 2017; 2020; Tso et al., 2016; Zhao and Illman, 2017; 2018). This is because SimSLE recursively updates the mean and covariance of the estimated hydraulic parameters during each iteration, and the posterior mean and covariance from the previous iteration serve as a prior for the next iteration (Zha et al., 2017).

On the other hand, it is important to recognize that the setups of wells at the NCRS were

slightly different during pumping/injection tests for the HT survey and cross-hole flowmeter surveys. Specifically, individual screens in multilevel-screened wells were isolated using FLUTe systems (FLUTe Ltd.) during the HT tests to monitor water-level variations at different depths. The FLUTe systems were removed when conducting the cross-hole flowmeter survey, and multilevel-screened wells not used for pumping (PW3 and PW5 when pumping at PW1-3; PW1 and PW3 when pumping at PW5-3) acted as vertical conduits for groundwater flow. Therefore, geostatistical inversions of flux and steady-state head data were carried out separately using SimSLE. Following the concept of prior geological information, the integration of cross-hole flowmeter data in this study was achieved through adopting the estimated *K* distribution from SimSLE inversion of flux data as the initial *K* field for SimSLE inversion of HT head data. In other words, HT analysis of head data started from the mean *K* field conditioned on the flux data, resulting in the final estimated *K* field conditioned on both flux and head data. A similar procedure was used by Zhao and Illman (2022b) to integrate high-resolution *K* estimates from HPT surveys with HT analysis of head data. In their study, HPT-derived *K* field was utilized as the prior mean *K* field for SimSLE inversion of head data, yielding significantly improved delineation of *K* heterogeneity.

6.2.4 Scenario and Cases

The geostatistical inversions of flux data were performed under two scenarios depending on whether prior geological information is incorporated for inverse modeling or not. Scenario 1 refers to the inversion of flux data only, in which SimSLE starts with a homogeneous field of $K = 8.0 \times 10^{-6}$ m/s. This *K* value was obtained by taking the geometric mean of *K* estimates from transient analyses of drawdown data at individual monitoring ports induced by pumping at PW1-3 (Berg and Illman, 2011b). Scenario 2 incorporated prior geological information, in which, SimSLE started with the distributed mean *K* field (Fig. E5). Due to the existing

numerical boreholes for flowmeter test simulation, the estimated K tomograms from these two scenarios were treated as intermediate characterization results and utilized as initial K fields for SSHT analyses of head response data.

Later, four cases were carried out to characterize spatial K distribution at the NCRS through SSHT analysis of head response data with different prior information (initial K field): Case 1: starting with a homogeneous K field; Case 2: integrating cross-hole flowmeter data; Case 3: prior structural information based on the geological model; and Case 4: integrating cross-hole flowmeter data with prior structural information. Specifically, geostatistical inversions of Cases 1 and 3 started with homogeneous ($K = 8.0 \times 10^{-6}$ m/s) and zonation (Figure S5) K fields, respectively. For Cases 2 and 4, which integrated cross-hole flowmeter data, the initial K fields were the estimated K fields from SimSLE inversions of flux data in Scenarios 1 and 2, respectively. The prediction of steady-state drawdown data from seven independent pumping/injection tests (PW1-3, PW1-5, PW1-7, PW3-1, PW3-4, PW5-4, and PW5-5) then validated the estimated K tomograms from all four cases.

6.3 Results

In the following sections, we first evaluate the inversion results of flux data. Then, characterization results from Cases 1 to 4 with SSHT analyses of head data utilizing different prior information are presented and compared. Subsequently, calibration and validation results from all four cases are assessed by scatterplots of simulated versus observed drawdowns and quantitative analyses of fitting errors (L_1, L_2, and R^2).

6.3.1 Inversion of Flux Data

During each cross-hole flowmeter survey, a packer system was placed to isolate the

pumped interval within the well when conducting steady-state pumping, while the remaining two multilevel-screened wells acted as vertical conduits for groundwater flow. Due to the discrepancy in well configurations during two pumping tests (PW1-3 and PW5-3), we applied the loop-iteration method for geostatistical inversions of flux data. During each loop, the inversion started with the inclusion of the flux data derived from flowmeter measurements when pumping at PW5-3, followed by those measured when pumping at PW1-3. The estimated K tomogram from the previous dataset was utilized as the initial K field for the inversion of the next dataset, and the process continued until the mean square error (L_2) between the simulated and observed vertical fluxes stabilized. This is similar to the Sequential Successive Linear Estimator (SSLE) proposed by Yeh and Liu (2000). However, we inverted two sets of flux data in a loop and did not utilize the estimated residual covariance from the previous dataset as the conditional parameter covariance for the inversion of the next dataset. One reason that we did not use residual parameter covariance is because we adopt numerical boreholes with fixed K values for SimSLE analysis of flux data, and their settings (spatial distribution of fixed K values) are different for the inversions of flux data obtained from two cross-hole flowmeter tests. Fig. E6 shows the changes of L_2 as a function of iteration during the geostatistical inversions of flux data under two scenarios, and the selected iteration for each scenario is labelled as a green dot.

Fig. 6.3a and Fig. 6.3b illustrate the estimated K distributions at two cross-sections (A-B and C-D, as shown in Fig. 6.1a) under Scenarios 1 and 2, respectively. Fig. 6.3a shows that the geostatistical inversion of only the flux data obtained from cross-hole flowmeter surveys revealed heterogeneity patterns of K mainly between the tested wells, yielding an estimated K tomogram that roughly captured the two main aquifers at the NCRS and the low K aquitard between them. After incorporating prior geological information, Fig. 6.3b shows the refinement of inter/intra-layer heterogeneity patterns of K at the site in comparison to the geology-based zonation model (Fig. E5). It is interesting to note that a relatively high K zone was revealed at

the bottom of PW1, which was previously characterized as a low *K* area through HT analyses of steady-state and transient head response data (Zhao and Illman, 2017; 2018). The revealed high *K* zone, however, shows some agreement to the investigated lithology, where relatively high permeable sand and silt deposits are observed at the bottom of PW1 (as shown in Fig. 6.2). This result implies that geostatistical analysis of cross-hole flowmeter measurements with prior geological information is capable in not only refining *K* estimates between wells but also reveal small-scale *K* heterogeneities near wells.

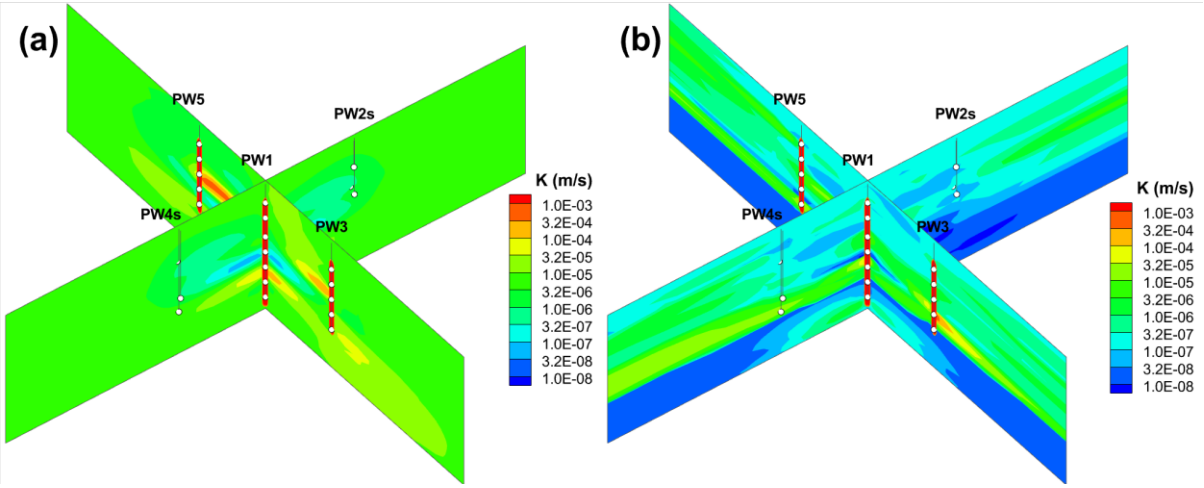

Fig. 6.3: Estimated *K* distributions at two diagonal cross-sections A-B and C-D through geostatistical inverse modeling of cross-hole flowmeter data: (a) without prior geological information; (b) with prior geological information.

To further evaluate the characterization results through geostatistical inversion of flux data, the estimated *K* values within 1 m in diameter of three multilevel-screened wells were extracted from both scenarios and plotted as functions of the elevation (Z), as shown in Fig. E7. This figure reveals the estimated vertical distribution of *K* corresponding well with the lithology at the well locations. However, estimated *K* values without prior geological information are about 1 to 2 orders larger than those estimated with prior geological information. This might be attributed to the fact that the initial *K* field for Scenario 1 was obtained through HT analyses of hydraulic test data while treating the *K* to be homogeneous. The estimated effective *K* field is

more affected by the high *K* zones that promote groundwater flow. In contrast, the utilization of prior geological information (Scenario 2) coupled with laboratory permeameter test results significantly improves estimates of the low *K* zones. The improved characterization of low *K* zones with overall smaller *K* estimates (Scenario 2) also yields better predictions of vertical fluxes within boreholes, as shown in Fig. 6.4. Such comparison results imply that the smaller *K* estimates may be more representative of actual hydraulic conditions at the NCRS.

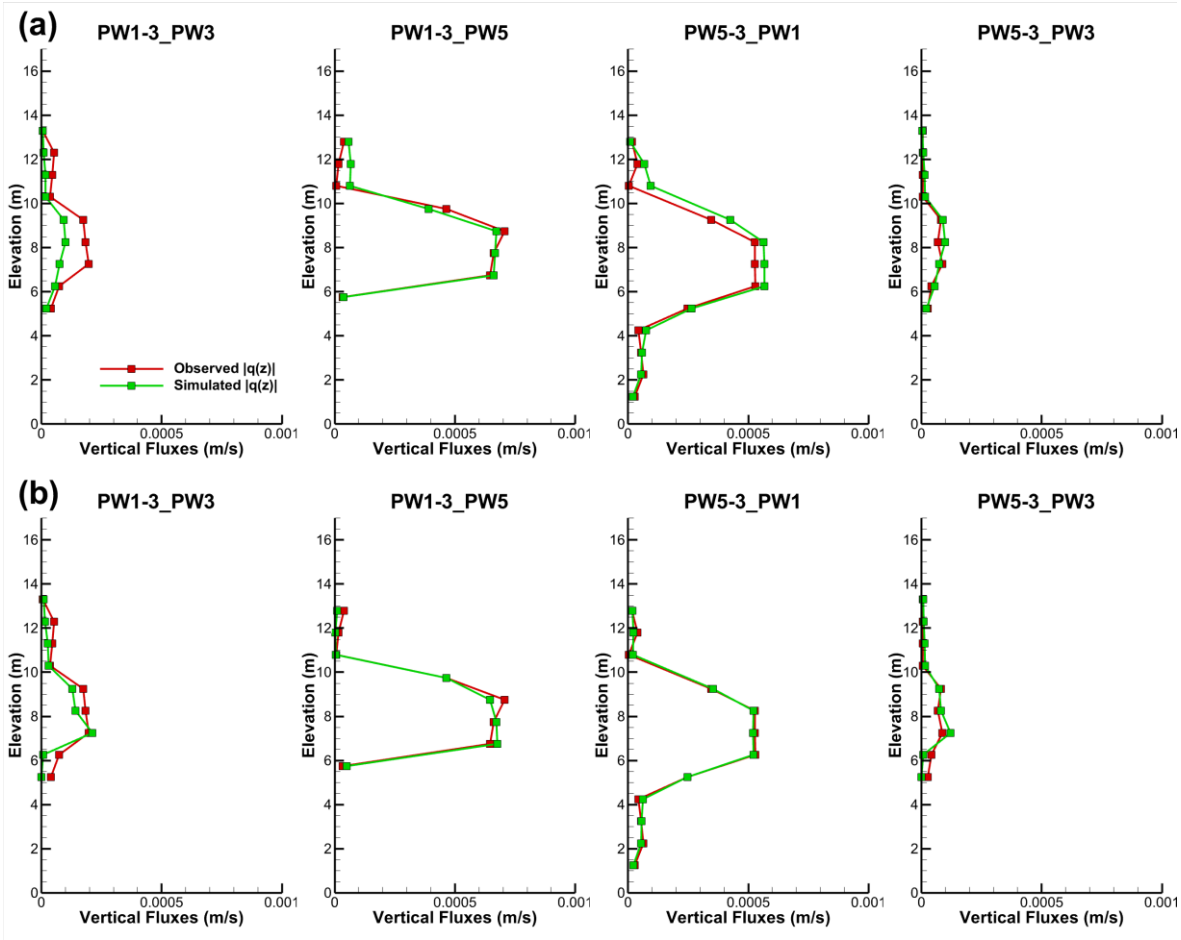

Fig. 6.4: Comparison of the simulated versus observed vertical flux magnitudes at multilevel-screened wells induced by steady-state pumping at PW1-3/PW5-3 using *K* tomograms estimated through geostatistical inverse analysis of cross-hole flowmeter data for (a) Scenario 1 and (b) Scenario 2.

The estimated *K* tomograms through geostatistical inversions of flux data (Scenarios 1 and 2) were subsequently utilized as initial *K* fields for SSHT analyses of head data (Cases 2 and

4). For Scenario 1, K values assigned to the elements representing numerical boreholes were replaced by the mean K value (K = 8.0 x 10^{-6} m/s) of the entire site, resulting in the initial K field for Case 2. For Scenario 2, K values of the numerical boreholes were replaced by the mean K values of corresponding geological units from permeameter tests, yielding the initial K field for Case 4.

6.3.2 *Inversion of Steady-state Head Data*

SSHT analyses of head response data with different initial K fields yield K distributions of varying heterogeneity details. Fig. 6.5 illustrates the estimated K distributions at two diagonal cross-sections (A-B and C-D sections in Figure 1a) from all four cases. The corresponding variation of L_2 norm that indicate the convergence of each case is presented in Fig. E8. Generally, Fig. 6.5 shows that more heterogeneity details about K were revealed when additional types of data were integrated for inverse modeling. In comparison to the inversion of head data only (Case 1), the integration of flux with head data (Case 2) revealed more heterogeneity details about K within the 15 m well square. Specifically, the thickness of the bottom high K zone was narrowed, and an additional low K layer was revealed above the high K zone. Such low K layer indicates the presence of glacial till at an elevation of 326 m to the bottom of the simulation domain (323 m) at the NCRS (see sub-figure for PW1 in Fig. 6.2b), which is not captured through HT analysis of only head response data. After incorporating prior geological information for inverse modeling, Cases 3 and 4 refined K heterogeneity throughout the site. The comparison between these two cases shows that the integration of flux with head data (Case 4) yields clearer delineation of high K zones within the 15 m well square. Furthermore, a relatively high K zone was revealed at the bottom of domain, corresponding to the presence of sand and silt materials within the glacial till based on the lithology at PW1.

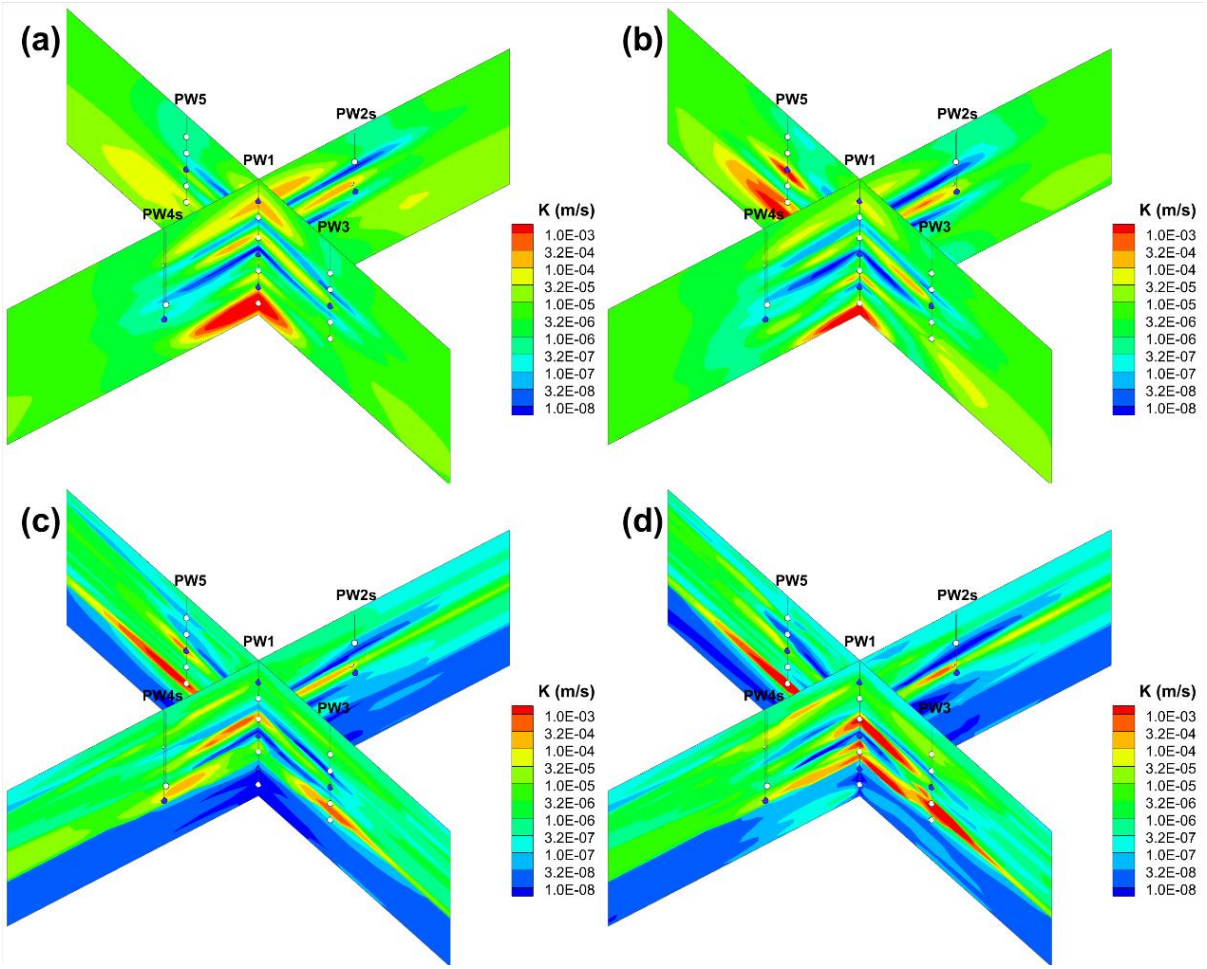

Fig. 6.5: Cross-sections of estimated K tomograms along A-B and C-D through the integration of different types of data for SSHT analysis: (a) head response data (Case 1); (b) head and cross-hole flowmeter data (Case 2); (c) head data with prior geological information (Case 3); (d) head and cross-hole flowmeter data with prior geological information (Case 4).

Estimated K values within the 15 m × 15 m wellfield from four cases are qualitatively and quantitively compared using scatterplots and statistical properties (L_1, L_2, and R^2). Since the true K field at the NCRS is unknown, we choose the estimated K tomogram from Case 4, which shows the greatest details about K heterogeneity among four cases, as the reference field. Instead of evaluating the accuracy of estimated K values, these comparisons aim to show the ability of different cases in revealing K heterogeneity information within the wellfield. To clearly address the effect of flux data on delineating K heterogeneities, the 15 m × 15 m wellfield is subdivide into three zones: 0 - 4 m (323 –327 masl) represents the bottom glacial

till at the NCRS with few flux data at PW1, 4 - 13 m (327 – 336 masl) indicates the muti-aquifer/aquifer system with most flux data, 13 - 17 m (336 –-340 masl) refers to the top aquitard with only one flux data at the top of PW3 during each cross-hole flowmeter test. Fig. 6.6 illustrates the comparison results in individual subdivided zones.

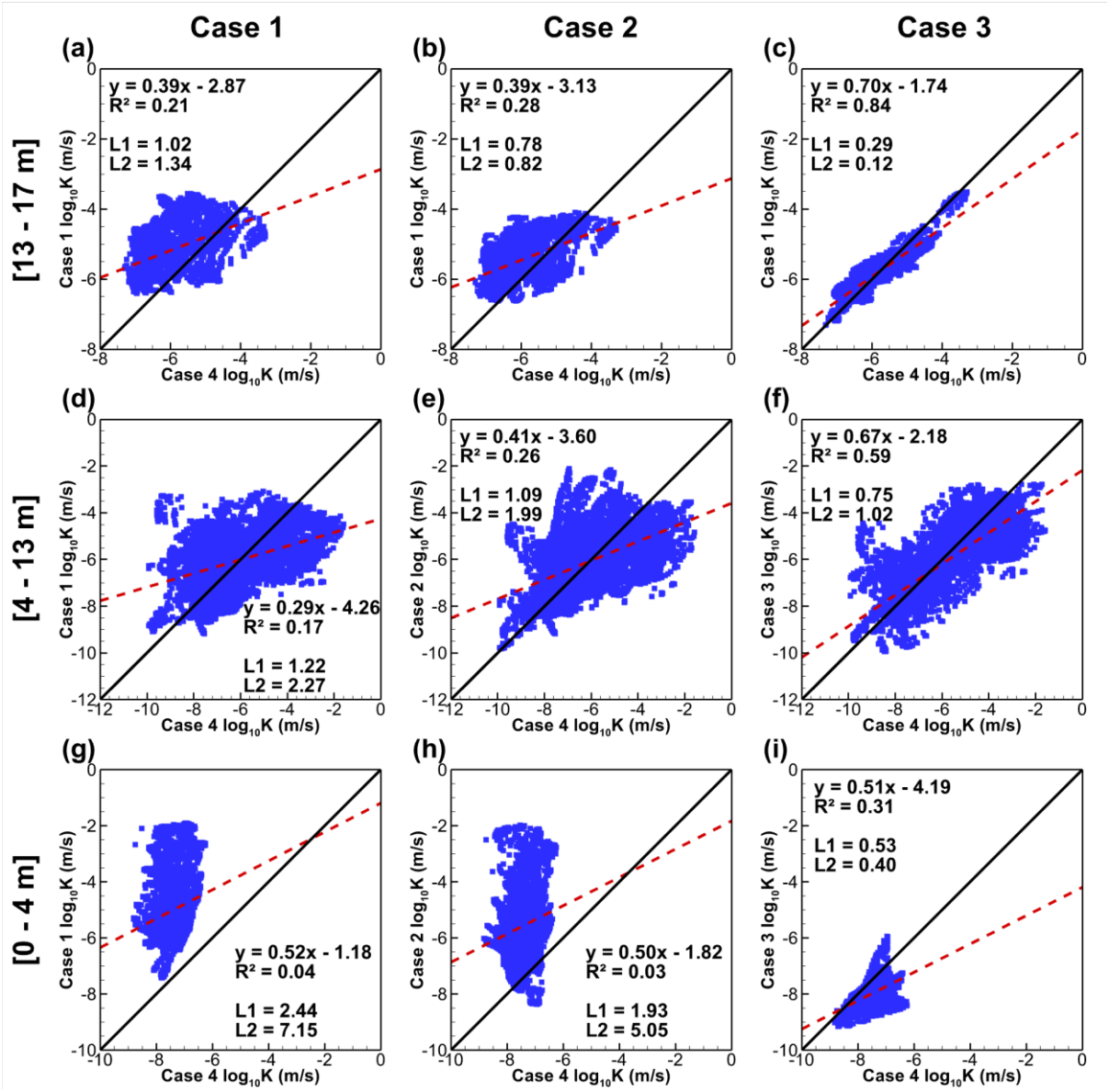

Fig. 6.6: Scatterplots of *K* estimates obtained from Cases 1, 2, 3 versus Case 4 within the 15 m well square. (a) – (c) refers to the top aquitard, (d) – (f) indicates the muti-aquifer/aquitard layers, and (g) – (i) represents the bottom layer of glacial till.

In each scatterplot, a linear model that fits to all data points is provided along with the statistical properties. Examination of Fig. 6.6a – 6.6c reveals that for the upper aquitard, the integration of flux data (Case 2) slightly "improved" K estimates in comparison to the inversion of only head data (Case 1) in terms of the computed L_1, L_2, and R^2 values. Such an "improvement" was reduced as prior geological information was incorporated into inverse modeling, yielding the K estimates from Case 3 that were almost the same as those obtained from Case 4 (Fig. 6.6c). Fig. 6.6d – 6.6f show that when a certain amount of flux data was integrated for inverse modeling, Case 2 yields K estimates were more similar to Case 4 in comparison to Case 1, especially for low K values. Furthermore, the comparison of estimated K values from Case 3 to Case 4 (Fig. 6.6f) shows that the estimated K field within the 15 m × 15 m wellfield from these two cases share similar heterogeneity patterns (data points surround the 45-degree line with relatively large R^2); however, the large scattering of data points indicates small-scale variability of K estimates between Cases 3 and 4. Similar findings were also observed for the bottom glacial till, in which, integrating flux data improved the estimation of low K values when no prior structural information is incorporated (Fig. 6.6g and Fig. 6.6h) and refines K estimates when prior structural information is available (Fig. 6.6i).

We then compared the K estimates from four HT cases to the permeameter K values obtained through laboratory analyses of core samples at the nine well locations (PWs 1-5 and CMTs 1-4), as shown in Fig. 6.7.

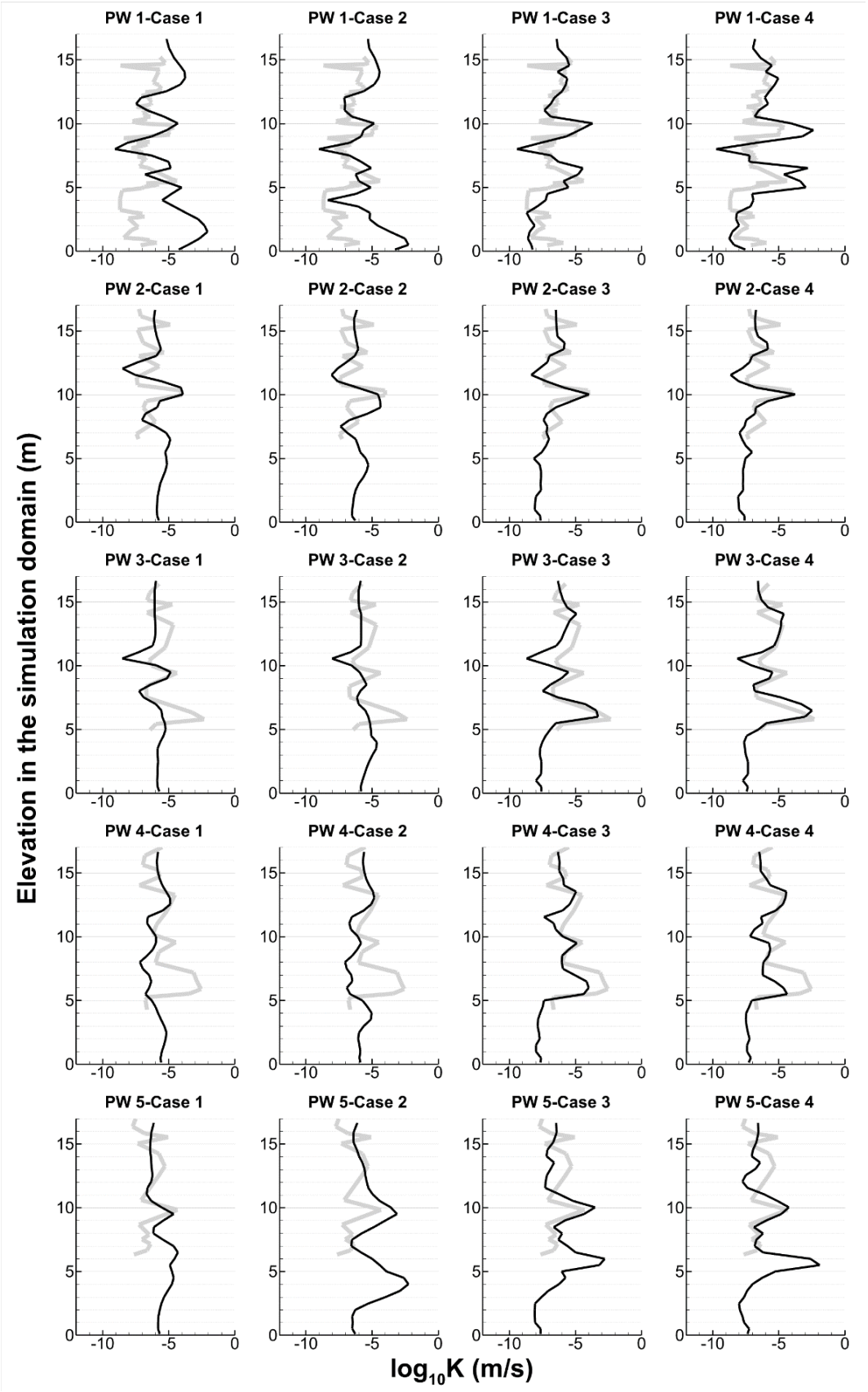

Fig. 6.7: Comparison of $\log_{10}K$ profiles along nine boreholes obtained from permeameter tests and estimated through HT analysis of head data with different prior information (Cases 1 - 4).

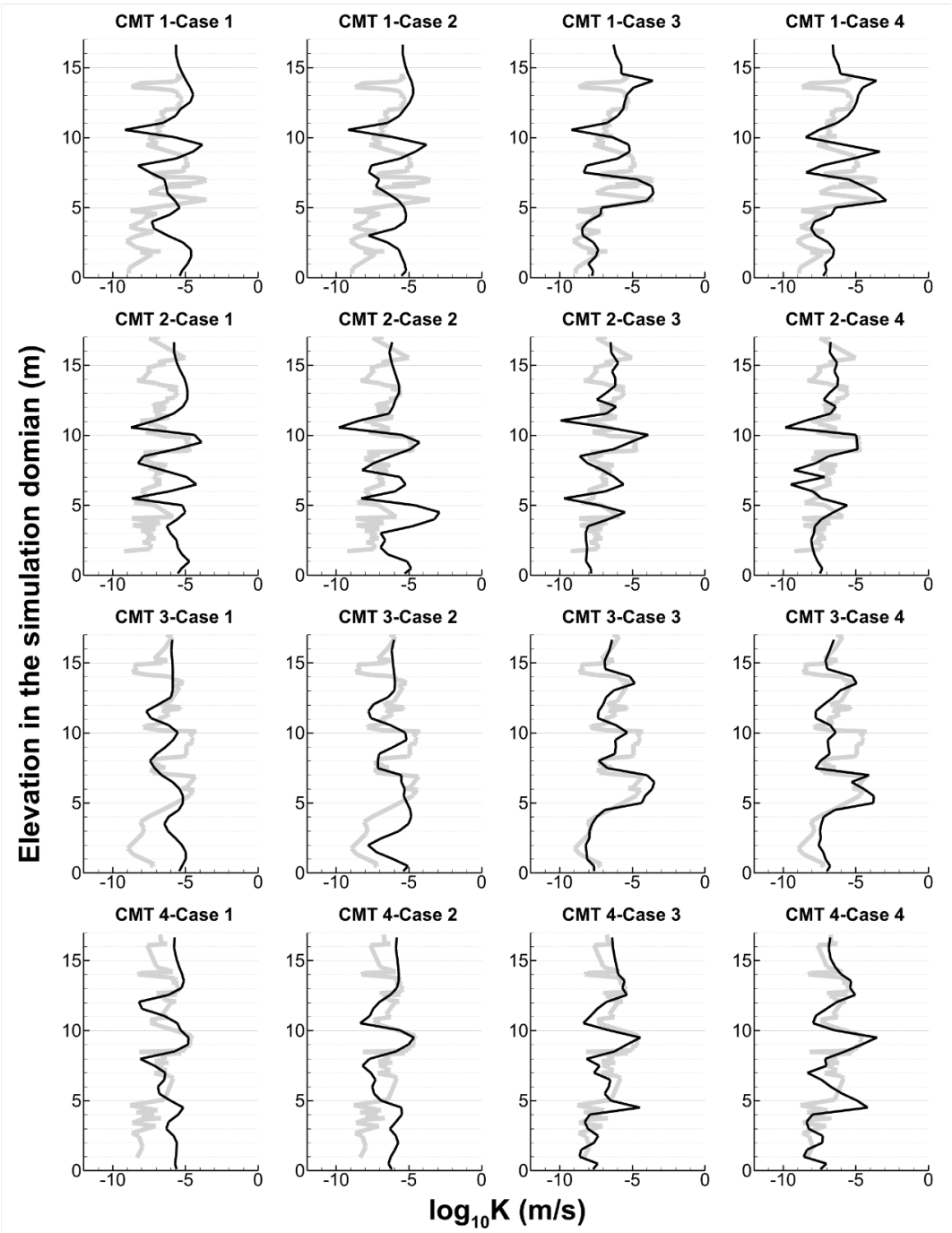

Fig. 6.7 continued

Specifically, geostatistical inversion of head data (Case 1) is capable in revealing general high and low *K* features between 4 and 13 meters above the model domain. The estimated *K* profiles at most well locations are relatively smooth compared to permeameter *K* estimates. After integrating flux data, Case 2 yielded similar vertical variations of *K* in the upper aquitard (13 – 17 m) and the multi-aquitard/aquifer system (4 – 13 m) as Case 1; however, it improved

the estimation of low K values in the bottom aquitard (0 – 4 m) at almost all well locations where permeameter K estimates are available for comparison (PW1, CMTs 1 – 4). This again emphasizes the usefulness of cross-hole flowmeter data collected from multilevel-screened wells in revealing low K zones at the site.

When prior geological information is incorporated for inverse modeling, Cases 3 and 4 yield consistently improved fits between estimated and permeameter K values, while misfits are still observed above the 13 m at most well locations due to the lack of observation data from HT tests. On the other hand, no apparent difference regarding heterogeneity patterns is observed between Case 3 and Case 4. This is consistent with the conclusion derived from the comparison of K estimates within the 15 m × 15 m wellfield obtained from Cases 3 and 4 (Fig. 6.6). When prior geological information is incorporated for inverse modeling, the integration of flux data from cross-hole flowmeter surveys retains heterogeneity patterns while enhancing small-scale variability of K estimates.

Fig. E9 illustrates the maps of computed ln$K$ residual variances from four cases of SSHT analyses, which are used to evaluate the uncertainty of K estimates. Larger residual variance values indicate higher uncertainty and vice versa. For each case, small ln$K$ residual variances exist mainly around pumping/injection ports (blue circles in Fig. E9), while residual variances become larger away from these ports. Due to the inclusion of abundant hydraulic head data for inverse modeling, the estimated uncertainty maps from all four cases are comparable. However, slight improvement could still be visible when additional data types were integrated for inverse modeling.

6.3.3 *Model Calibration and Validation*

Fig. 6.8 shows the calibration scatterplots of the observed versus simulated drawdown data from four SSHT cases. In each scatterplot, data points of individual pumping/injection tests are distinguished using different symbols and colors, and a linear model that fits through all data points is presented along with the statistical properties (L_1, L_2, and R^2).

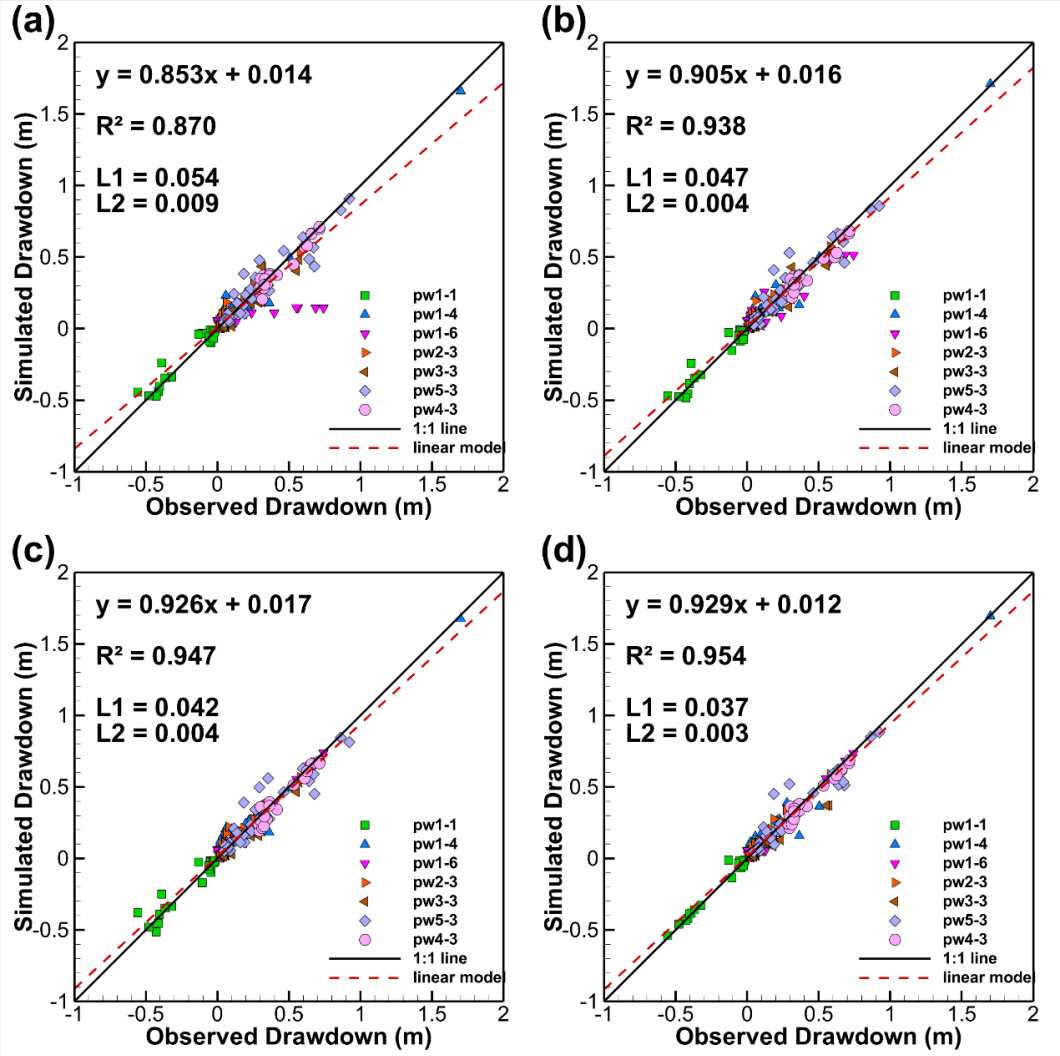

Fig. 6.8: Calibration scatterplots of observed versus simulated drawdowns for Cases 1 to 4 through (a) to (d), respectively.

Fig. 6.8 reveals that Case 4 yields the best calibration result in terms of the smallest L_1 and L_2 norms and the largest R^2 values, followed by Cases 3 and 2. In contrast, Case 1 yields the

worst match between the observed and the simulated drawdowns. This comparison indicates that the reliability of the estimated K tomogram improved when additional datasets that carry non-redundant information of K heterogeneity are integrated for site characterization using HT.

The estimated K tomograms from Cases 1 to 4 were further evaluated via their predictions of drawdown data from seven independent pumping/injection tests not utilized in the calibration effort, and Fig. 6.9 shows the validation scatterplots.

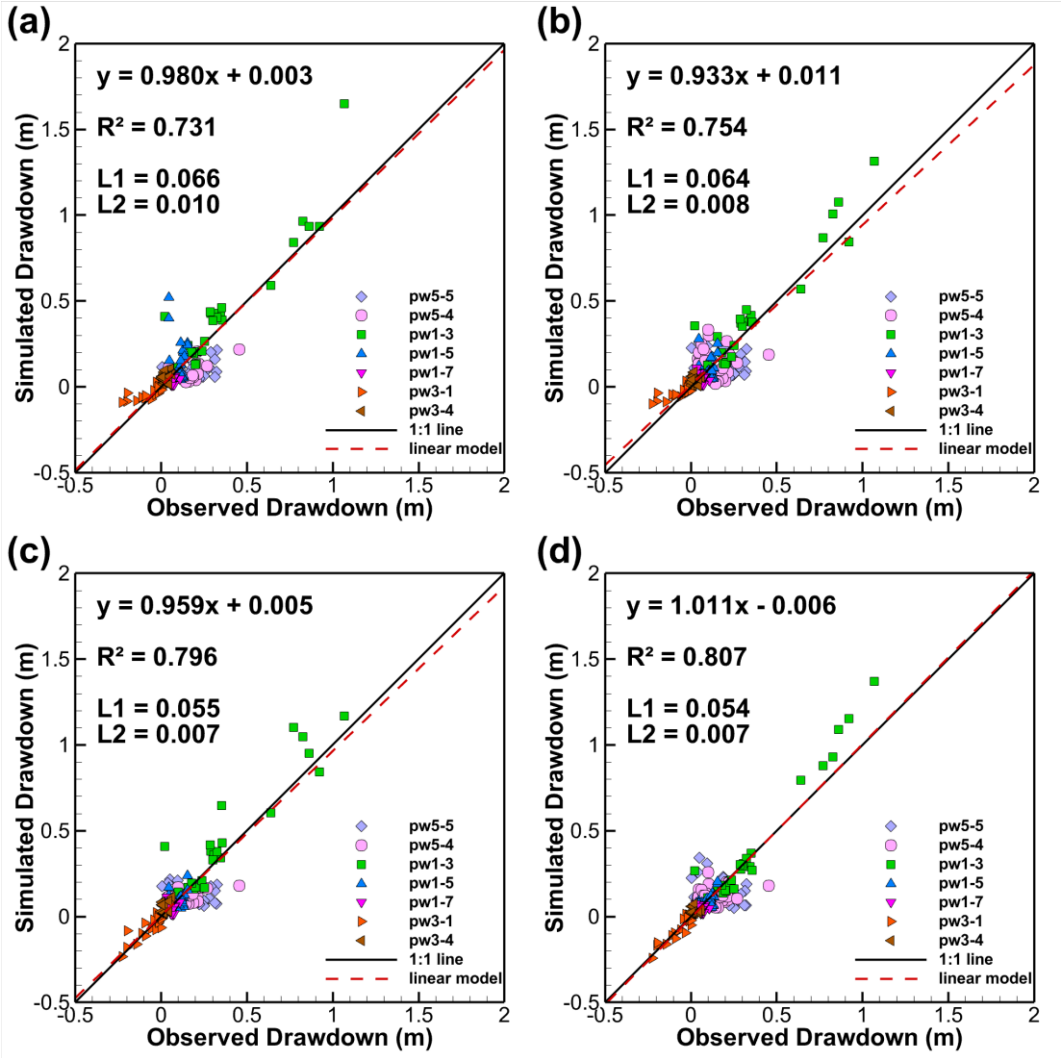

Fig. 6.9: Validation scatterplots of observed versus simulated drawdowns for Cases 1 to 4 through (a) to (d), respectively.

According to the computed statistical properties (L_1, L_2, and R^2), the performances of different cases in predicting independent hydraulic test data share the same order as the

calibration results. In detail, Case 4 performs the best, followed by Cases 3 and 2, while Case 1 yields the worst prediction results in terms of bias and scattering between the observed and simulated drawdowns. On the other hand, it is surprising to notice that integrating flux data with SSHT analysis of head data (Case 2) yields comparable overall validation results compared to SimSLE inversion of steady-state head data with prior geological information (Case 3). It should be noted that the prior geological information in this study was obtained from 18 borehole logs with extensive laboratory analyses of core samples, while the cross-hole flowmeter data were only taken from three multilevel-screened wells during two steady-state pumping tests. Considering different efforts required in obtaining cross-hole flowmeter measurements and reliable structural information, the similar performances of Cases 2 and 3 in predicting independent pumping/injection tests emphasize that the measured vertical flow within boreholes during the cross-hole flowmeter survey is valuable and should be included in the geostatistical analysis for subsurface heterogeneity characterization.

On the other hand, the L_1 and L_2 norms of the four cases for individual pumping/injection tests utilized for model validation are summarized in Table 6.2. This table is color-coded, such that the smallest L_1 and L_2 values in each test are highlighted as green, followed by light green, light yellow, and yellow for increasing values.

Table 6.2 indicates that incorporating additional datasets for model calibration improves predictions of independent hydraulic tests, implying the enhanced delineation of detailed heterogeneity in Case 1 to Case 4 is reliable. Besides, the usefulness of cross-hole flowmeter measurements for site heterogeneity characterization is verified through the improved validation results for Cases 1 to 2 and Cases 3 to 4 for most individual tests. However, such improvement is unclear for the pumping tests conducted at PW5-4 and PW5-5. In contrast, the estimated K tomograms through the integration of flux data with SSHT analysis of head data

(Cases 2 and 4) yield larger discrepancies between the observed and simulated drawdowns. Such discrepancies might be attributed to biased estimation of K values at the bottom of PW5 flux data, implying that attention must be paid to the accuracy of refined K heterogeneities at small scales after integrating flux data with SSHT analysis of head data.

Table 6.2: Summary of L_1 and L_2 norms for individual hydraulic tests utilized for model validation. Orange means worse result while green indicates better result.

	L_1							
	PW1-3	PW1-5	PW1-7	PW3-1	PW3-4	PW5-4	PW5-5	OVERALL
Case 1	0.101	0.067	0.025	0.036	0.040	0.085	0.097	0.066
Case 2	0.095	0.044	0.028	0.035	0.029	0.109	0.103	0.064
Case 3	0.095	0.035	0.025	0.025	0.037	0.070	0.091	0.055
Case 4	0.080	0.033	0.022	0.020	0.026	0.081	0.108	0.054
	L_2							
	PW1-3	PW1-5	PW1-7	PW3-1	PW3-4	PW5-4	PW5-5	OVERALL
Case 1	0.026	0.013	0.001	0.003	0.002	0.010	0.013	0.010
Case 2	0.015	0.004	0.001	0.003	0.001	0.015	0.015	0.008
Case 3	0.020	0.002	0.001	0.001	0.002	0.008	0.013	0.007
Case 4	0.014	0.001	0.001	0.001	0.001	0.010	0.018	0.007

6.4 Discussion

6.4.1 Flowmeter Survey at Multilevel-screened Well

At the NCRS, flowmeter surveys were also performed at individual multilevel-screened wells (PW1, PW3, and PW5) by quasi-steady state pumping at the water surface and measuring vertical flow along the same well. Based on pumping and measuring locations, the collected vertical flow rates under such scenario are named single-hole flowmeter data. Fig. E10 illustrates the summarized single-hole flowmeter data at multilevel-screened wells. Same as Fig. 6.2, ambient, pump-induced, and N.F. logs are plotted in each sub-figure, along with the well screen intervals and the lithology at each well location.

Fig. E10 reveals that significant bypass occurs when performing single-hole flowmeter surveys at all three multilevel-screened wells. Specifically, the reduced magnitude of measured vertical flow at screen intervals implies that a portion of groundwater bypasses the flowmeter through surrounding highly permeable sand packs. Bypass effects in PW1 and PW5 are apparent since the maximum measured vertical flow is much smaller than the assigned pumping rate (1 LPM) at these two wells. Bypass effects can be corrected by multiplying a constant factor to account for leakage (Paillet, 2004) and eliminating unreasonable measurement points. However, such correction may introduce unknown artifacts into the analysis and result in smoothed profiles that fail to capture detailed *K* variations in the vertical direction. Consequently, these single-hole flowmeter data were excluded for the analysis presented in this study.

Compared to single-hole flowmeter data, Fig. 6.2 shows the measured cross-hole flowmeter data at multilevel-screened wells with sand packs are less likely to be affected by the bypass effect, yielding the computed N.F. logs corresponding well the lithology at the highly heterogeneous NCRS. Furthermore, results from this study indicate that geostatistical interpretation of cross-hole flowmeter data is capable in revealing major *K* heterogeneity patterns within the well site.

Such a comparison implies that at sites where traditional single-hole flowmeter survey is restricted for characterizing *K* variations due to the effects of well configuration (i.e., sand/gravel packs) and/or highly heterogeneous condition of hydraulic properties, cross-hole flowmeter survey coupled with sophisticated interpretation methods may be appliable in obtaining reasonable estimation of 3-D *K* distribution.

6.4.2 *Geostatistical Inverse Analysis of Cross-hole Flowmeter Measurements and Its Implementation*

Cross-hole flowmeter data have rarely been used to characterize *K* heterogeneity in unconsolidated deposits such as the multi-aquifer/aquitard system at the NCRS. This may be attributed to the lack of conventional approach that interpret such datasets for providing *K* estimates. Results from this study show that geostatistical inverse analysis of cross-hole flowmeter measurements can map the 3-D distribution of *K* at the site. The estimated vertical distribution of *K* at well locations and high/low *K* zones between tested wells agree well with the known geology, and the estimated *K* tomograms satisfactorily predict vertical fluxes within multilevel-screened boreholes induced by steady-state pumping at a different well location.

The proposed approach in this study can also be applied to interpret single-hole flowmeter data for *K* heterogeneity characterization if high quality flow rate measurements are available. In comparison to conventional interpretation approaches (e.g., Cooper and Jacob, 1946; Kabala, 1994; Molz et al., 1989), numerical simulation of flowmeter tests and geostatistical inversion of the measured flow rates yield estimated *K* distributions not restricted by stratified sublayers and horizontal groundwater flow assumptions. Furthermore, the approach of geostatistical inverse modeling does not require depth-averaged *K* values to derive absolute *K* estimates from the measured flow logs. The depth-averaged *K* values from independent pumping tests may also result in biased *K* estimates due to the discrepancy in influence area between flowmeter and pumping tests.

Despite that geostatistical inverse analysis of cross-hole flowmeter measurements is found to be capable in delineating major high/low *K* zones at a highly heterogeneous site with appropriate estimation of *K* values, the presented inversions rely on the validity of numerical boreholes utilized in forward and inverse analyses. Here, we further examine their validity by

comparing the simulated vertical fluxes using the estimated K tomograms from geostatistical inversions of flux data (Fig. 6.3), but with different K values assigned to the elements representing numerical boreholes. Specifically, the original K values assigned to vertical conduits (1×10^{-2} m/s) and cased intervals (1×10^{-10} m/s) are multiplied and divided, respectively, by factors of 0.1, 5, and 10, yielding three different sets of K values representing numerical boreholes. Fig. E11 illustrates the comparison of observed and simulated vertical fluxes along multilevel-screened wells (PW1, PW3, and PW5) during two steady-state pumping tests (PW1-3 and PW5-3).

Examination of Fig. E11 shows that the estimated K tomogram from Scenario 1 (Fig. 6.3a) yields significantly lowered magnitude of vertical fluxes when assigning K values close to the site geological material (factor 0.1), while it leads to similar predictions of vertical fluxes when setting numerical borehole K values with a larger contrast (factors 5 and 10). However, variation patterns of vertical fluxes along boreholes are comparable to each other when different sets of K values are assigned to elements representing numerical boreholes for simulation. Examination of Fig. E11b reveals that when prior geological information is incorporated for geostatistical inversion of flux data (Scenario 2), the estimated K tomogram (Fig. 6.3b) yields comparable predictions of vertical fluxes when different K values are assigned to numerical boreholes, while unexpected variation patterns are only observed at PW3 when K of vertical conduits and cased intervals are higher and lower, respectively, by a factor of 10. These comparison results indicate that different K values assigned to numerical boreholes show slight impacts on the prediction of flux magnitudes, but they retain variation patterns of vertical fluxes along boreholes under most tested scenarios. Therefore, we may conclude that the settings of numerical boreholes might affect the estimation of K values; however, they could still be utilized to reveal K heterogeneity patterns.

Based on the above analyses, this study refrains from predicting hydraulic test data using the estimated *K* fields from geostatistical inverse analyses of cross-hole flowmeter tests. Instead, these *K* fields are used as initial *K* fields for the integration of cross-hole flowmeter data with SSHT analysis of head data. However, a more rigorous representation of flowmeter surveys including borehole geometry and corresponding hydraulic properties for numerical modeling is suggested for future studies.

6.4.3 *Usefulness of Cross-hole Flowmeter Data for K heterogeneity Characterization*

This study employs three types of data (geological information, cross-hole flowmeter measurements, and steady-state head responses) to characterize the spatial distribution of *K* at the NCRS. Different types of data are integrated sequentially as initial *K* fields for inverse modeling. The joint inversion of different datasets may lead to the same characterization results, while using initial *K* guesses conditioned on the previous dataset as the prior information is more efficient in terms of inversion. The usefulness of integrating cross-hole flowmeter data for mapping spatial *K* distribution is evaluated through the comparison of calibration and validation results from four cases.

The comparison results between Cases 1 and 2 show that integrating flux data helps to reveal more heterogeneity details and improve model calibration and validation results. This implies that the cross-hole flowmeter measurements at the NCRS carry non-redundant heterogeneity information about *K* compared to the existing HT data collected from a series of spatially varying pumping/injection tests. Although cross-hole flowmeter measurements are only available at limited number of wells through two steady-state pumping tests, results in this study show that the revealed high/low *K* zones (Fig. 6.3a) through geostatistical inversion of the corresponding flux data can be applied to constrain the inversion of head data for improved

characterization of *K* heterogeneity. Further improvements might be observed if cross-hole flowmeter measurements are taken from a series of spatially varying pumping/injection tests, like the traditional HT survey for 3-D mapping of hydraulic parameters.

When geological data are incorporated as prior information for inverse modeling, Cases 3 and 4 yield significantly improved delineation of *K* heterogeneity throughout the simulation domain, implying that the importance of prior structure information for characterizing spatial distribution of hydraulic parameters using HT. Although the estimated hydraulic parameter values may remain uncertain where there is lack of observation data, the heterogeneity information derived from various field surveys (i.e., lithology, geophysical survey, etc.) plays an important role in adequately predicting groundwater flow and solute transport at a large-scale site. However, attention must be paid regarding the reliability of obtained structural information, especially when the observation data utilized for inverse modeling are collected from sparsely distributed locations with limit number of aquifer test, as demonstrated in the completed studies described in previous chapters.

When comparing the characterization results of Case 1 with Case 2 and those of Case 3 with Case 4, we notice that the influence of flux data reduced when prior geological information is incorporated for inverse modeling. This finding corroborates the work by Tso et al. (2016), which pointed out that flux data carry information mainly about the connectivity between pumping and observation locations, and the prior geological information with distributed initial mean *K* field for inverse modeling is superior to the heterogeneity information derived from flux data. Tso et al. (2016) also demonstrated that the influence of prior structural information would be relieved as more non-redundant data were used in the HT analysis. However, the comparison of characterization results between Cases 2 and 3 reveals that prior geological information shows superior influence on *K* estimates to the flux data in our study due to the

limited number of cross-hole flowmeter measurements.

Overall, the integration of flux data has shown to improve the overall validation results in terms of L1, L2, and R2 (Fig. 6.9). However, such improvement is still unclear regarding individual independent pumping/injection tests. As indicated in Table 6.2, the estimated K tomograms from Case 2 and Case 4 yield larger discrepancies of the observed and simulated drawdowns of pumping tests at PW5-4 and PW5-5 in comparison to those obtained from Case 1 and Case 3, respectively. Such worsen prediction results may be due to the biased K estimates at the bottom of PW5 obtained through geostatistical inversion of flux data, which are utilized as initial K guesses for integrating flux data with SSHT analyses of head data (Cases 2 and 4). We attribute biased estimation of *K* values to the measurement errors of cross-hole flowmeter data the bottom of PW5; however, future investigation is required to address this issue. This indicates that despite the benefit of utilizing cross-hole flowmeter measurements as additional datasets in HT analysis, attention must be paid to the accuracy of field flowmeter measurements to avoid biased estimation of hydraulic parameters. Once field flowmeter survey can be progressively conceptualized for numerical modeling, joint interpretation of flowmeter and head data is suggested for future HT studies to enhance the heterogeneity characterization of hydraulic parameters.

7 Summary and Conclusions

Summary and conclusions of the four completed HT studies in this thesis are provided individually as followings.

Study I: *Comparative Study of Transient Hydraulic Tomography with Varying Parameterization and Zonations: Laboratory Sandbox Investigation*

In this study, a synthetic heterogeneous aquifer constructed in a laboratory sandbox (Illman et al., 2010) is characterized through transient hydraulic tomography (THT) with varying parameterized and zonation models, including one effective parameter model, four geology-based zonation models of varying accuracy and resolution, and five geostatistical models with different prior structural information. All models are calibrated using two datasets of different pumping and observation densities through simultaneous inversion of transient head data in each dataset. In total, 20 sets of K and S_s tomograms are obtained for the synthetic aquifer characterization. The performances of these K and S_s tomograms are validated through the prediction of drawdowns from 16 independent pumping tests. Results from this study lead to following findings and conclusions:

1. Treating the synthetic aquifer as a homogeneous medium, the simultaneous inversion of multiple pumping tests yields effective K and S_s estimates that are more representative of the aquifer in comparison to those generated from small scale estimates. Although the

calibrated effective parameter provides zero resolution on aquifer heterogeneity, it can be used as a starting point for zonation or highly parameterized model calibration.

2. All geology-based zonation models are well calibrated, yielding quite different patterns of estimated high and low value zones of K and S_s among different models. However, only the calibrated zonation model with good stratigraphic information is found to be adequate in predicting independent pumping tests. These results indicate that reliable (if not accurate) structural information is required when calibrating zonation models for site characterization.

3. By calibrating the model using only hydraulic head data, we find that the highly parameterized geostatistical model yields significantly improved calibration and validation results in comparison to effective and zonation models when a large amount of head data is available for inverse modeling (Case 1). However, when the number of head data is reduced (Case 2), the zonation model based on good stratigraphy information yields slightly improved prediction results in comparison to the geostatistical model. These results imply that for a given amount of head data, the number of parameters to be estimated should be considered carefully to avoid the issue of overparameterization.

4. Prior geological information shows slight impact on characterization results when a great number of head data is available for geostatistical inverse modeling, but it becomes important after reducing head data for model calibration. When the number of pumping tests and observation intervals is limited, “accurate” or simplified structural information is suggested to be incorporated for geostatistical inverse modeling to reveal more

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

5. We also compared the results from steady state and transient inversions of the same pumping test data. Forward simulations of 16 pumping tests conducted with the K tomogram obtained from the steady state inversion together with an estimate of S_s obtained from single-hole tests yielded significantly biased transient drawdown predictions. Therefore, our results suggest that in order to obtain accurate predictions of transient drawdowns from independent tests, K and S_s tomograms from THT analyses are necessary.

Study II: *Three-dimensional Hydraulic Tomography Analysis of Long-term Municipal Wellfield Operations: Validation with Synthetic Flow and Solute Transport*

In this study, a synthetic 3D multi-aquifer/aquitard system is characterized using HT-based approaches through the interpretation of long-term water-supply pumping/injection records (municipal well data). In particular, pumping/injection rate records from 13 water-supply wells and simulated hydraulic head observations at 28 monitoring locations are interpreted to map subsurface heterogeneities in hydraulic conductivity (K) and specific storage (S_s). To investigate the performance of different modeling approaches and the effect of data selection

on inverse modeling, the synthetic system is successively characterized using four groundwater models (effective parameter model, geological model, and two geostatistical models with different prior information) through the interpretation of five datasets consisting of different time durations and periods within a given year. The estimated K and S_s tomograms from all investigated models are then applied to predict municipal well data with the existing water-supply wells and independent pumping test data from additional water-supply wells for model validation. Additional model runs are performed to investigate the ability of estimated K and S_s tomograms in predicting solute transport in subsurface conditions for a stronger form of validation of HT results. This study results in the following findings and conclusions:

1. Results from all investigated models reveal that HT analysis of long-term pumping/injection and water-level records is feasible and yields reliable K and S_s estimates where hydraulic data are available. In comparison to traditional subsurface characterization with dedicated pumping tests, the utilization of such data is able to reveal large-scale heterogeneities of hydraulic parameters and yield K and S_s estimates representative of aquifer properties during existing pumping/injection event, while reducing cost and time requirements for site characterization.

2. To avoid the effect of uncertain initial conditions on inverse modeling when using long-term records for site heterogeneity characterization, pumping/injection records prior to the observation data should be accounted for during model calibration. To minimize the effect of uncertain initial conditions, while maintaining computational efficiency for inverse

modeling, preliminary characterization of well hydrographs at the study site is suggested to select an appropriate length of prior pumping/injection records.

3. The calibration of the effective parameter model yields K and S_s estimates that are more representative to the effective hydraulic parameters of the upper layers, where most monitoring wells are screened with sufficient hydraulic head data. The utilization of these values yields significantly biased predictions of hydraulic head variations at monitoring wells, implying the importance of considering heterogeneity for subsurface characterization. With well identified geological layers and well estimated initial hydraulic parameters, the calibrated geological model is found able to provide relatively adequate predictions of drawdown variations. However, additional hydraulic data at different geological layers are still required to obtain reliable estimates of hydraulic parameters for each hydrostratigraphic unit.

4. Stratigraphy information is verified to be of critical importance for large-scale heterogeneity characterization, in which hydraulic data are typically sparsely located with limited number of monitoring wells. The geostatistical inversion of hydraulic head data only, is able to reveal heterogeneity details where head data are concentrated; however, the overall smooth patterns and poor predictions of independent pumping test data cause the estimated K and S_s tomograms to fail to represent site specific heterogeneities. After incorporating the geological data as prior information, the geostatistical model reveals greater detail of subsurface heterogeneity and yields K and S_s tomograms comparable to

the “true” fields. The estimated K and S_s tomograms provide adequate predictions of not only the municipal well data with the existing water-supply wells, but also independent pumping test data from additional pumping wells, implying that these estimated hydraulic parameter fields can be used to guide the construction of new water-supply wells.

5. The effect of data selection on inverse modeling is investigated by manually selecting different datasets, on the basis of duration and period for model calibration. Based on the comparison results, continuous data points with large water-level variations are suggested to be incorporated for large-scale heterogeneity characterization using the geostatistical model with geological information. However, new approaches need to be developed for big data synthesis and intelligent data selection for inverse modeling.

6. Synthetic conservative solute transport simulations conducted with various estimated hydraulic parameter fields (effective parameter model, geological model, and geostatistical models with different prior information) reveal that solute migration is strongly impacted by the heterogeneity of hydraulic parameters (K and S_s). Although the calibrated geological model provides adequate predictions of head variations at monitoring wells, it yields poor predictions of contaminant transport due to the neglect of intralayer heterogeneities and poor estimation of K and S_s values at the source layer where hydraulic head data are few for model calibration. On the other hand, the geostatistical inversion of the municipal well data incorporated with geological information yields K and S_s tomograms that can provide adequate predictions of not only drawdown variations at monitoring wells but also solute

transport in subsurface conditions, indicating the superior application of this approach for large-scale heterogeneity characterization using the long-term water-supply pumping/injection records.

Study III: *Large-scale Three-dimensional Hydraulic Tomography Analyses of Long-term Municipal Wellfield Operations*

In this study, the long-term municipal well records at the Mannheim wellfield, a municipal water-supply wellfield located in the Regional Municipality of Waterloo, Ontario, Canada, are obtained and subjected to THT analyses using geology-based zonation and geostatistical models for large-scale subsurface heterogeneity characterization. Uncertainties regarding the interpreted initial and boundary conditions are addressed to avoid estimation of biased hydraulic parameters. The estimated K and S_s tomograms from both models are then validated through predictions of water-level variations that have not been used for calibration efforts. This study resulted in following findings and conclusions:

1. With well addressed uncertainties in groundwater modeling, existing long-term hydrographs that have been affected by multiple production wells at variable flow rates can be adopted for the characterization of subsurface heterogeneity through transient HT. In comparison to traditional HT surveys with dedicated cross-hole pumping tests, the utilization of such existing datasets is capable of revealing large-scale heterogeneity features of K and S_s, while maintaining cost efficiency for large-scale site characterization.

2. In this study, initial and lateral boundary conditions for groundwater flow modeling are estimated through kriging of available static water-level measurements. To minimize their uncertainties, a certain simulation period prior to the municipal well records is incorporated for model spin-up (uncertain initial condition) and water-level variations in each observation well based on the corresponding datum head data are utilized for model calibration (uncertain lateral boundary conditions). Such proposed data processing and analysis strategies are recommended when analyzing existing hydrographs for subsurface heterogeneity characterization to increase the stability of inverse models, its convergence, and to avoid biased estimation results.

3. Existing municipal well records provide abundant hydraulic response data that have been affected by long-term operations of municipal water-supply wells at variable flow rates. However, as it is impractical to incorporate all data points, existing hydrographs in observation wells should be assessed carefully to select appropriate head data for inverse modeling. Following the synthetic study, daily observation data during a six-month period with large water-level variations are extracted and utilized for the analyses presented in this study. The utilization of such datasets is found to yield reliable K estimates where observation wells are concentrated; however, it yields relatively poor S_s estimates. As a result, advanced temporal sampling strategies for HT analysis of long-term municipal well records are still required to obtain reliable estimates of both K and S_s, while maintaining computational efficiency for inverse modeling.

4. With sufficient hydraulic response data, the geology-based zonation yields satisfactory estimation of hydraulic parameter within the water-supply aquifer (AFB2), while the geostatistical inverse model with prior geological information reveals great details of large-scale heterogeneity in hydraulic parameters, yielding significantly improved calibration and validation results. Such comparison results advocate that geostatistical inversion should be utilized as a second step over traditional geology-based zonation model when analyzing municipal well records for large-scale heterogeneity characterization. On the other hand, smooth K and S_s tomograms estimated for the upper and lower geological units as well as the estimated high uncertainties of the corresponding K and S_s estimates imply that additional head data from these geological units are still required to obtain improved characterization results for the entire simulation domain. Alternatively, other site data could potentially be integrated to improve results within those units.

5. Model conceptualization is a fundamental step for constructing groundwater flow models. To maintain computational efficiency for inverse modeling, several simplifications (saturated subsurface condition, no surface water feature, constant recharge rate) are made to the conceptual model adopted for this study. Such simplifications are evaluated to have negligible impacts on the analyses presented in this study; however, it might be very important for the long-term simulation of water-level variations for other field problems depending on their hydrologic characteristics. On the other hand, some uncertainties associated with model conceptualization that cannot be fully characterized through preliminary examination of hydraulic head data (e.g., unknown stimuli impacting water-

level variations) can be revealed through the assessment of model calibration and validation results. This finding implies that the characterization results can be used to improve the construction of groundwater flow models through the improvement of model conceptualization.

Study IV: *Three-dimensional Steady-state Hydraulic Tomography Analysis with Integration of Cross-hole Flowmeter Data at a Highly Heterogeneous Site*

Flowmeter surveys have been used to estimate the vertical variation of horizontal *K* at well locations with pumping-induced flow logs coupled with depth-averaged *K* estimates. Spatially interpolating the estimated *K* values at different locations could lead to a 3-D *K* distribution for groundwater flow modeling. Such heterogeneity characterization approach, however, suffers from several drawbacks. For instance, estimated *K* values through conventional analyses of flowmeter data could be biased in representing realistic conditions due to the implemented assumptions of stratified layers and horizontal groundwater flow. Furthermore, the classical interpretation methods (i.e., kriging) often yield overall smooth *K* tomograms when point estimates are sparsely distributed. Thus, we propose numerical inverse modeling of flowmeter data using HT for *K* heterogeneity characterization. Briefly, HT takes many snapshots of a 3-D heterogeneous *K* field at different locations, angles, and perspectives, using excitations created by pumping or injection of water and the recorded the groundwater responses at monitoring wells as images. It then synthesizes these snapshots using SimSLE to yield the 3-D estimated *K* fields that can reproduce these snapshots. SimSLE is an inverse algorithm that

exploits prior geologic information (prior mean, variance, and correlation scales) as soft constraints to adjust the unknown *K* field to ensure the simulated 3-D groundwater flow dynamics conformed to the HT snapshots. It further ensures that the estimated 3-D *K* fields honor local-scale *K* values as well as head and flux data at sample locations.

This study's major accomplishment is integrating cross-hole flowmeter measurements as a new type of data with SimSLE inversions of head response data to improve *K* heterogeneity characterization at the NCRS. We first examine the quality of cross-hole flowmeter measurements at the NCRS to select proper data points for the analyses. Numerical boreholes are adopted to simulate cross-hole flowmeter surveys, and the measured vertical flow rates within multilevel-screened wells are subjected to geostatistical inverse analyses to estimate spatial *K* distributions under two scenarios (with/without prior geological information). The estimated *K* tomograms are then adopted as initial *K* fields for SimSLE inversions of head response data to characterize 3-D spatial *K* distribution at the NCRS.

Four cases (Cases 1 – 4) integrating different types of data (geological information, cross-hole flowmeter measurements, and steady-state head response data) for inverse modeling are examined. Their performances are qualitatively and quantitively assessed in terms of model calibration and validation. Our study leads to following major findings and conclusions:

1. Despite the fact that different types of flowmeters that have been developed to increase the accuracy of vertical flow rate measurements within a borehole, the measured flow logs still significantly are affected by well configuration. These results suggest for the preliminary

examination of flowmeter data before its interpretation for *K* estimates. At the NCRS, induced flowmeter measurements at multilevel-screened wells are highly influenced by bypass effects. However, such bypass effects are not observed during two cross-hole flowmeter surveys (steady-state pumping at PW1-3/PW5-3), yielding the cross-hole flow logs at multilevel-screened wells (PW1, PW3, and PW5) consistent with the lithology at well locations. These initial analyses suggest the potential feasibility of interpreting cross-hole flowmeter for *K* estimates.

2. With proper conceptualization of groundwater flow within wells, cross-hole flowmeter data collected during steady-state pumping tests could be interpreted using a geostatistical inverse modeling approach to image the heterogeneity patterns of *K*.

3. In addition to revealing vertical distributions of *K* at well locations, the estimated *K* tomogram is also found to be capable in capturing major high/low *K* zones within the 15 m by 15 m wellfield. After incorporating prior geological information, inter/intra-layer variation patterns of *K* heterogeneity are both refined. Characterization results indicate that geostatistical interpretation of cross-hole flowmeter data is able to reveal major *K* heterogeneity patterns and provide a preliminary estimation of *K* values, making it possible to integrate such information for subsurface heterogeneity characterization through SimSLE inversion of steady-state head response data.

4. The usefulness of cross-hole flowmeter data for mapping 3-D spatial *K* distribution is verified based on the field study at the NCRS underlain by a highly heterogeneous multiple-

aquifer/aquitard system. Specifically, the collected cross-hole flowmeter data are examined to carry non-redundant information of *K* heterogeneity compared to the existing HT head data and structural information, and the integration of such datasets improves characterization of *K* heterogeneity details as well as model calibration and validation. At locations where flowmeter measurements are available, the integration of flux data with SSHT analysis of head data reveals more heterogeneity patterns of *K*, and it refines small-scale variability of *K* estimates when prior geological information is additionally incorporated for inverse modeling. Results from this study emphasize the meaningful usage of cross-hole flowmeter data in porous media for *K* heterogeneity characterization.

8 Recommendations for Future HT Investigations

This thesis has shown that the integration of diverse data sources into hydraulic tomography analyses has revolutionized our understanding of spatial distribution of hydraulic parameters. Traditional subsurface heterogeneity characterization methods, relying on individual data types, struggle to capture the intricate spatial variability of hydraulic properties. The integration of various types of data, however, yields refined characterization of hydraulic properties with greater heterogeneity details. Based on the completed four studies, some recommendations are derived for future HT investigations:

1. Future applications of HT should continue to emphasize the integration of multiple data types. Incorporating data from various sources, such as geological maps and logs, geophysical surveys, tracer/heat tests, and other emerging imaging techniques, will provide a more comprehensive view of subsurface heterogeneity of hydraulic properties. Efforts should focus on refining methodologies for effectively fusing these datasets to capture heterogeneity details across different scales.

2. The use of machine learning and data analytics holds promise for enhancing HT analyses. Future applications can leverage these tools to automate data processing, identity patterns in complex datasets, and even improve inversion algorithms. Incorporating artificial intelligence techniques can help in streamlining the interpretation of large volumes of data for subsurface heterogeneity characterization using the approach of HT.

3. Advancements in sensor technology to enable real-time monitoring of subsurface processes. Future HT applications could incorporate such monitoring systems to (1) continuously refine characterization results through assimilating information from new datasets and (2) provide dynamic insights into changes in hydraulic properties over time at mining, remediation, and other sites. This approach is particularly valuable for understanding transient phenomena and improving predictive models.

4. Field application of HT at sites of diverse geological settings (e.g., vadose zones, facture rocks) is crucial to showcase the versatility of HT for addressing real-world challenges in groundwater management and contaminant remediation.

5. Technological advancements are rapid and can aid in improving HT. Future HT applications should remain adaptable to changes in sensor technology, computational resource, and data collection/management strategies. Regular updates to methodologies and tools will ensure that HT remains at the forefront of subsurface heterogeneity characterization.

In conclusion, the future application of HT holds immense potential for advancing subsurface characterization. By embracing multi-source integration and technological innovation, HT can continue to play a vital role in characterizing heterogeneity of hydraulic parameters and supporting sustainable groundwater management and environmental assessments.

References

Ackerer, P., Carrera, J., Delay, F., 2023. Identification of aquifer heterogeneity through inverse methods. Comptes Rendus. Géoscience, 355(S1): 1-14. DOI:10.5802/crgeos.162

Ahmed, A.S., Zhou, J., Jardani, A., Revil, A., Dupont, J., 2015. Image-guided inversion in steady-state hydraulic tomography. Advances in Water Resources, 82: 83-97. DOI:10.1016/j.advwatres.2015.04.001

Alcolea, A., Carrera, J., Medina, A., 2006. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. Advances in water resources, 29(11): 1678-1689. DOI:10.1016/j.advwatres.2005.12.009

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

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

Bajc, A.F., Shirota, J., 2007. Three-dimensional mapping of surficial deposits in the Regional Municipality of Waterloo, southwestern Ontario. Groundwater Resources Study 3, Ontario Geological Survey.

Batu, V., 1998. Aquifer hydraulics: A comprehensive guide to hydrogeologic data analysis. John Wiley & Sons, Inc., New York.

Bennett, R.R., Meyer, R.R., 1952. Geology and ground-water resources of the Baltimore area, Maryland Geological Survey.

Berg, S.J., Illman, W.A., 2011a. Capturing aquifer heterogeneity: Comparison of approaches through controlled sandbox experiments. Water Resour. Res., 47(9). DOI:10.1029/2011WR010429

Berg, S.J., Illman, W.A., 2011b. Three-dimensional transient hydraulic tomography in a highly heterogeneous glaciofluvial aquifer-aquitard system. Water Resour. Res., 47(10). DOI:10.1029/2011WR010616

Berg, S.J., Illman, W.A., 2012. Improved predictions of saturated and unsaturated zone drawdowns in a heterogeneous unconfined aquifer via transient hydraulic tomography: Laboratory sandbox experiments. J. Hydrol., 470–471: 172-183. DOI:10.1016/j.jhydrol.2012.08.044

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

Berg, S.J., Illman, W.A., 2015. Comparison of hydraulic tomography with traditional methods at a highly heterogeneous site. Ground Water, 53(1): 71-89. DOI:10.1111/gwat.12159

Bohling, G.C., Butler, J.J., Zhan, X., Knoll, M.D., 2007. A field assessment of the value of steady shape hydraulic tomography for characterization of aquifer heterogeneities. Water Resour. Res., 43(5). DOI:10.1029/2006wr004932

Bohling, G.C., Zhan, X., Butler, J.J., Zheng, L., 2002. Steady shape analysis of tomographic

pumping tests for characterization of aquifer heterogeneities. Water Resour. Res., 38(12). DOI:10.1029/2001wr001176

Bomana, G.K., Molz, F.J., Boonec, K.D., 1997. Borehole flowmeter application in fluvial sediments: Methodology, results, and assessment. Groundwater, 35(3): 443-450. DOI:10.1111/j.1745-6584.1997.tb00104.x

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. DOI:10.1111/gwat.13039

Brauchler, R., Hu, R., Dietrich, P., Sauter, M., 2011. A field assessment of high-resolution aquifer characterization based on hydraulic travel time and hydraulic attenuation tomography. Water Resour. Res., 47(3). DOI:10.1029/2010WR009635

Brauchler, R., Liedl, R., Dietrich, P., 2003. A travel time based hydraulic tomographic approach. Water Resour. Res., 39(12). DOI:10.1029/2003wr002262

Burnett, R.D., Frind, E.O., 1987. Simulation of contaminant transport in three dimensions: 2. Dimensionality effects. Water Resources Research, 23(4): 695-705. DOI:10.1029/WR023i004p00695

Cardiff, M., Barrash, W., Kitanidis, P., Malama, B., Revil, A., Straface, S., Rizzo, E., 2009. A potential-based inversion of unconfined steady-state hydraulic tomography. Ground Water, 47(2): 259-270. DOI:10.1111/j.1745-6584.2008.00541.x

Cardiff, M., Barrash, W., Kitanidis, P.K., 2012. A field proof-of-concept of aquifer imaging using 3-D transient hydraulic tomography with modular, temporarily-emplaced equipment. Water Resour. Res., 48(5). DOI:10.1029/2011WR011704

Cardiff, M., Barrash, W., Kitanidis, P.K., 2013. Hydraulic conductivity imaging from 3-D transient hydraulic tomography at several pumping/observation densities. Water Resour. Res., 49(11): 7311-7326. DOI:10.1002/wrcr.20519

Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., Slooten, L.J., 2005. Inverse problem in hydrogeology. Hydrogeol. J, 13(1): 206-222. DOI:10.1007/s10040-004-0404-7

Carrera, J., Neuman, S.P., 1986a. Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. Water Resources Research, 22(2): 199-210. DOI:10.1029/WR022i002p00199

Carrera, J., Neuman, S.P., 1986b. Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability, and solution algorithms. Water Resour. Res., 22(2): 211-227. DOI:10.1029/WR022i002p00211

Carrera, J., Neuman, S.P., 1986c. Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data. Water Resources Research, 22(2): 228-242. DOI:10.1029/WR022i002p00228

Castagna, M., Becker, M.W., Bellin, A., 2011. Joint estimation of transmissivity and storativity in a bedrock fracture. Water Resour. Res., 47(9). DOI:10.1029/2010wr009262

Certes, C., de Marsily, G., 1991. Application of the pilot point method to the identification of aquifer transmissivities. Advances in Water Resources, 14(5): 284-300. DOI:10.1016/0309-1708(91)90040-U

CH2MHILL, Papadopulo&Associates, S.S., 2003. Alder Creek Groundwater Study - Final

Report, Regional Municipality of Waterloo.
Cooper, H.H., Jacob, C.E., 1946. A generalized graphical method for evaluating formation constants and summarizing well field history. Am. Geophys. Union Trans., 27: 526-534. DOI:10.1029/TR027i004p00526
de Marsily, G., Lavedan, G., Boucher, M., Fasamino, G., 1984. Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model, Geostatistics for natural resources characterization. NATO advanced Study Institute, pp. 831-849. DOI:10.1007/978-94-009-3701-7_16
Dinwiddie, C.L., Foley, N.A., Molz, F.J., 1999. In-well hydraulics of the electromagnetic borehole flowmeter. Groundwater, 37(2): 305-315. DOI:10.1111/j.1745-6584.1999.tb00988.x
Doherty, J., 2003. Ground water model calibration using pilot points and regularization. Groundwater, 41(2): 170-177. DOI:10.1111/j.1745-6584.2003.tb02580.x
Doherty, J., 2015. Calibration and uncertainty analysis for complex environmental models. Watermark Numerical Computing Brisbane, Australia.
Fienen, M.N., Kitanidis, P.K., Watson, D., Jardine, P., 2004. An application of Bayesian inverse methods to vertical deconvolution of hydraulic conductivity in a heterogeneous aquifer at Oak Ridge National Laboratory. Mathematical Geology, 36(1): 101-126. DOI:10.1023/B:MATG.0000016232.71993.bd
GolderAssociatesLtd., 2011. Tire 3 water budget and local area risk assessment: mannheim well fields characterization. 08-1112-0020, Regional Municipality of Waterloo.
Gottlieb, J., Dietrich, P., 1995. Identification of the permeability distribution in soil by hydraulic tomography. Inverse Probl, 11: 353-360. DOI:10.1088/0266-5611/11/2/005
Hantush, M.S., Jacob, C.E., 1955. Non-steady radial flow in an infinite leaky aquifer. Eos, Transactions American Geophysical Union, 36(1): 95-100. DOI:10.1029/TR036i001p00095
Harp, D.R., Vesselinov, V.V., 2011. Identification of pumping influences in long-term water level fluctuations. Ground water, 49(3): 403-414. DOI:10.1111/j.1745-6584.2010.00725.x
Hendricks-Franssen, H.J., 2001. Inverse stochastic modelling of groundwater flow and mass transport. Universidad Politecnica de Valencia (Spain).
Hu, R., Brauchler, R., Herold, M., Bayer, P., 2011. Hydraulic tomography analog outcrop study: Combining travel time and steady shape inversion. Journal of Hydrology, 409(1): 350-362. DOI:10.1016/j.jhydrol.2011.08.031
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). DOI:10.1029/2011WR010698
Hyder, Z., Butler Jr, J.J., McElwee, C.D., Liu, W., 1994. Slug tests in partially penetrating wells. Water Resources Research, 30(11): 2945-2957. DOI:10.1029/94WR01670
Illman, W.A., Berg, S.J., Yeh, T.C.J., 2012. Comparison of approaches for predicting solute transport: Sandbox experiments. Ground Water, 50(3): 421-431. DOI:10.1111/j.1745-

6584.2011.00859.x

Illman, W.A., Berg, S.J., Zhao, Z., 2015. Should hydraulic tomography data be interpreted using geostatistical inverse modeling? A laboratory sandbox investigation. Water Resour. Res., 51(5): 3219-3237. DOI:10.1002/2014WR016552

Illman, W.A., Craig, A.J., Liu, X., 2008. Practical issues in imaging hydraulic conductivity through hydraulic tomography. Ground Water, 46(1): 120-132. DOI: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. J. Hydrol., 341(3–4): 222-234. DOI: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 Resour. Res., 45(1). DOI:10.1029/2007WR006715

Illman, W.A., Zhu, J., Craig, A.J., Yin, D., 2010. Comparison of aquifer characterization approaches through steady state groundwater model validation: A controlled laboratory sandbox study. Water Resour. Res., 46(4): n/a-n/a. DOI:10.1029/2009WR007745

Jiménez, S., Brauchler, R., Bayer, P., 2013. A new sequential procedure for hydraulic tomographic inversion. Advances in water resources, 62: 59-70. DOI:10.1016/j.advwatres.2013.10.002

Jiménez, S., Brauchler, R., Hu, R., Hu, L., Schmidt, S., Ptak, T., Bayer, P., 2015. Prediction of solute transport in a heterogeneous aquifer utilizing hydraulic conductivity and specific storage tomograms. Water Resources Research, 51(7): 5504-5520. DOI:10.1002/2014WR016402

Kabala, Z., 1994. Measuring distributions of hydraulic conductivity and specific storativity by the double flowmeter test. Water Resources Research, 30(3): 685-690. DOI:10.1029/93WR03104

Karrow, P.F., 1979. Geology of the University of Waterloo Campus, University of Waterloo, Waterloo, Ontario, Canada.

Karrow, P.F., 1993. Quaternary Geology, Stratford-Conestogo Area. Report 283, Ontario Geological Survey.

Kitanidis, P.K., 1995. Quasi-linear geostatistical theory for inversing. Water Resour. Res., 31(10): 2411-2419. DOI:10.1029/95WR01945

Kitanidis, P.K., Lee, J., 2014. Principal Component Geostatistical Approach for large - dimensional inverse problems. Water resources research, 50(7): 5428-5443. DOI:10.1002/2013WR014630

Kuhlman, K.L., Hinnell, A.C., Mishra, P.K., Yeh, T.-C.J., 2008. Basin-scale transmissivity and storativity estimation using hydraulic tomography. Groundwater, 46(5): 706-715. DOI:10.1111/j.1745-6584.2008.00455.x

LaVenue, A.M., Pickens, J.F., 1992. Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model. Water Resources Research, 28(6): 1543-1569. DOI:10.1029/92WR00208

Lavenue, M., de Marsily, G., 2001. Three-dimensional interference test interpretation in a fractured aquifer using the pilot point inverse method. Water Resources Research, 37(11): 2659-2675. DOI:10.1029/2000WR000289

Li, W., Andreas, E., A., C.O., Harry, V., 2008. Three-dimensional geostatistical inversion of flowmeter and pumping test data. Groundwater, 46(2): 193-201. DOI:doi:10.1111/j.1745-6584.2007.00419.x

Liu, S., Yeh, T.C.J., Gardiner, R., 2002. Effectiveness of hydraulic tomography: Sandbox experiments. Water Resour. Res., 38(4). DOI:10.1029/2001WR000338

Liu, X., Illman, W.A., Craig, A.J., Zhu, J., Yeh, T.C.J., 2007. Laboratory sandbox validation of transient hydraulic tomography. Water Resour. Res., 43(5). DOI:10.1029/2006wr005144

Lu, Z., Vesselinov, V.V., Lei, H., 2018. Identifying arbitrary parameter zonation using multiple level set functions. Journal of Computational Physics, 364: 257-273. DOI:10.1016/j.jcp.2018.03.016

Luo, N., Illman, W.A., 2016. Automatic estimation of aquifer parameters using long-term water supply pumping and injection records. Hydrogeol. J, 24(6): 1443-1461. DOI:10.1007/s10040-016-1407-x

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. DOI: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. DOI:10.1016/j.jhydrol.2017.09.045

Mao, D., Liu, Z., Wang, W., Li, S., Gao, Y., Xu, Z., Zhang, C., 2018. An application of hydraulic tomography to a deep coal mine: Combining traditional pumping tests with water inrush incidents. Journal of Hydrology, 567: 1-11. DOI:10.1016/j.jhydrol.2018.09.058

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. Journal of the society for Industrial and Applied Mathematics, 11(2): 431-441. DOI:10.1137/0111030

Martin, P.J., Frind, E.G., 1998. Modeling a complex multi-aquifer system: The Waterloo Moraine. Groundwater, 36(4): 679-690. DOI:10.1111/j.1745-6584.1998.tb02843.x

Mathias, S.A., Butler, A.P., 2006. Linearized Richards' equation approach to pumping test analysis in compressible aquifers. Water Resources Research, 42(6). DOI:10.1029/2005WR004680

McCall, W., Christy, T.M., Evald, M.K., 2017. Applying the HPT-GWS for hydrostratigraphy, water quality and aquifer recharge investigations. Groundwater Monitoring & Remediation, 37(1): 78-91. DOI:10.1111/gwmr.12193

Mishra, P.K., Neuman, S.P., 2011. Saturated-unsaturated flow to a well with storage in a compressible unconfined aquifer. Water Resources Research, 47(5). DOI:10.1029/2010WR010177

Moench, A.F., 1997. Flow to a well of finite diameter in a homogeneous, anisotropic water

table aquifer. Water Resources Research, 33(6): 1397-1407. DOI:10.1029/97WR00651

Molz, F.J., Boman, G., Young, S.C., Waldrop, W., 1994. Borehole flowmeters: Field application and data analysis. Journal of Hydrology, 163(3-4): 347-371. DOI:10.1016/0022-1694(94)90148-1

Molz, F.J., Morin, R.H., Hess, A.E., Melville, J.G., Güven, O., 1989. The impeller meter for measuring aquifer permeability variations: evaluation and comparison with other tests. Water Resources Research, 25(7): 1677-1683. DOI:10.1029/WR025i007p01677

Molz, F.J., Young, S.C., 1993. Development and application of borehole flowmeters for environmental assessment. The Log Analyst, 34(01).

Neuman, S.P., 1973. Calibration of distributed parameter groundwater flow models viewed as a multiple-objective decision process under uncertainty. Water Resources Research, 9(4): 1006-1021. DOI:10.1029/WR009i004p01006

Neuman, S.P., 1974. Effect of partial penetration on flow in unconfined aquifers considering delayed gravity response. Water Resources Research, 10(2): 303-312. DOI:10.1029/WR010i002p00303

Neuman, S.P., 1987. On Methods of Determining Specific Yield. Groundwater, 25(6): 679-684. DOI:10.1111/j.1745-6584.1987.tb02208.x

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 & Technology, 43(10): 3720-3727. DOI:10.1021/es8024098

Paillet, F.L., 2000. A field technique for estimating aquifer parameters using flow log data. Groundwater, 38(4): 510-521. DOI:10.1111/j.1745-6584.2000.tb00243.x

Paillet, F.L., 2004. Borehole flowmeter applications in irregular and large-diameter boreholes. Journal of Applied Geophysics, 55(1-2): 39-59. DOI:10.1016/j.jappgeo.2003.06.004

Paradis, D., Gloaguen, E., Lefebvre, R., Giroux, B., 2016. A field proof-of-concept of tomographic slug tests in an anisotropic littoral aquifer. J. Hydrol., 536: 61-73. DOI:10.1016/j.jhydrol.2016.02.041

Paradis, D., Lefebvre, R., Morin, R.H., Gloaguen, E., 2011. Permeability profiles in granular aquifers using flowmeters in direct-push wells. Groundwater, 49(4): 534-547. DOI:10.1111/j.1745-6584.2010.00761.x

Pool, M., Carrera, J., Alcolea, A., Bocanegra, E.M., 2015. A comparison of deterministic and stochastic approaches for regional scale inverse modeling on the Mar del Plata aquifer. Journal of Hydrology, 531: 214-229. DOI:10.1016/j.jhydrol.2015.09.064

RamaRao, B.S., LaVenue, A.M., De Marsily, G., Marietta, M.G., 1995. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments. Water Resources Research, 31(3): 475-493. DOI:10.1029/94WR02258

RegionalMunicipalityofWaterloo, 2014. Region of Waterloo: hydrogeology and source water - WRAS database design manual.

Rehfeldt, K.R., Boggs, J.M., Gelhar, L.W., 1992. Field study of dispersion in a heterogeneous aquifer: 3. Geostatistical analysis of hydraulic conductivity. Water Resources Research, 28(12): 3309-3324. DOI:10.1029/92wr01758

Rehfeldt, K.R., Hufschmied, P., Gelhar, L.W., Schaefer, M.E., 1989. Measuring hydraulic conductivity with the borehole flowmeter, United States.

Robin, M., Gutjahr, A., Sudicky, E., Wilson, J., 1993. Cross-correlated random field generation with the direct Fourier transform method. Water Resources Research, 29(7): 2385-2397. DOI:10.1029/93WR00386

Rohatgi, A., 2021. WebPlotDigitizer, https://automeris.io/WebPlotDigitizer/.

Ruud, N., Kabala, Z., Molz, F., 1999. Evaluation of flowmeter-head loss effects in the flowmeter test. Journal of Hydrology, 224(1-2): 55-63. DOI:10.1016/S0022-1694(99)00119-5

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). DOI:10.1029/2006wr005522

Schöniger, A., Illman, W.A., Wöhling, T., Nowak, W., 2015. Finding the right balance between groundwater model complexity and experimental effort via Bayesian model selection. Journal of Hydrology, 531: 96-110. DOI:10.1016/j.jhydrol.2015.07.047

Seequent, 2020. Leapfrog Geo 6.0, pp. 3D Geological Modelling Software.

Sharmeen, R., Illman, W.A., Berg, S.J., Yeh, T.C.J., Park, Y.J., Sudicky, E.A., Ando, K., 2012. Transient hydraulic tomography in a fractured dolostone: Laboratory rock block experiments. Water Resources Research, 48(10). DOI:10.1029/2012WR012216

Shepherd, R.G., 1989. Correlations of Permeability and Grain Size. Groundwater, 27(5): 633-638. DOI:10.1111/j.1745-6584.1989.tb00476.x

Stallman, R.W., 1956. Numerical analysis of regional water levels to define aquifer hydrology. Eos, Transactions American Geophysical Union, 37(4): 451-460. DOI:10.1029/TR037i004p00451

Straface, S., Yeh, T.C., Zhu, J., Troisi, S., Lee, C., 2007. Sequential aquifer tests at a well field, Montalto Uffugo Scalo, Italy. Water Resour. Res., 43(7). DOI:10.1029/2006WR005287

Sudicky, E.A., 1986. A natural gradient experiment on solute transport in a sand aquifer: Spatial variability of hydraulic conductivity and its role in the dispersion process. Water Resour. Res., 22(13): 2069-2082. DOI:10.1029/WR022i013p02069

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

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

Theis, C.V., 1935. The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using groundwater storage. Am. Geophys. Union Trans., 16: 519-524. DOI:10.1029/TR016i002p00519

Tikhonov, A.N., 1963. On the regularization of ill-posed problems, Doklady Akademii Nauk. Russian Academy of Sciences, pp. 49-52.

Tong, X., Illman, W.A., Berg, S.J., Luo, N., 2021. Hydraulic tomography analysis of municipal-

well operation data with geology-based groundwater models. Hydrogeology Journal: 1-19. DOI:10.1007/s10040-021-02320-4

Tso, C.-H., Zha, Y., Jim Yeh, T.-C., Wen, J.-C., 2016. The relative importance of head, flux, and prior information in hydraulic tomography analysis. Water Resources Research, 52(1): 3-20. DOI:10.1002/2015WR017191

Wang, Y.-L., Yeh, T.-C.J., Wen, J.-C., Huang, S.-Y., Zha, Y., Tsai, J.-P., Hao, Y., Liang, Y., 2017. Characterizing subsurface hydraulic heterogeneity of alluvial fan using riverstage fluctuations. Journal of Hydrology, 547: 650-663. DOI:10.1016/j.jhydrol.2017.02.032

Williams, J., Paillet, F.L., 2002. Using flowmeter pulse tests to define hydraulic connections in the subsurface: A fractured shale example. Journal of Hydrology, 265(1-4): 100-117. DOI:10.1016/S0022-1694(02)00092-6

Wu, C.-M., Yeh, T.-C.J., Zhu, J., Lee, T.H., Hsu, N.-S., Chen, C.-H., Sancho, A.F., 2005. Traditional analysis of aquifer tests: Comparing apples to oranges? Water Resour. Res., 41(9). DOI:10.1029/2004wr003717

Xiang, J., Yeh, T.-C.J., Lee, C.-H., Hsu, K.-C., Wen, J.-C., 2009. A simultaneous successive linear estimator and a guide for hydraulic tomography analysis. Water Resour. Res., 45(2). DOI:10.1029/2008WR007180

Yeh, T.-C.J., Lee, C.-H., Hsu, K.-C., Illman, W.A., Barrash, W., Cai, X., Daniels, J., Sudicky, E., Wan, L., Li, G., Winter, C.L., 2008. A view toward the future of subsurface characterization: CAT scanning groundwater basins. Water Resources Research, 44(3). DOI:10.1029/2007wr006375

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). DOI:10.1029/2008wr007233

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. DOI:10.1029/95WR02869

Yeh, T.C.J., Lee, C.-H., 2007. Time to change the way we collect and analyze data for aquifer characterization. Groundwater, 45(2): 116-118. DOI:10.1111/j.1745-6584.2006.00292.x

Yeh, T.C.J., Liu, S., 2000. Hydraulic tomography: Development of a new aquifer test method. Water Resour. Res., 36(8): 2095-2105. DOI:10.1029/2000wr900114

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. DOI:10.1111/j.1745-6584.1993.tb00597.x

Yin, D., Illman, W.A., 2009. Hydraulic tomography using temporal moments of drawdown recovery data: A laboratory sandbox study. Water Resour. Res., 45(1). DOI:10.1029/2007WR006623

Yu, D., Yang, J., Shi, L., Zhang, Q., Huang, K., Fang, Y., Zha, Y., 2019. On the uncertainty of initial condition and initialization approaches in variably saturated flow modeling. Hydrology and Earth System Sciences, 23(7): 2897-2914. DOI:10.5194/hess-23-2897-2019

Zha, Y., Yeh, T.-C.J., Illman, W.A., Mok, C.M.W., Tso, C.-H.M., Carrera, B.A., Wang, Y.-L., 2019. Exploitation of pump-and-treat remediation systems for characterization of hydraulic heterogeneity. Journal of Hydrology, 573: 324-340. DOI:10.1016/j.jhydrol.2019.03.089

Zha, Y., Yeh, T.-C.J., Illman, W.A., Tanaka, T., Bruines, P., Onoe, H., Saegusa, H., 2015. What does hydraulic tomography tell us about fractured geological media? A field study and synthetic experiments. J. Hydrol., 531, Part 1: 17-30. DOI:10.1016/j.jhydrol.2015.06.013

Zha, Y., Yeh, T.-C.J., Mao, D., Yang, J., Lu, W., 2014. Usefulness of flux measurements during hydraulic tomographic survey for mapping hydraulic conductivity distribution in a fractured medium. Advances in Water Resources, 71: 162-176. DOI:10.1016/j.advwatres.2014.06.008

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. DOI:10.1002/2016WR019185

Zha, Y., Yeh, T.C.J., Illman, W.A., Tanaka, T., Bruines, P., Onoe, H., Saegusa, H., Mao, D., Takeuchi, S., Wen, J.C., 2016. An application of hydraulic tomography to a large-scale fractured granite site, Mizunami, Japan. Ground Water, 54(6): 793-804. DOI:10.1111/gwat.12421

Zha, Y., Yeh, T.C.J., Illman, W.A., Zeng, W., Zhang, Y., Sun, F., Shi, L., 2018. A reduced-order successive linear estimator for geostatistical inversion and its application in hydraulic tomography. Water Resources Research, 54(3): 1616-1632. DOI:10.1002/2017WR021884

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. DOI: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. DOI:10.1016/j.jhydrol.2018.02.024

Zhao, Z., Illman, W.A., 2022a. 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. DOI:10.1016/j.jhydrol.2022.128124

Zhao, Z., Illman, W.A., 2022b. Integrating hydraulic profiling tool pressure logs and hydraulic tomography for improved high-resolution characterization of subsurface heterogeneity. Journal of Hydrology: 127971. DOI:10.1016/j.jhydrol.2022.127971

Zhao, Z., Illman, W.A., Berg, S.J., 2016. On the importance of geological data for hydraulic tomography analysis: laboratory sandbox study. J. Hydrol. DOI:10.1016/j.jhydrol.2016.08.061

Zhao, Z., Illman, W.A., Yeh, T.-C.J., Berg, S.J., Mao, D., 2015. Validation of hydraulic tomography in an unconfined aquifer: A controlled sandbox study. Water Resour. Res.,

51(6): 4137-4155. DOI:10.1002/2015WR016910

Zhao, Z., 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: 129060. DOI:10.1016/j.jhydrol.2023.129060

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

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

Appendix A: Successive Linear Estimator

SLE treats the natural logarithm of *K* (ln*K*) of an aquifer as a stationary stochastic process with an unconditional mean $F = \langle lnK \rangle$ and the unconditional perturbation f, yielding $\ln K(x) = F(x) + f(x)$. The corresponding head distribution can be expressed in a similar manner and given by $\emptyset(\mathbf{x}) = H(\mathbf{x}) + h(\mathbf{x})$, where $H(\mathbf{x})$ is an unconditional mean and $h(\mathbf{x})$ is the unconditional perturbation. Yeh et al. (1996) stated that "*one possible solution that an inverse model that can produce is* $lnK(\mathbf{x})$ *and* $\emptyset(\mathbf{x})$ *fields that conditioned on the observed head and transmissivity values at sample locations and satisfy their underlying statistical properties (i.e., mean and covariance) and the governing flow equation*". By expressing the conditional *K* and $\emptyset$ fields as the sum of their conditional means and perturbations:

$$K_c(\mathbf{x}) = \langle K_c(\mathbf{x}) \rangle + f_c(\mathbf{x}) \qquad (Eq.\,\text{A. 1a})$$

$$\emptyset_c(\mathbf{x}) = \langle \emptyset_c(\mathbf{x}) \rangle + h_c(\mathbf{x}) \qquad (Eq.\,\text{A. 1b})$$

the conditional mean flow equation for steady-state condition can be expressed as:

$$\nabla \cdot [\, \langle K_c(\mathbf{x}) \rangle \nabla \langle (\emptyset_c(\mathbf{x}) \rangle] + \langle \nabla \cdot [f_c(\mathbf{x}) \nabla \mathrm{h_c}(\mathbf{x})] \rangle + \mathrm{Q}(\mathbf{x}) = 0 \qquad (Eq.\,\text{A. 2})$$

Since it is unable to correctly evaluate the second term of the left-hand side of Eq. A.2, the proposed SLE focuses on the first term by assuming that the second term is proportional to the conditional mean gradient. Thus, the conditional mean flow equation can be simplified as:

$$\nabla \cdot \left[\left\langle K_{eff}(\mathbf{x}) \right\rangle \nabla \langle (\emptyset_c(\mathbf{x}) \rangle \right] + Q(\mathbf{x}) = 0 \qquad (Eq.\,\text{A. 3})$$

where the conditional effective hydraulic conductivity $\langle K_{eff}(\mathbf{x})\rangle$ is a parameter the combines the conditional mean hydraulic conductivity $\langle K_c(\mathbf{x})\rangle$ and the ratio of the second term to the conditional mean gradient. Eq. A.3 is used to simulate steady-state head data using a finite-element groundwater flow model. The simulated head data are then compared to the observations and used as one basis for deriving *K* estimates.

The procedure of parameter estimation using SLE can be summarized as following steps:

Step 1: To derive the conditional effective *K* field that will produce a conditional mean head field in Eq. A.3, SLE starts the inversion with cokriging of observed f_i^* and h_j^* to construct a cokriged, mean-removed map of $\ln(K)$. The interpolation can be expressed as:

$$f(\mathbf{x}_0) = \sum_{i=1}^{\mathrm{n_f}} \lambda_{i0}\, f_i^*(\mathbf{x}_i) + \sum_{j=1}^{\mathrm{n_h}} \mu_{j0}\, h_j^*(\mathbf{x}_j) \qquad (Eq.\,\mathrm{A.4})$$

where $f(\mathbf{x}_0)$ is the cokriged f value at location $\mathbf{x}_0$. λ_{i0} and μ_{j0} are the cokriging weights associated with $\mathbf{x}_0$, which can be evaluated based on the unbiasedness and minimum variance of the linear estimator of the cokriging technique as follows:

$$\sum_{i=1}^{\mathrm{n_f}} \lambda_{i0}\, \mathrm{R_{ff}}(\mathbf{x}_k, \mathbf{x}_i) + \sum_{\mathrm{j}=1}^{\mathrm{n_h}} \mu_{j0}\, \mathrm{R_{hf}}(\mathbf{x}_k, \mathbf{x}_j) = \mathrm{R_{ff}}(\mathbf{x}_0, \mathbf{x}_k) \qquad (Eq.\,\mathrm{A.5a})$$

$$k = 1,2,\ldots,\mathrm{n_f}$$

$$\sum_{i=1}^{\mathrm{n_f}} \lambda_{i0}\, \mathrm{R_{hf}}(\mathbf{x}_h, \mathbf{x}_i) + \sum_{j=1}^{\mathrm{n_h}} \mu_{j0}\, \mathrm{R_{hh}}(\mathbf{x}_h, \mathbf{x}_j) = \mathrm{R_{hf}}(\mathbf{x}_0, \mathbf{x}_h) \qquad (Eq.\,\mathrm{A.5b})$$

$$h = 1,2,\ldots,\mathrm{n_h}$$

where $\mathrm{R_{ff}}$, $\mathrm{R_{hh}}$, and $\mathrm{R_{hf}}$ are covariances and cross-covariance of f and h, respectively.

Step 2: Solve the conditional mean flow Eq. A.3 to yield head distribution with cokriged *K* field at the iteration $r = 0$ or the updated *K* field when $r \geq 1$.

Step 3: Update $\ln K$ field iteratively using a linear estimator:

$$\widehat{\mathrm{Y}}_c^{(\mathrm{r}+1)}(\mathbf{x}_0) = \widehat{\mathrm{Y}}_c^{(\mathrm{r})}(\mathbf{x}_0) + \sum_{j=1}^{\mathrm{n_h}} \omega_{j0}^{(\mathrm{r})}\left[\emptyset_j^*(\mathbf{x}_j) - \emptyset_j^{(\mathrm{r})}(\mathbf{x}_j)\right] \qquad (Eq.\,\mathrm{A.}\,6)$$

where (r) and $(r+1)$ represent iteration number. $\widehat{\mathrm{Y}}_c(\mathbf{x}_0)$ is an estimated conditional mean of $\ln K$, which is equivalent to the sum of cokriged $f(x_0)$ (Eq. A.4) and *F* at $r = 0$. $\emptyset_j^*(\mathbf{x}_j)$ and $\emptyset_j(\mathbf{x}_j)$ are observed and model-generated head data at location $\mathbf{x}_j$. ω_{j0} is the weighting coefficient for the estimate at location $\mathbf{x}_0$ with respect to the head measurement at location $\mathbf{x}_j$.

The values of ω in Eq. A.6 are determined by solving the following system of equations:

$$\sum_{j=1}^{\mathrm{n_h}} \omega_{j0}^{(\mathrm{r})}\,\varepsilon_{\mathrm{hh}}^{(\mathrm{r})}(\mathbf{x}_h, \mathbf{x}_j) + \theta\delta_{\mathrm{ii}} = \varepsilon_{\mathrm{hy}}^{(\mathrm{r})}(\mathbf{x}_0, \mathbf{x}_h) \qquad (Eq.\,\mathrm{A.}\,7)$$

$$\mathrm{h} = 1,2,\ldots,\mathrm{n_h}$$

where $\varepsilon_{\mathrm{hh}}$ and $\varepsilon_{\mathrm{hy}}$ are the error covariance and error cross-covariance, respectively.

θ is a stabilizing factor, δ_{ii} is an identity matrix, and the term $\theta\delta_{\mathrm{ii}}$ is included to ensure numerical stable of inversion, especially for highly heterogeneous cases.

$\varepsilon_{\mathrm{hh}}$ and $\varepsilon_{\mathrm{hy}}$ in Eq. A.7 are approximated on the basis of the first-order analysis for a finite element groundwater flow model, in which, hydraulic head data at the *r*-th iteration can be written as a first-order Taylor series:

$$\emptyset = \widehat{\emptyset}_c^{(r)} + h^{(r)} = G\left(\widehat{\mathbf{Y}}_c^{(r)} + \mathbf{y}^{(r)}\right) \approx G\left(\widehat{\mathbf{Y}}_c^{(r)}\right) + \left.\frac{\partial G\left(\widehat{\mathbf{Y}}_c^{(r)}\right)}{\partial \ln K}\right|_{\widehat{\mathbf{Y}}_c^{(r)}} \mathbf{y}^{(r)} \qquad (Eq.\,\mathrm{A.\,8a})$$

$$h^{(r)} \approx \left.\frac{\partial G\left(\widehat{\mathbf{Y}}_c^{(r)}\right)}{\partial \ln K}\right|_{\widehat{\mathbf{Y}}_c^{(r)}} \mathbf{y}^{(r)} = \mathbf{J}^{(r)}\mathbf{y}^{(r)} \qquad (Eq.\,A.\,8\mathrm{b})$$

where $\mathbf{J}$ is the Jacobian matrix and can be evaluated using an adjoint state method.

The approximate covariance of $h^{(r)}$ and the cross-covariance between $h^{(r)}$ and $\mathbf{y}^{(r)}$ can then be derived using:

$$\varepsilon_{\mathrm{hh}}^{(\mathrm{r})} = \mathbf{J}^{(\mathrm{r})}\varepsilon_{\mathrm{yy}}^{(\mathrm{r})}(\mathbf{J}^{(r)})^T \qquad (Eq.\,\mathrm{A.\,9a})$$

$$\varepsilon_{\mathrm{hy}}^{(\mathrm{r})} = \mathbf{J}^{(\mathrm{r})}\varepsilon_{\mathrm{yy}}^{(\mathrm{r})} \qquad (Eq.\,\mathrm{A.\,9b})$$

where $\varepsilon_{\mathrm{yy}}^{(\mathrm{r})}$ is the error covariance of $\mathbf{y}$, which is given by:

when $r = 0$

$$\varepsilon_{\mathrm{yy}}^{(1)}(\mathbf{x}_0, \mathbf{x}_n) = \mathrm{R_{ff}}(\mathbf{x}_0, \mathbf{x}_n) - \sum_{\mathrm{i}=1}^{\mathrm{n_f}} \lambda_{i0}\,\mathrm{R_{ff}}(\mathbf{x}_i, \mathbf{x}_n) - \sum_{\mathrm{j}=1}^{\mathrm{n_h}} \mu_{j0}\,\mathrm{R_{fh}}\left(\mathbf{x}_j, \mathbf{x}_n\right) \qquad (Eq.\,\mathrm{A.\,10a})$$

when $r \geq 1$

$$\varepsilon_{yy}^{(r+1)}(\mathbf{x}_0, \mathbf{x}_n) = \varepsilon_{yy}^{(r)}(\mathbf{x}_0, \mathbf{x}_n) - \sum_{j=1}^{n_h} \omega_{i0}^{(r)} \varepsilon_{yh}^{(r)}(\mathbf{x}_i, \mathbf{x}_n) \qquad (Eq.\,\text{A.10b})$$

Step 4: Check the convergence of the inverse model with the updated $\widehat{\mathbf{Y}}_c(\mathbf{x})$. If converged, stop the iteration; otherwise, return to ***Step 3*** and start the next optimization iteration. The inverse model is considered to be converged once one of two criteria is met: (1) the absolute difference of the variance of the estimated *K* fields between two successive iterations is smaller than a prescribed tolerance; and (2) the mean squared error between the observed and simulated heads is stabilized.

Uncertainty of K estimates are evaluated using the conditional variance of estimated $\ln K$ values $\varepsilon_{yy}(\mathbf{x}_0, \mathbf{x}_0)$ (see Eq. A.10b). Smaller residual variances of $\ln K$ indicate higher confidence of the *K* estimates, and vice versa.

Appendix B: Supplementary Information for Study I

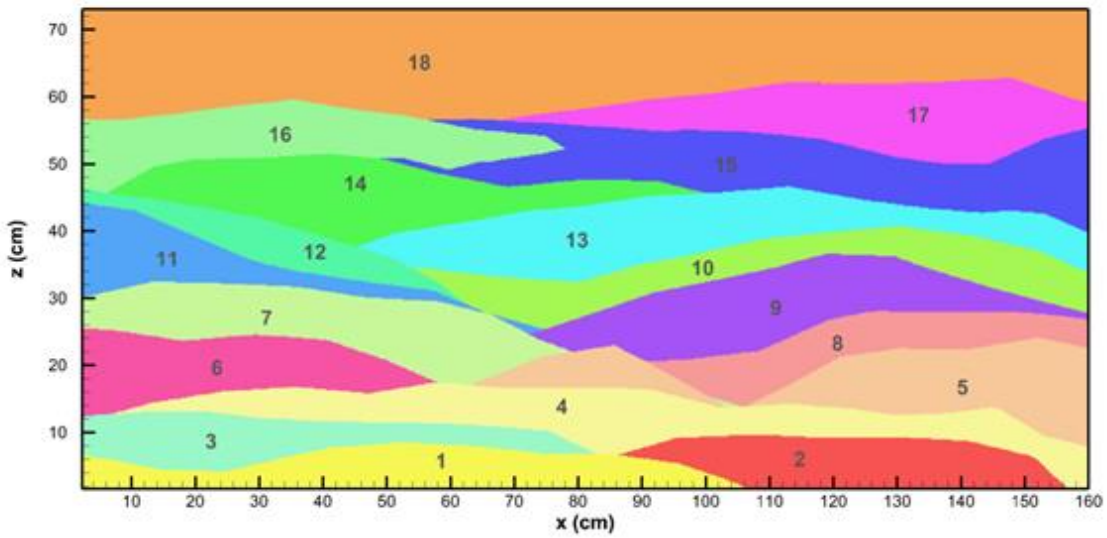

Fig. B1: "perfect" geological model of the synthetic heterogeneous aquifer constructed by delineating layer boundaries based on the photograph of the laboratory sandbox.

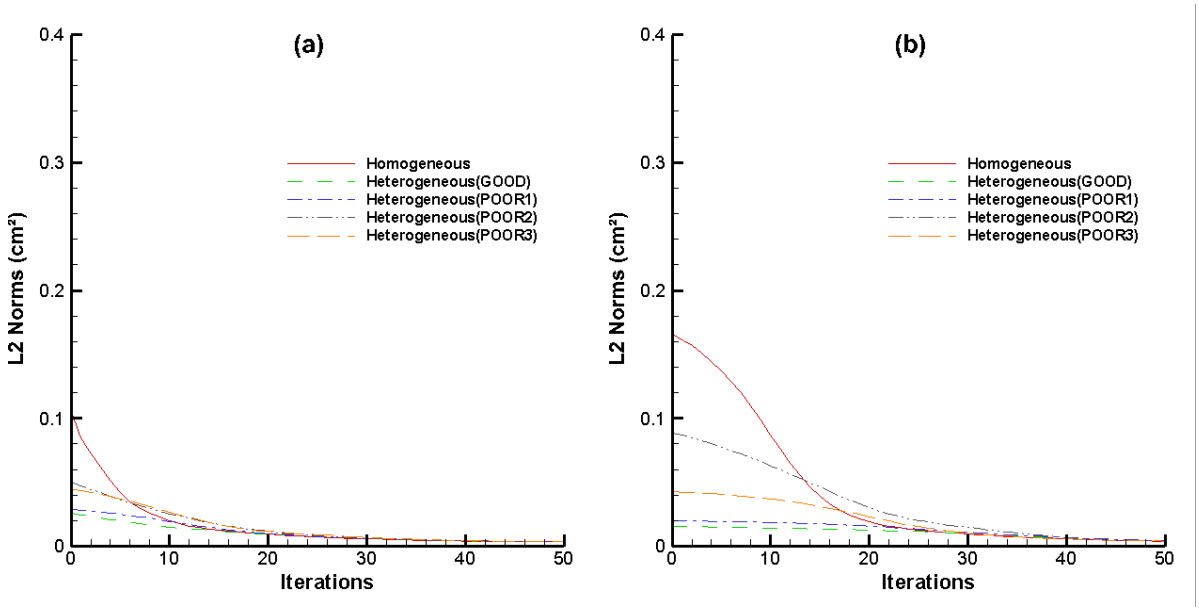

Fig. B2: L_2 norms (cm^2) as a function of iteration number when applying SimSLE for geostatistical models of different initial parameter fields. (a) Case 1, (b) Case 2.

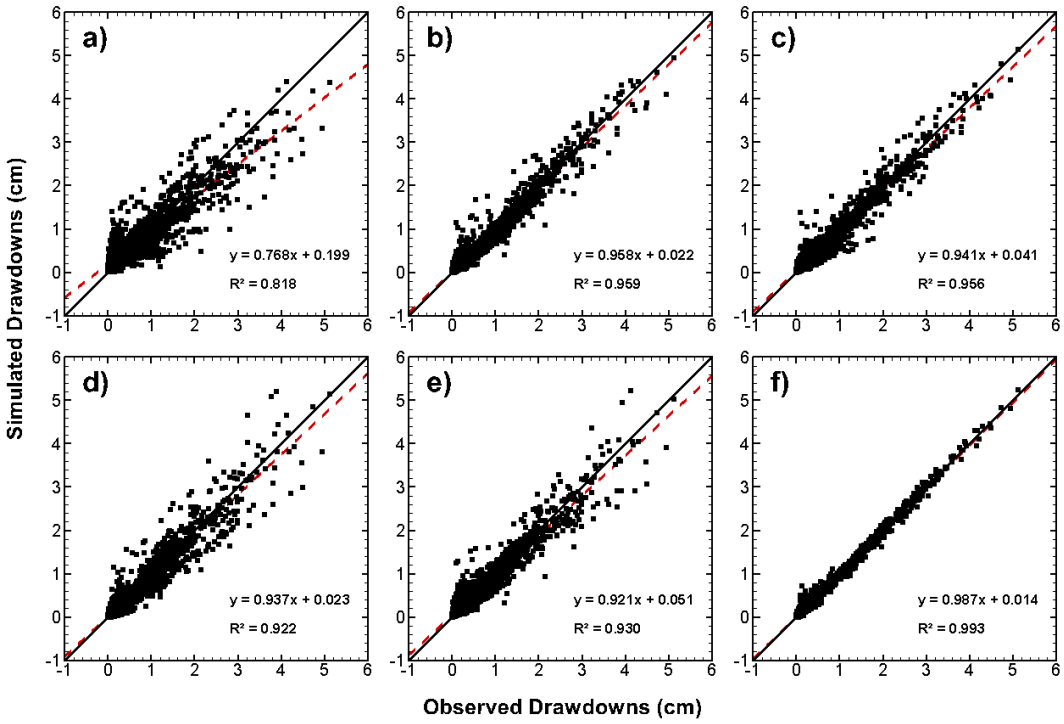

Fig. B3: Calibration scatterplots (Case 1) of simulated versus observed drawdowns for different modeling approaches. (a) effective model, (b)-(e) four geology-based zonation models: (b) GOOD, (c) POOR1, (d) POOR2, (e) POOR3; and (f) geostatistical model with homogeneous initial parameter fields.

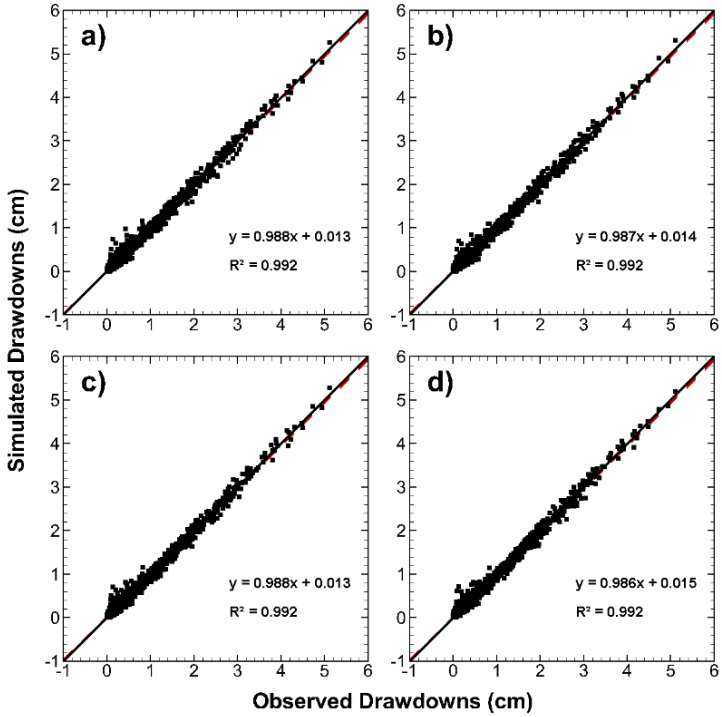

Fig. B4: Calibration scatterplots (Case 1) of simulated versus observed drawdowns for geostatistical models incorporated with four different types of geological information. (a) GOOD, (b) POOR1, (c) POOR2, and (d) POOR3.

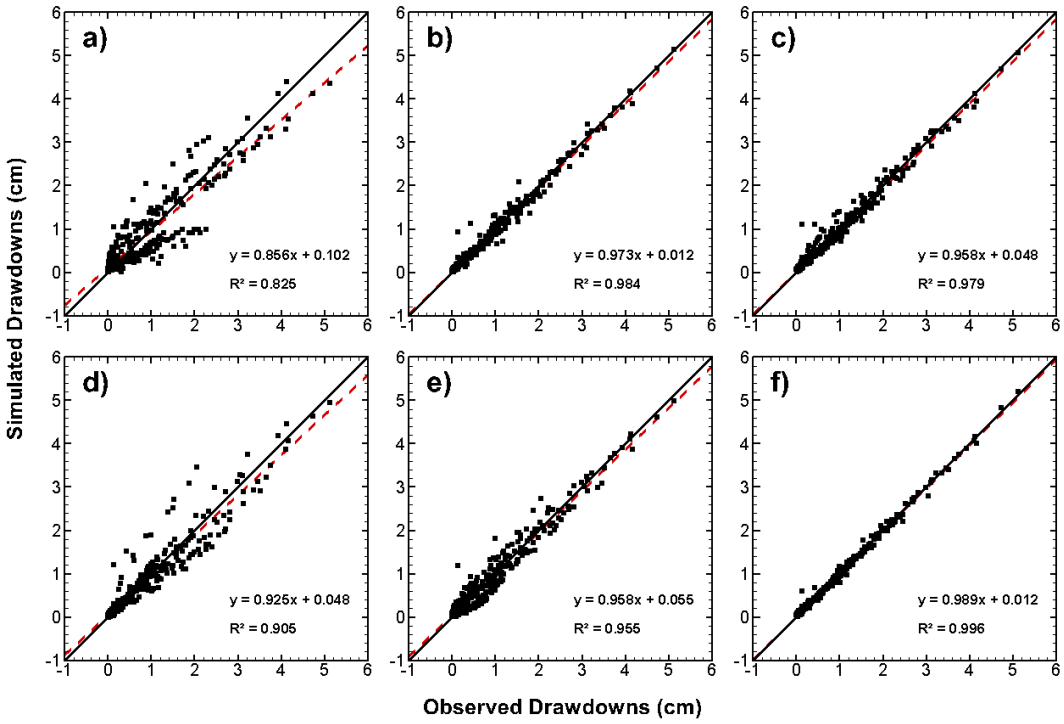

Fig. B5: Calibration scatterplots (Case 2) of simulated versus observed drawdowns for different modeling approaches. (a) effective model, (b)-(e) four geology-based zonation models: (b) GOOD, (c) POOR1, (d) POOR2, (e) POOR3; and (f) geostatistical model with homogeneous initial parameter fields.

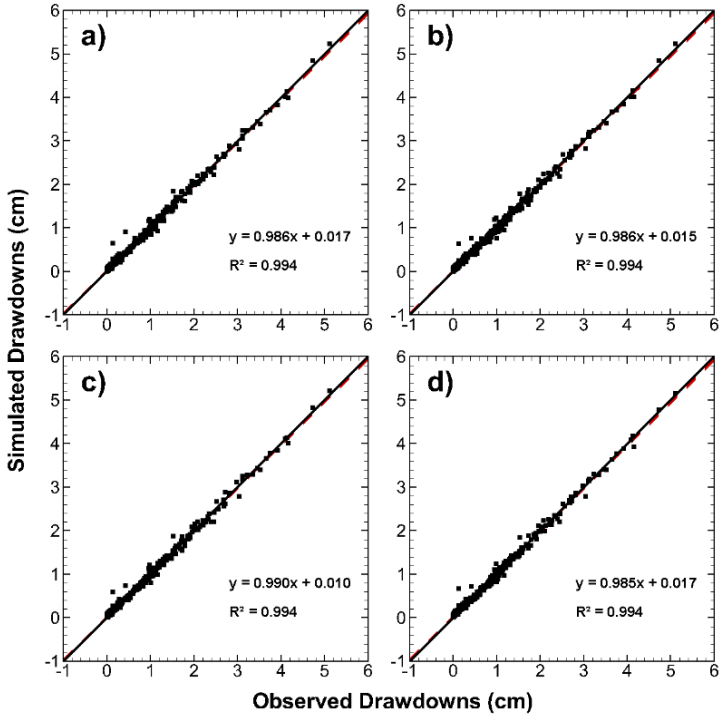

Fig. B6: Calibration scatterplots (Case 2) of simulated versus observed drawdowns for geostatistical models incorporated with four different types of geological information. (a) GOOD, (b) POOR1, (c) POOR2, and (d) POOR3.

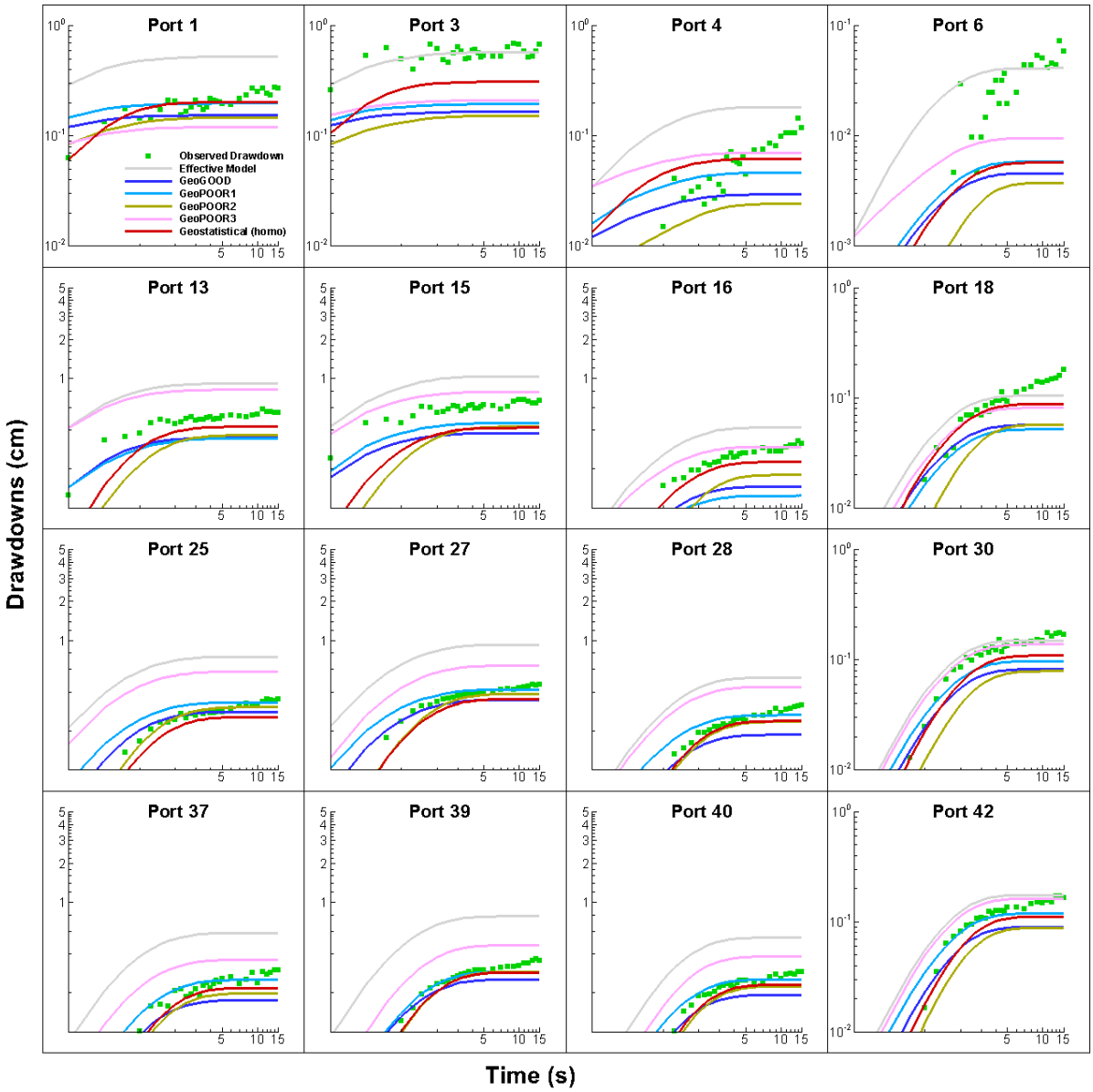

Fig. B7: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 8. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

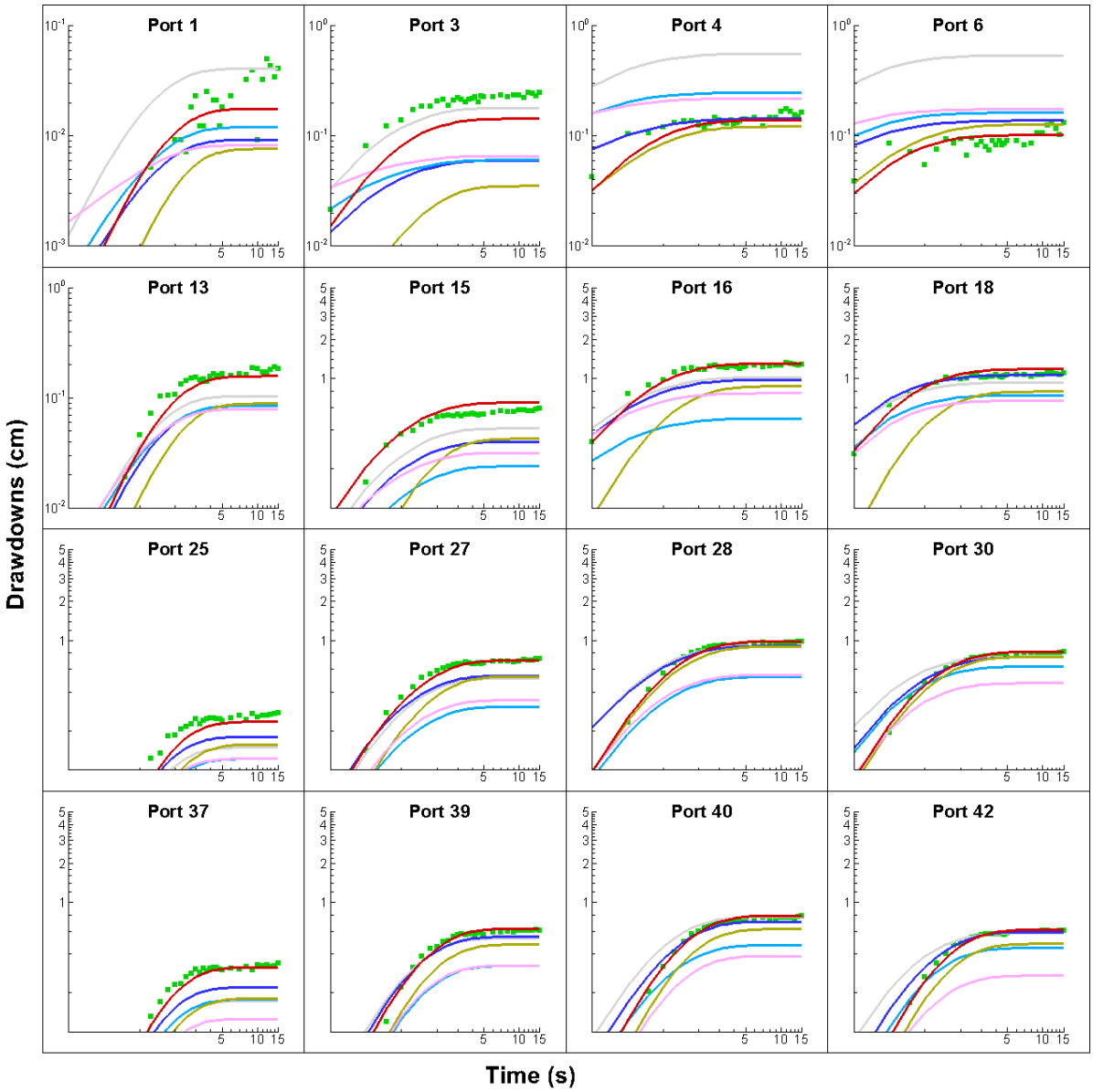

Fig. B8: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 11. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

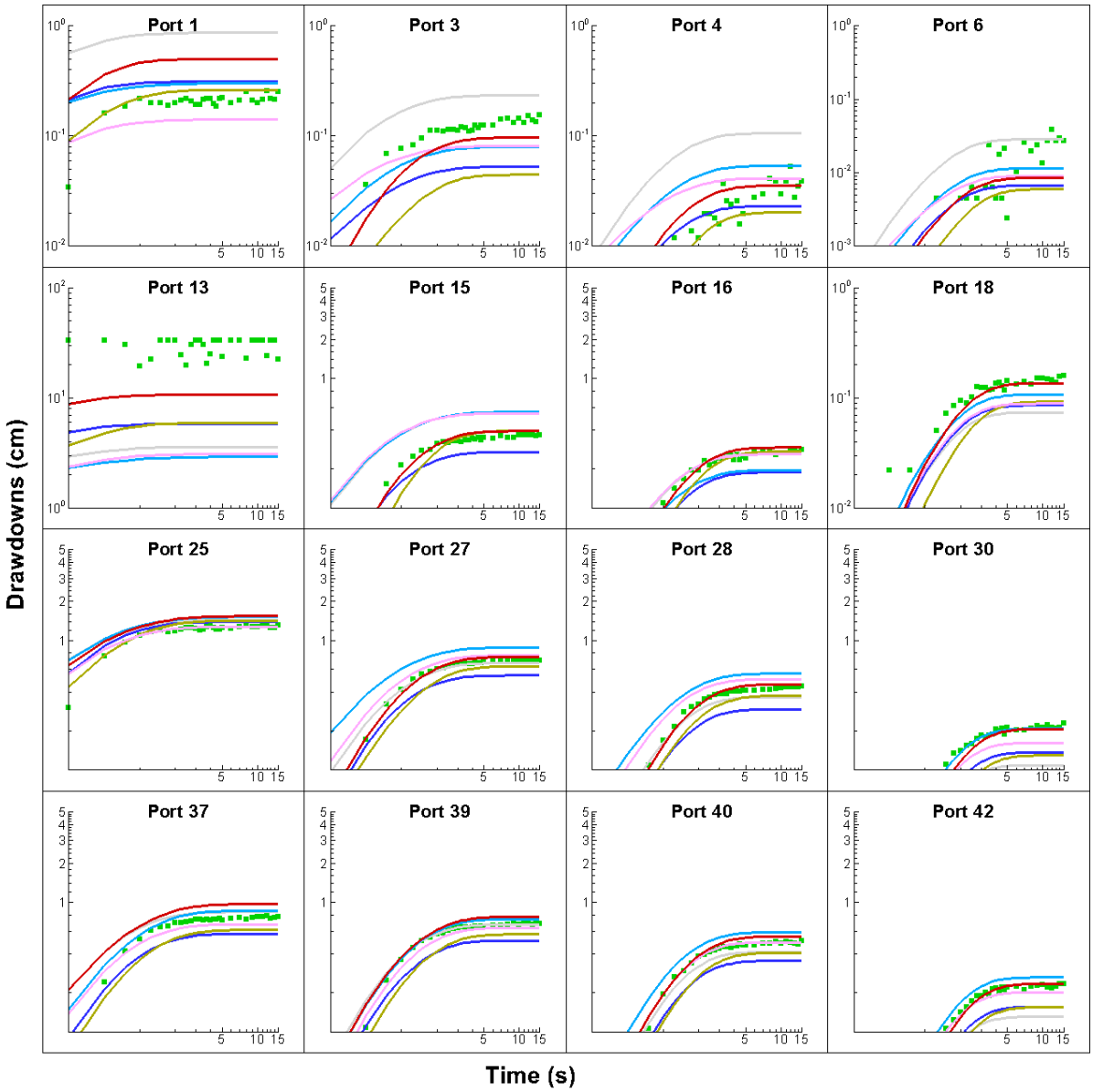

Fig. B9: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 13. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

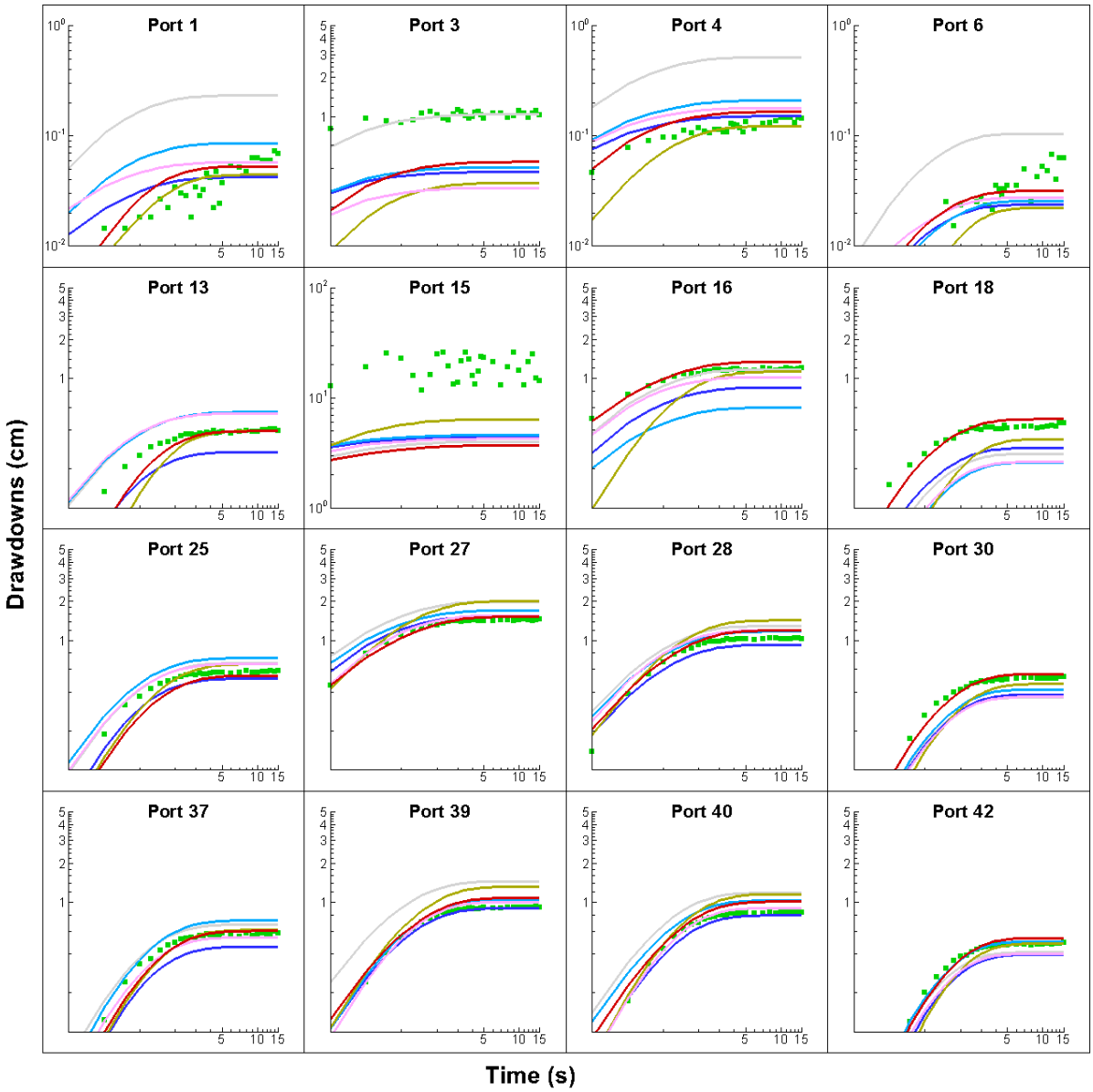

Fig. B10: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 15. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

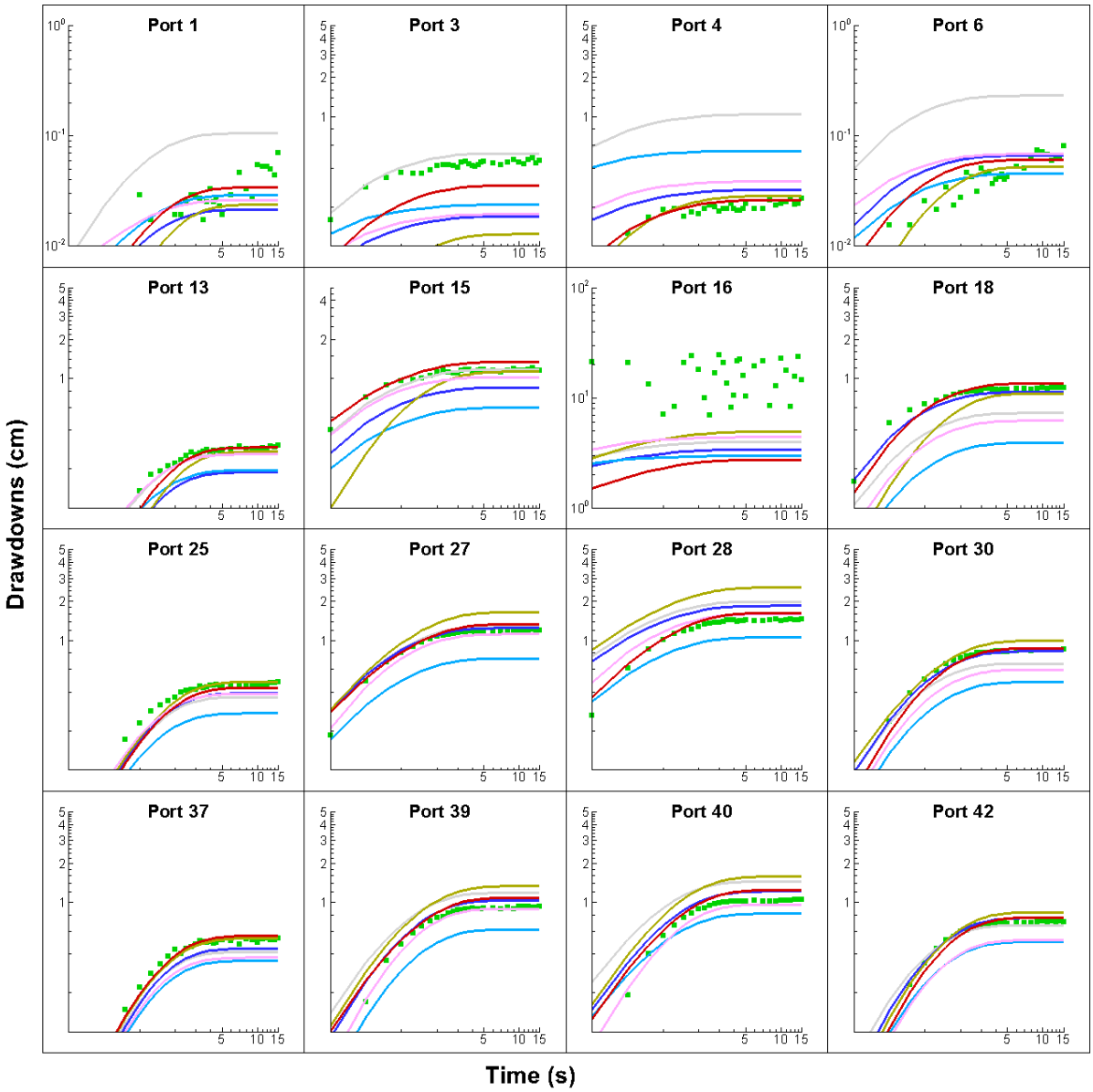

Fig. B11: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 16. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

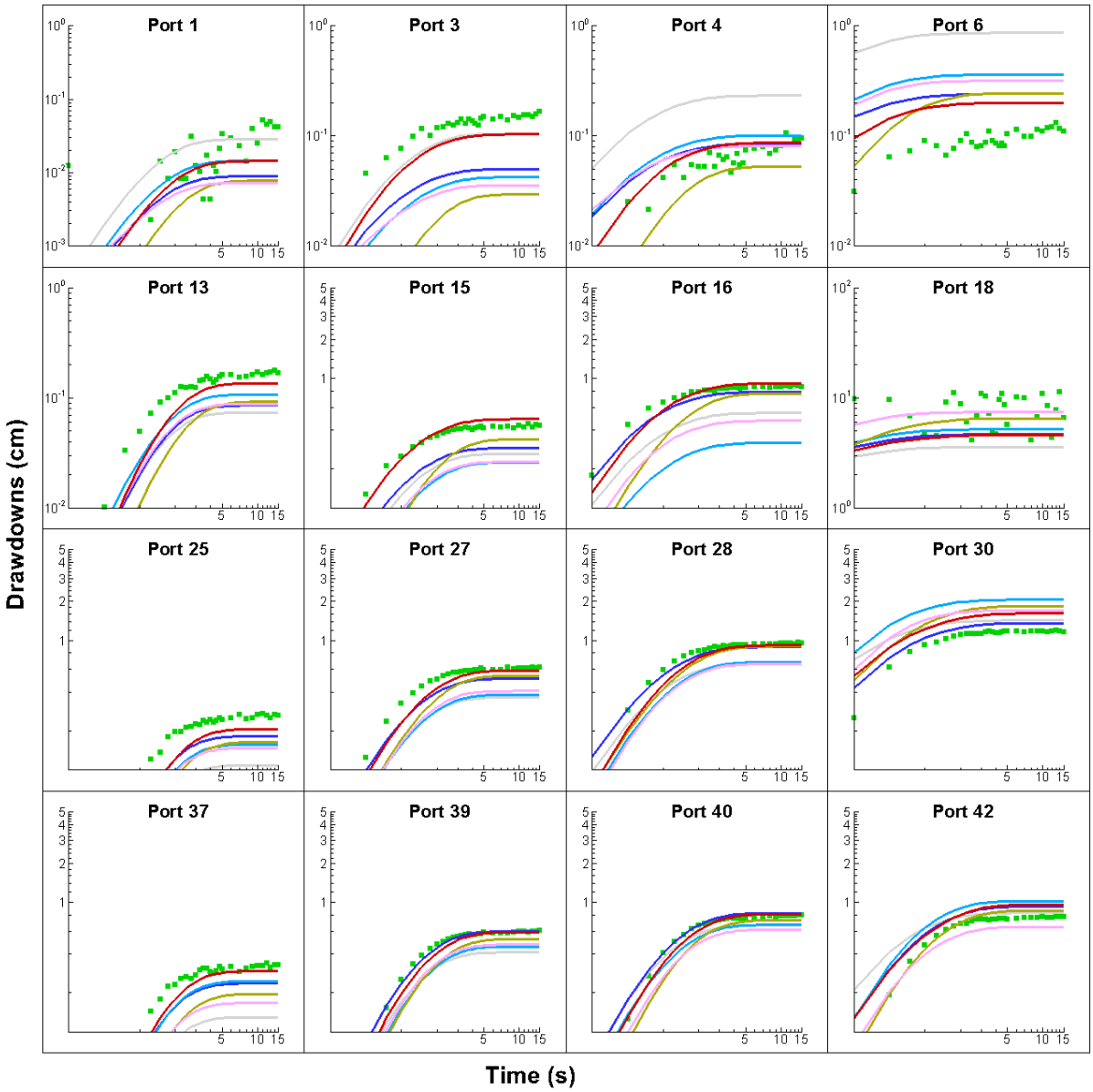

Fig. B12: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 18. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

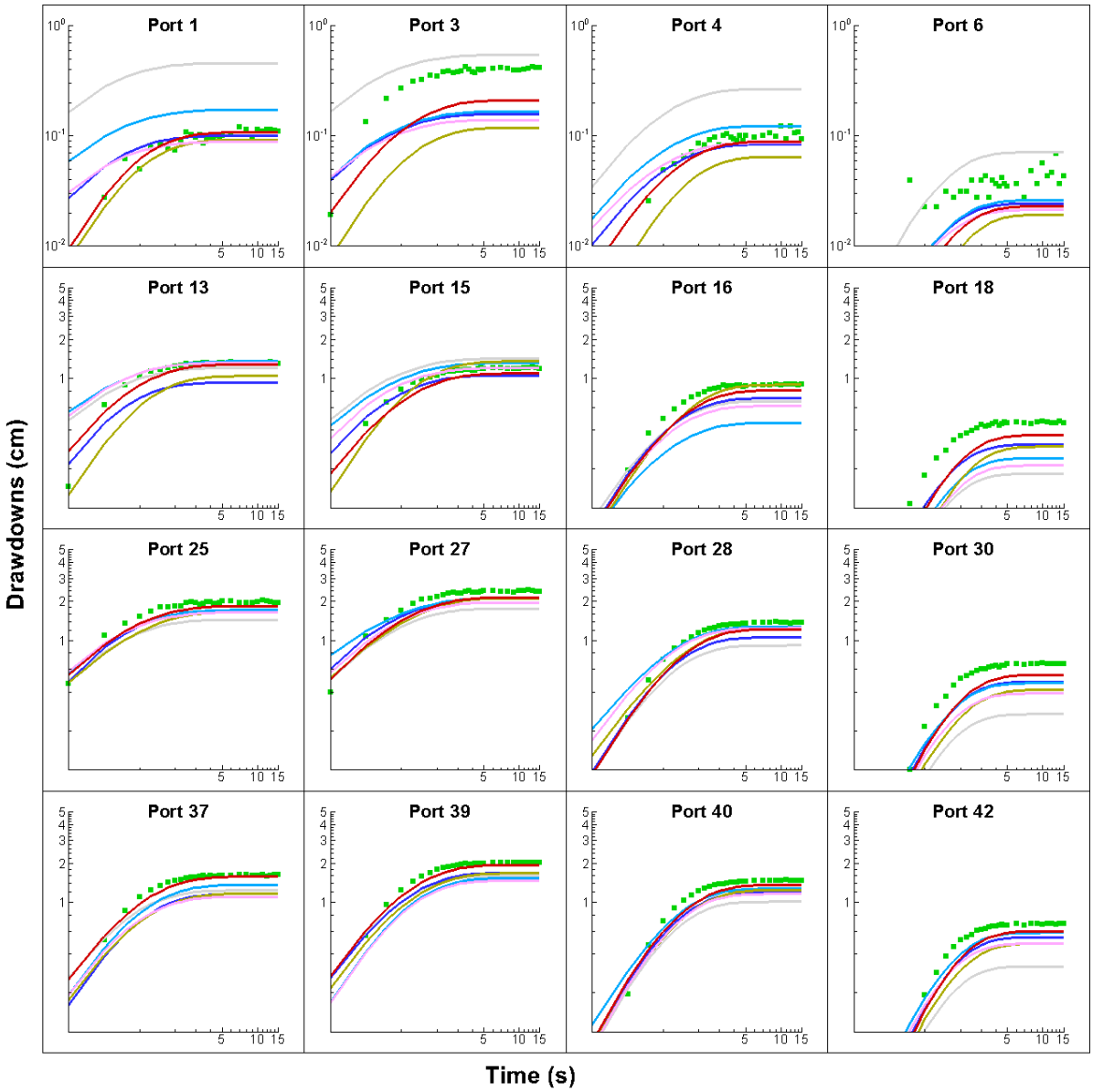

Fig. B13: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 20. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

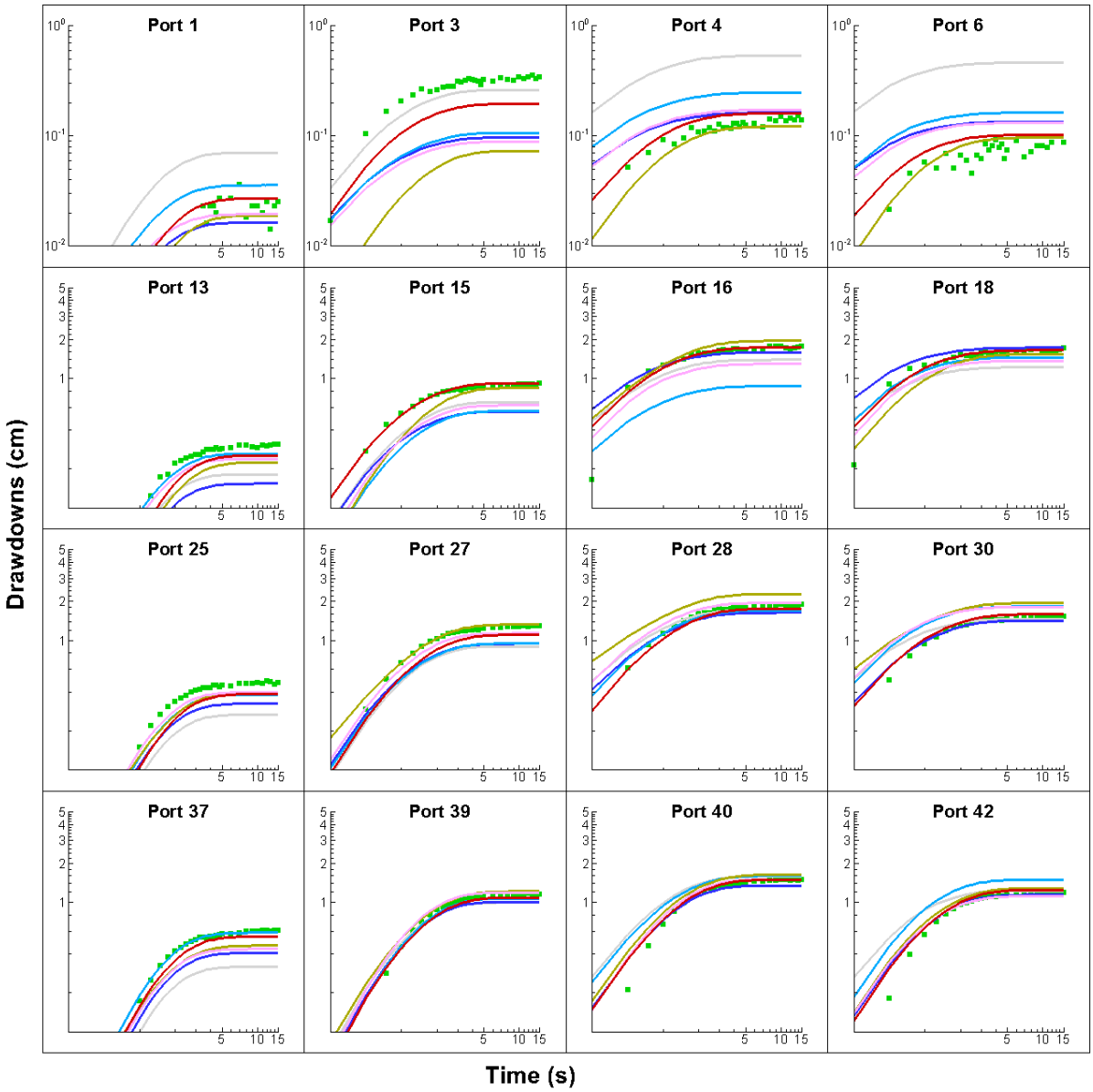

Fig. B14: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 23. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

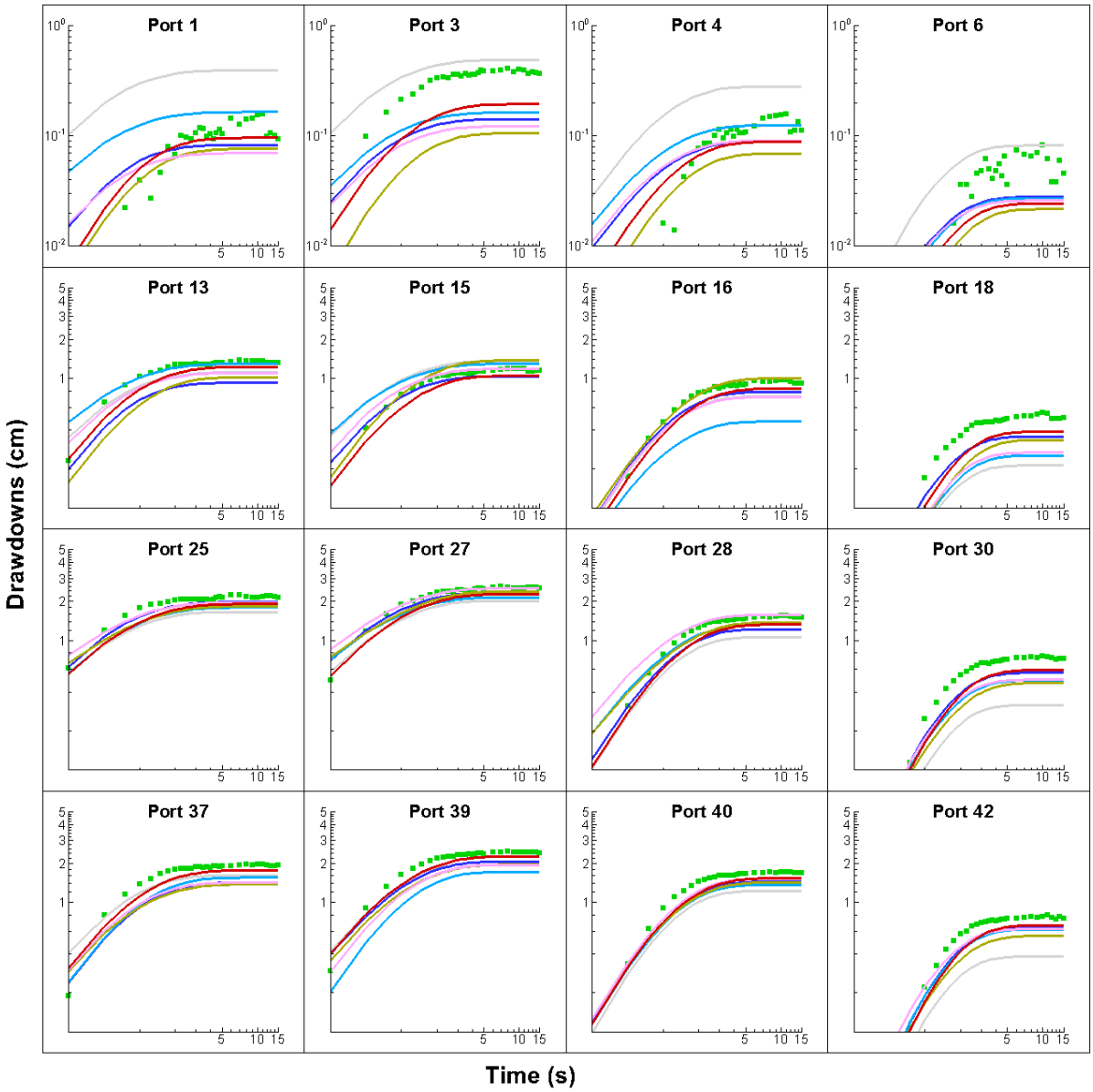

Fig. B15: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 26. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

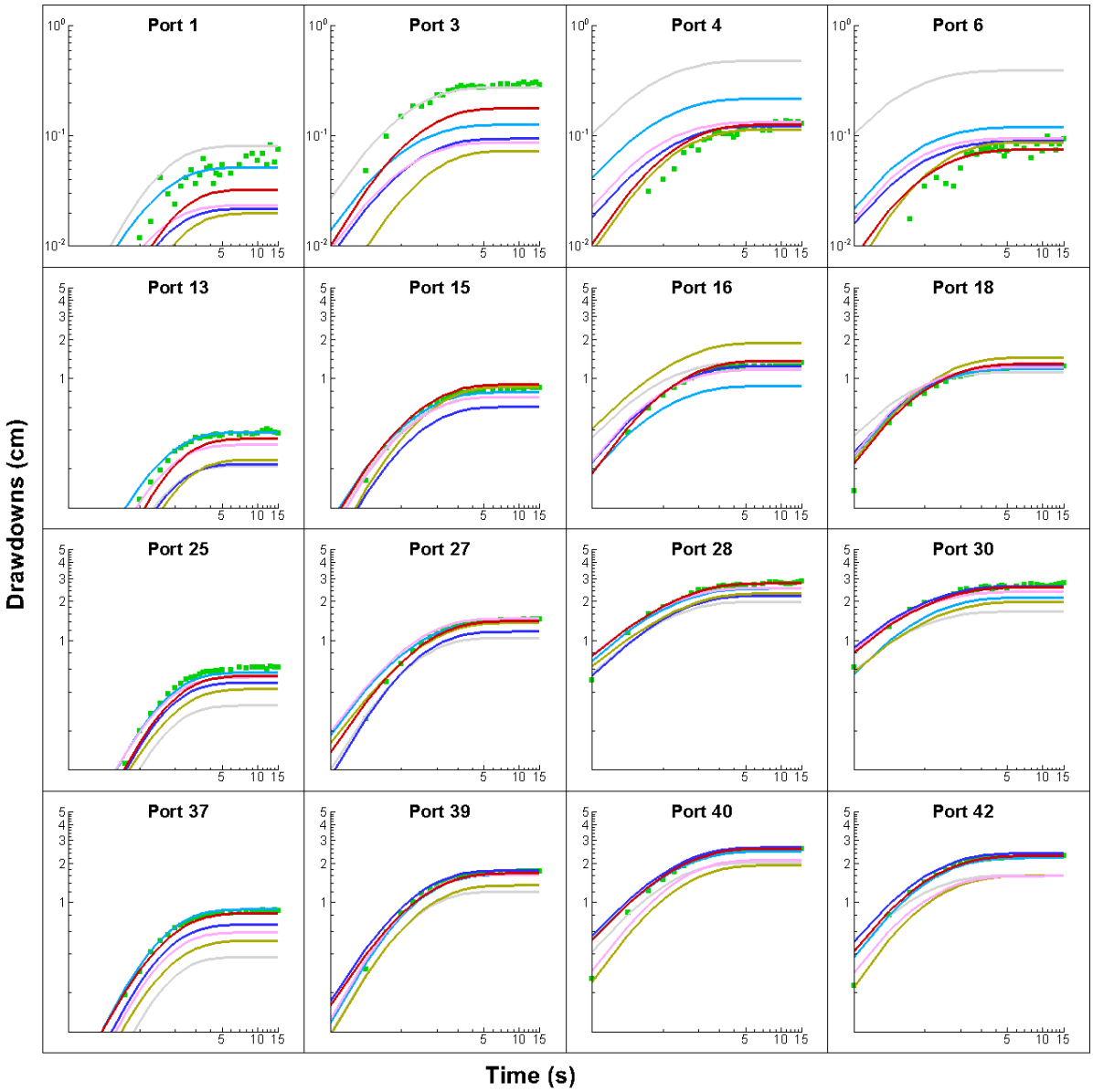

Fig. B16: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 29. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

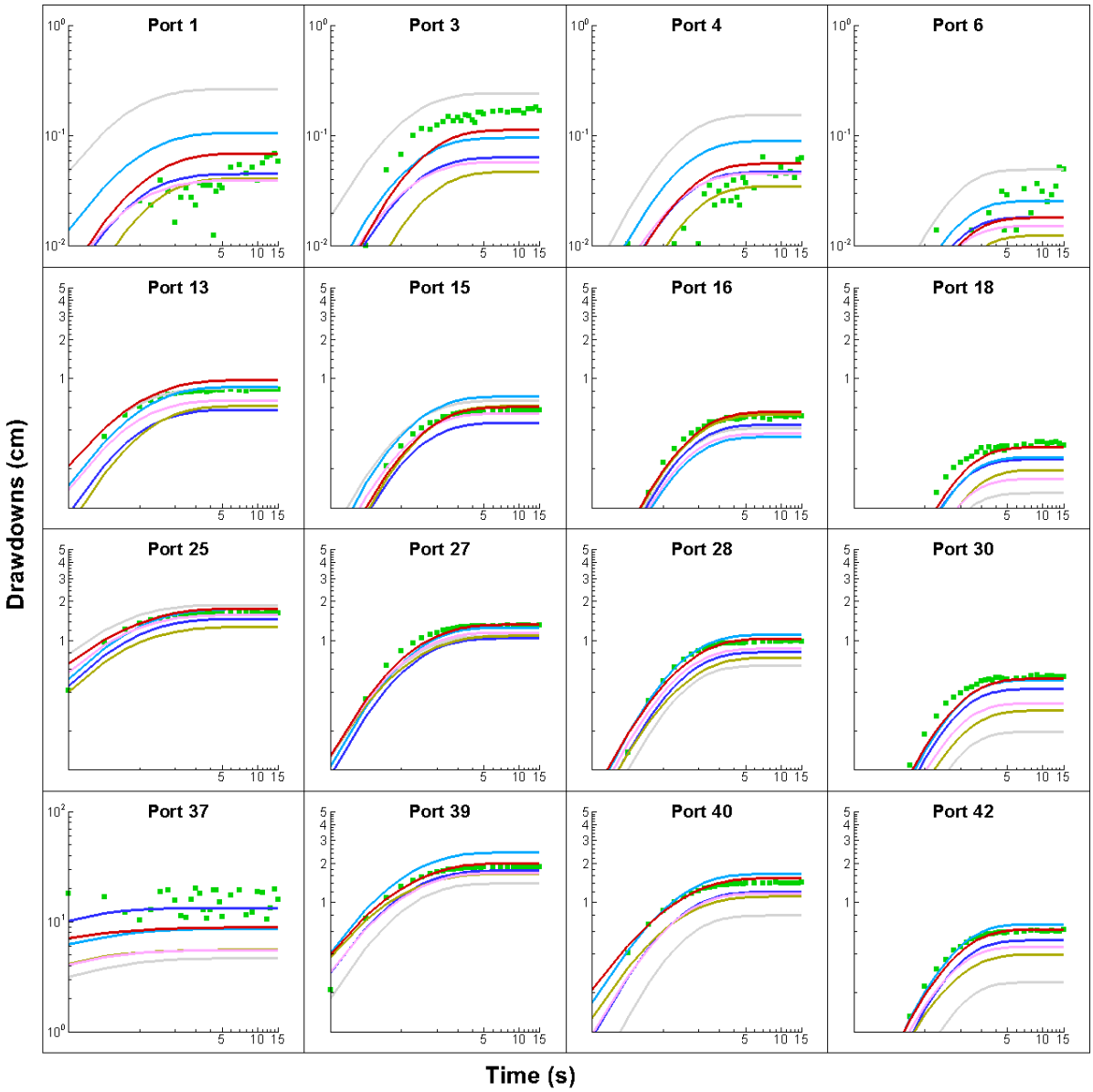

Fig. B17: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 37. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

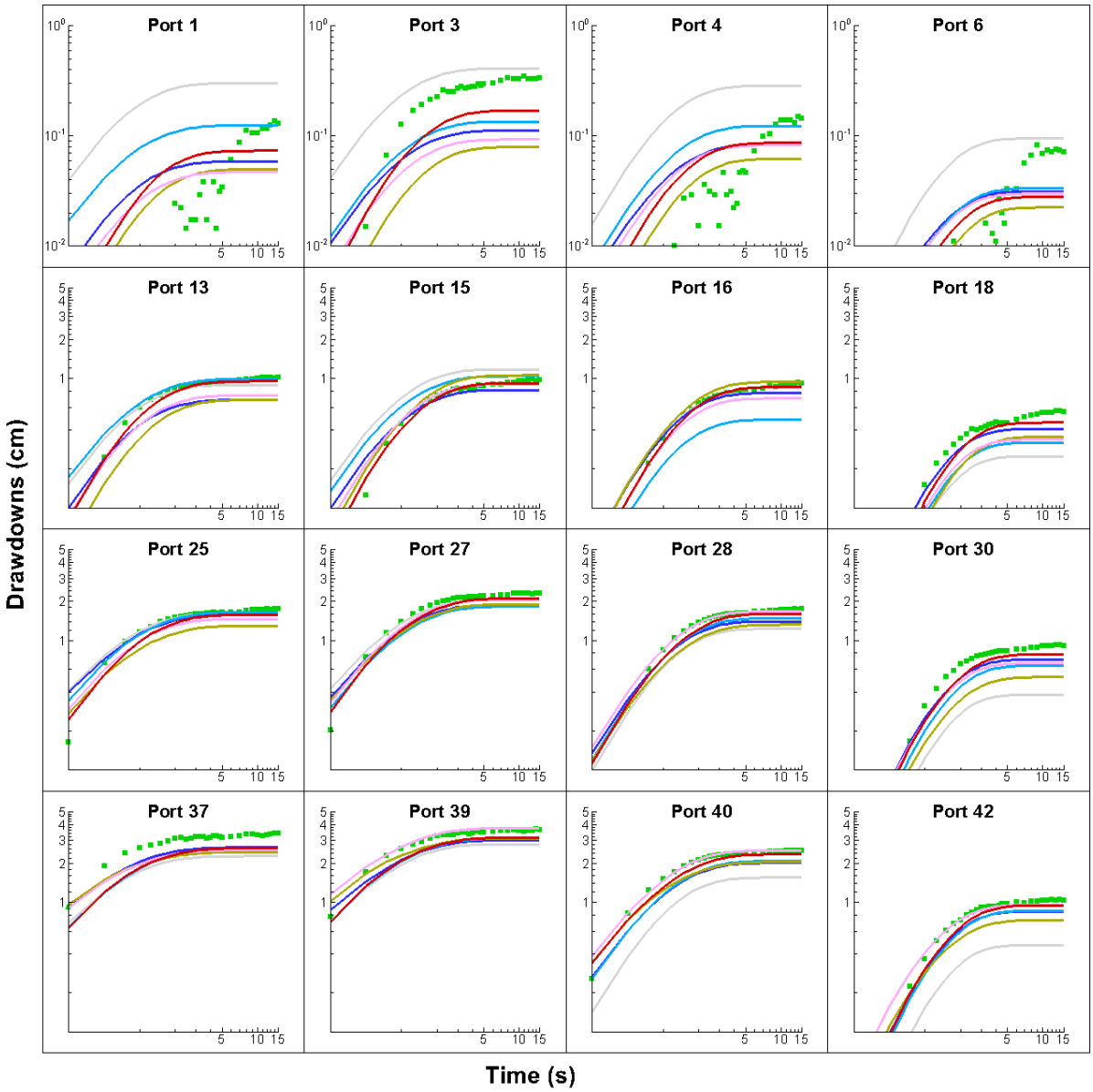

Fig. B18: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 38. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

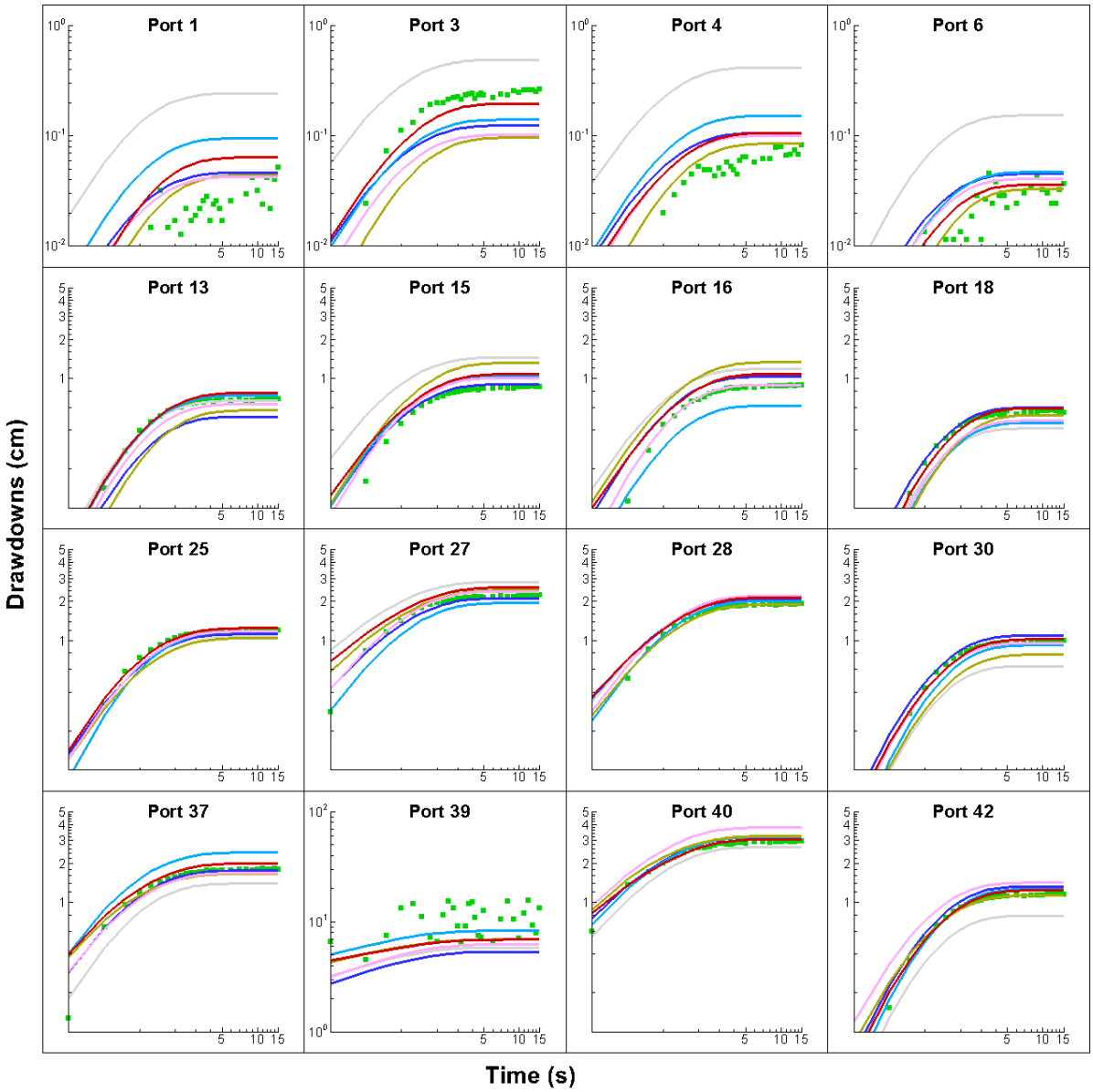

Fig. B19: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 39. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

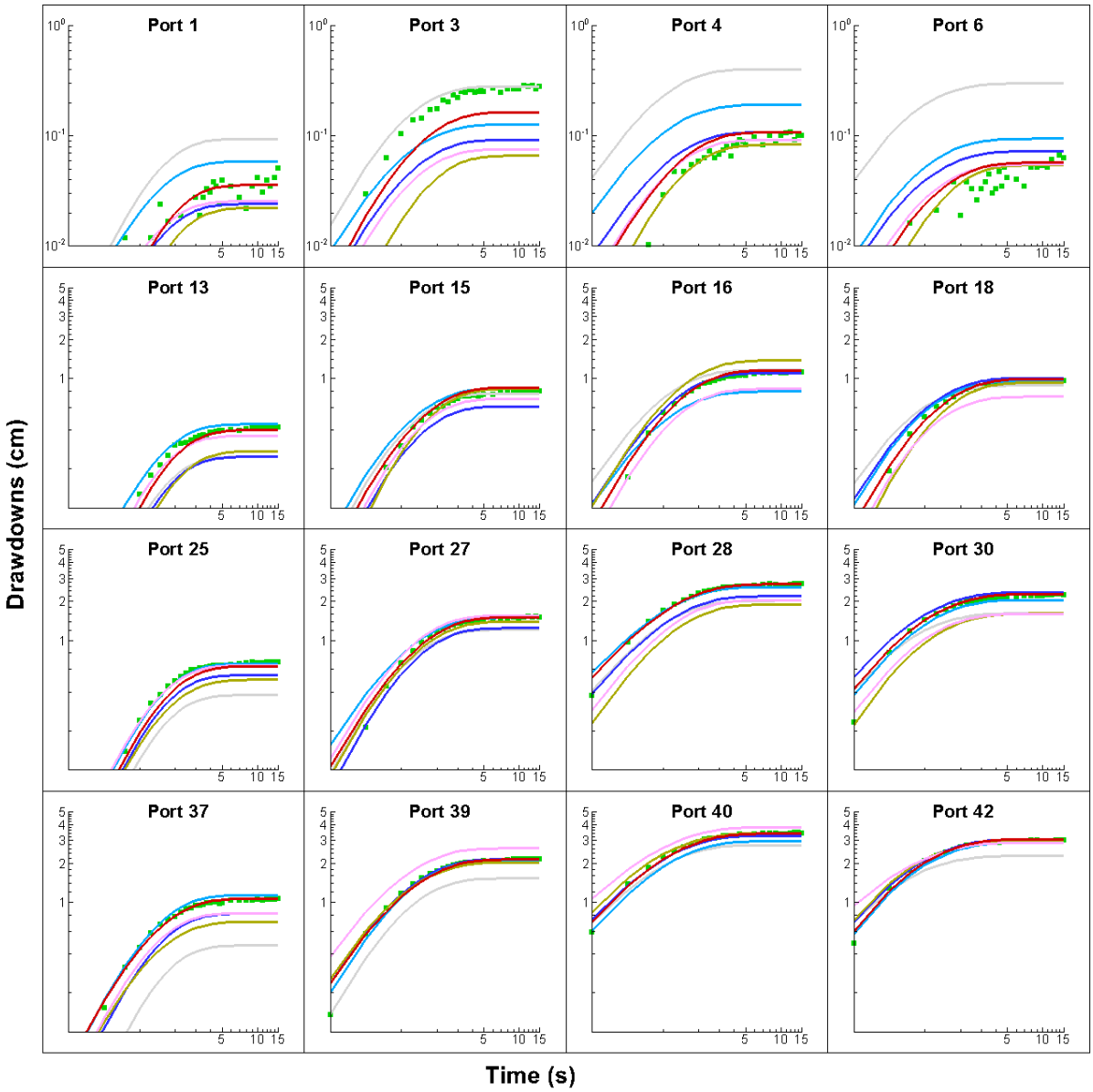

Fig. B20: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 41. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

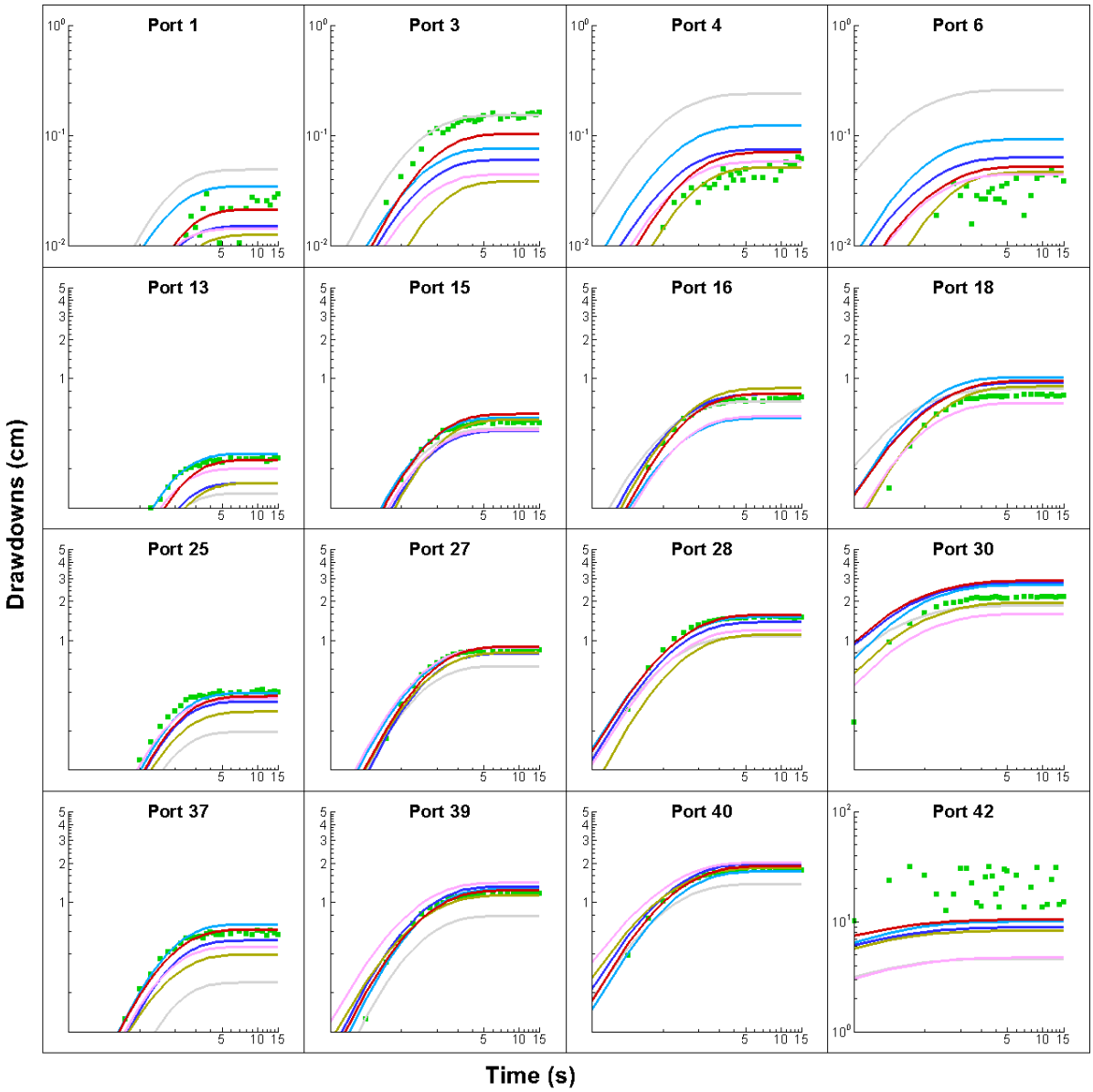

Fig. B21: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 42. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 8 pumping tests and 47 observation ports (Case 1).

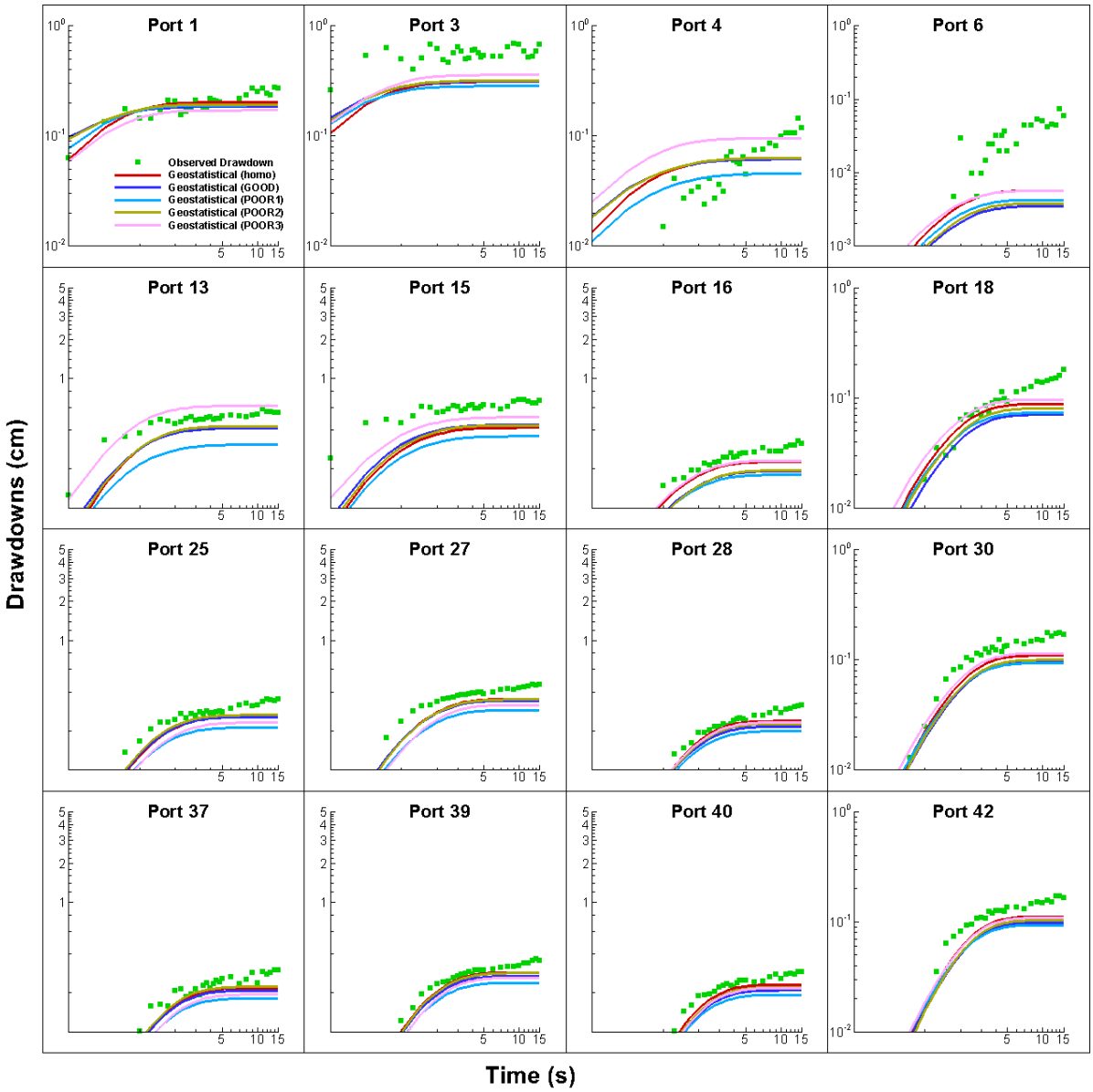

Fig. B22: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 8. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

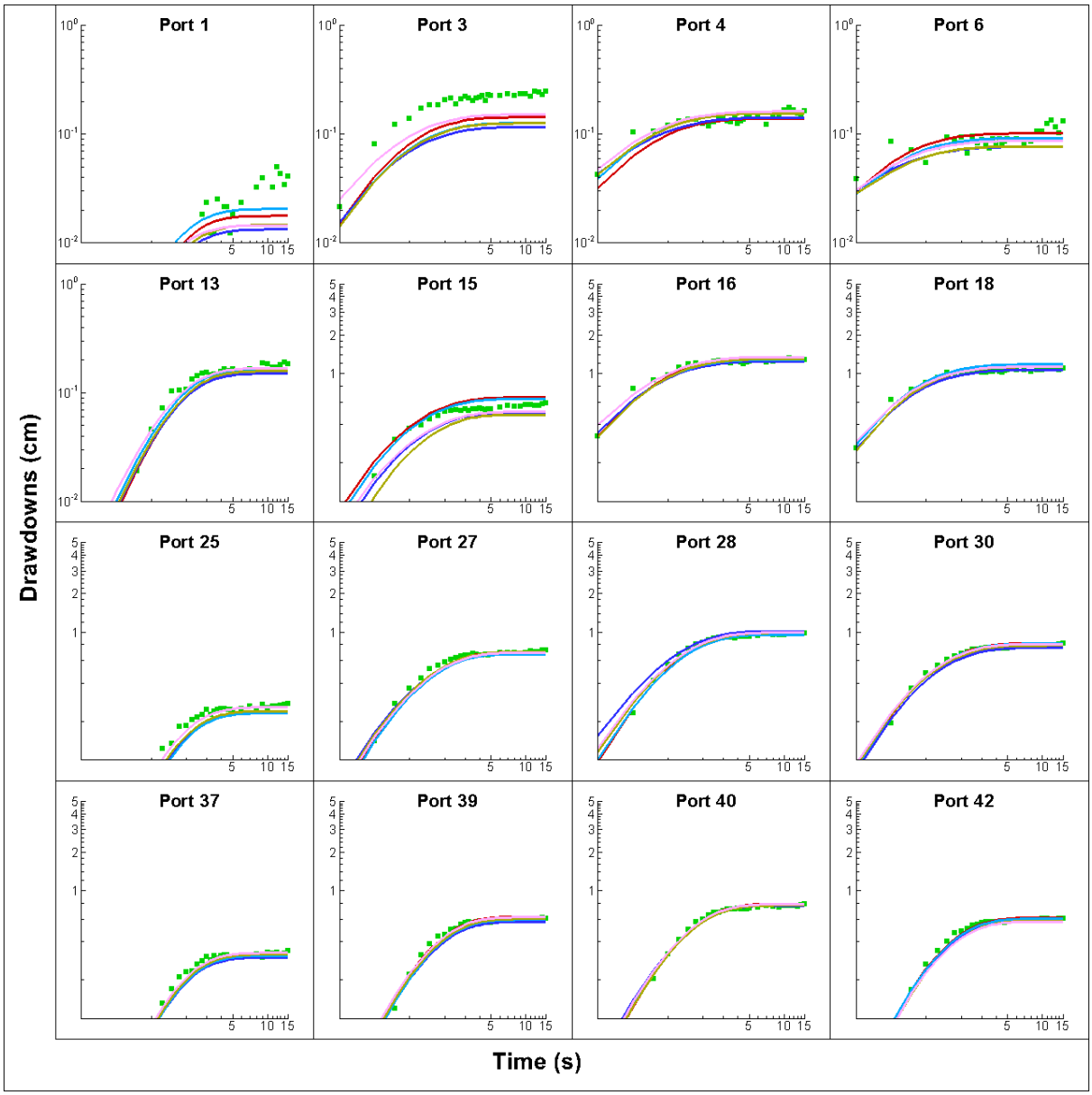

Fig. B23: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 11. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

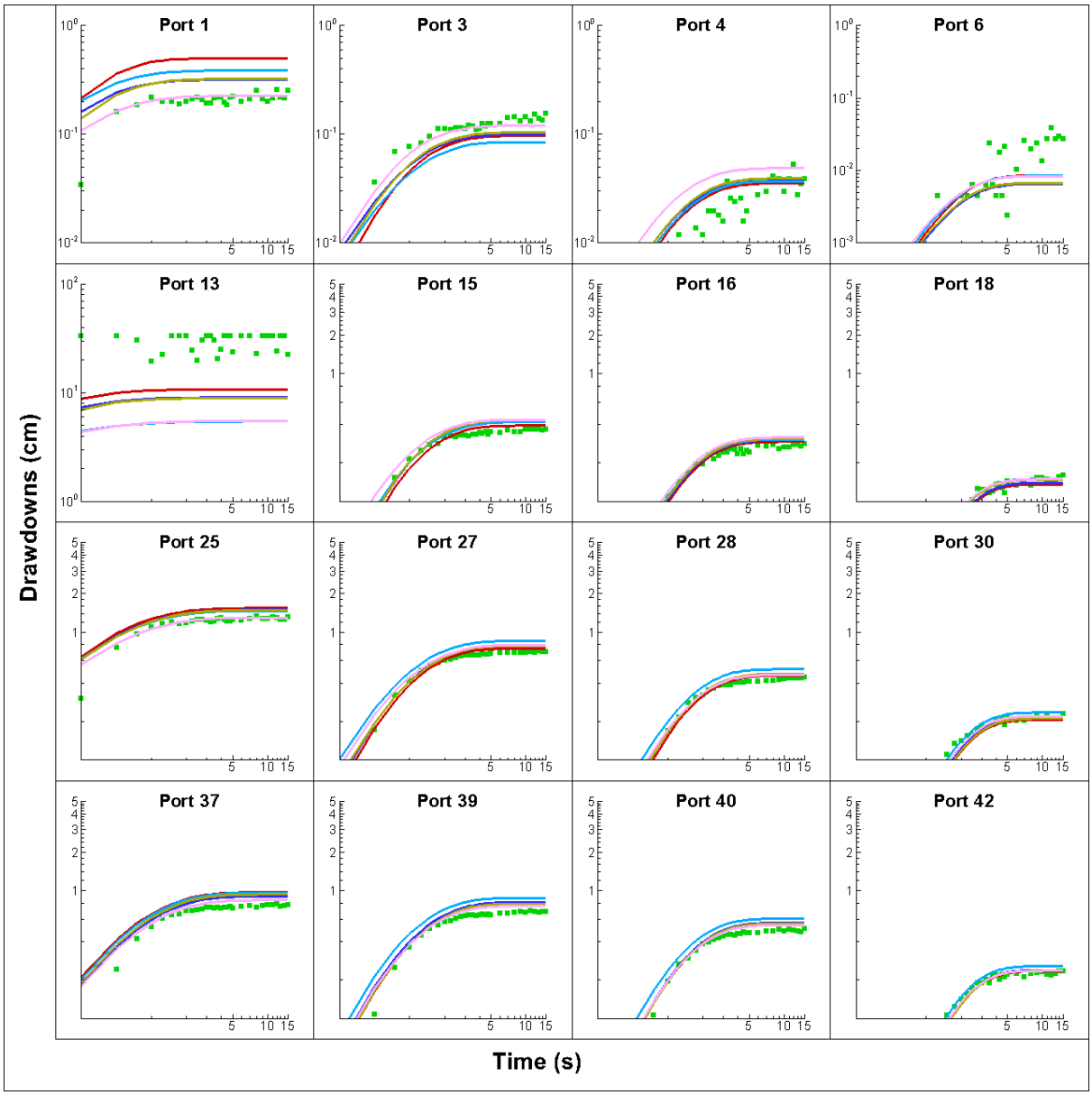

Fig. B24: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 13. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

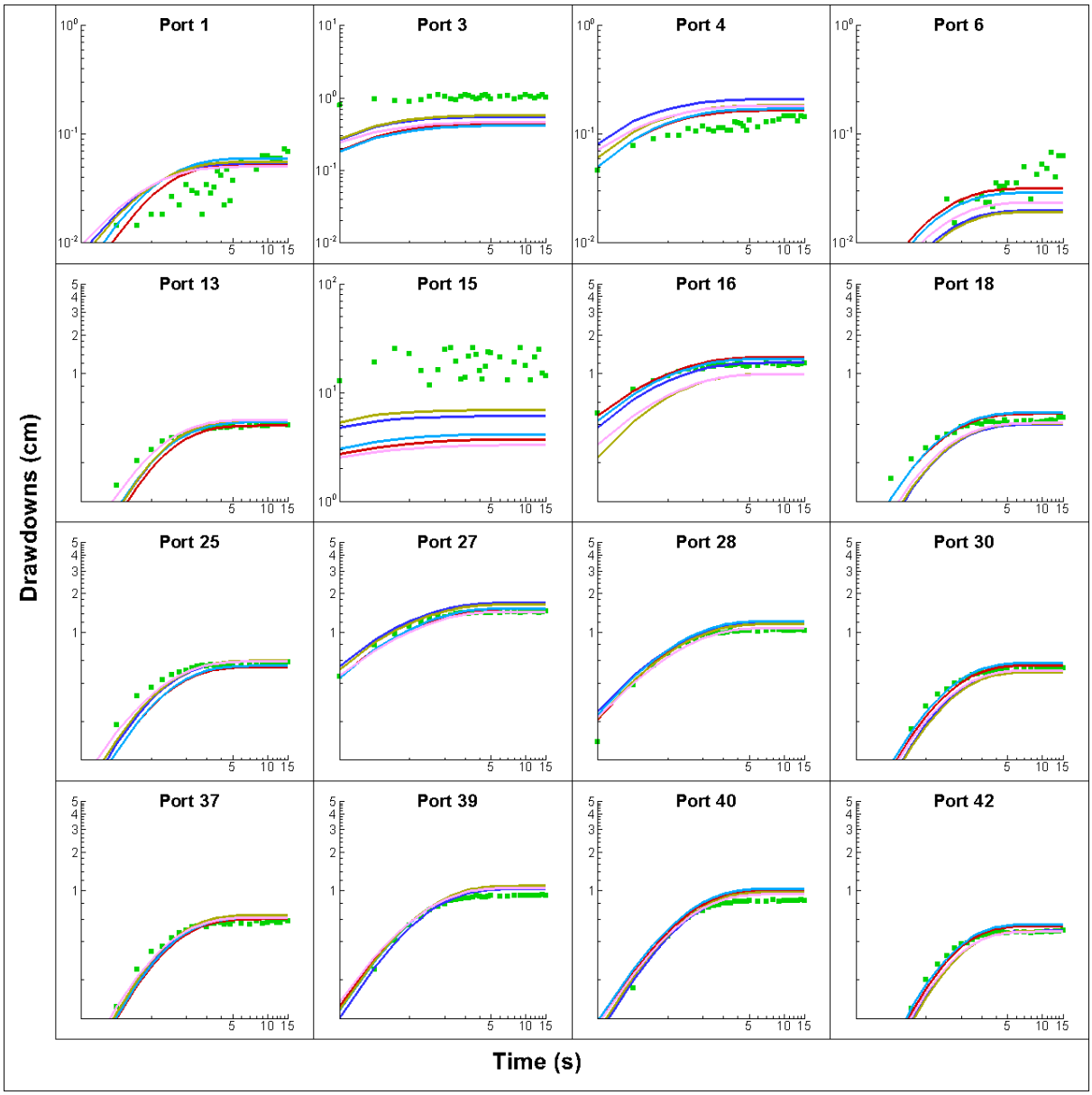

Fig. B25: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 15. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

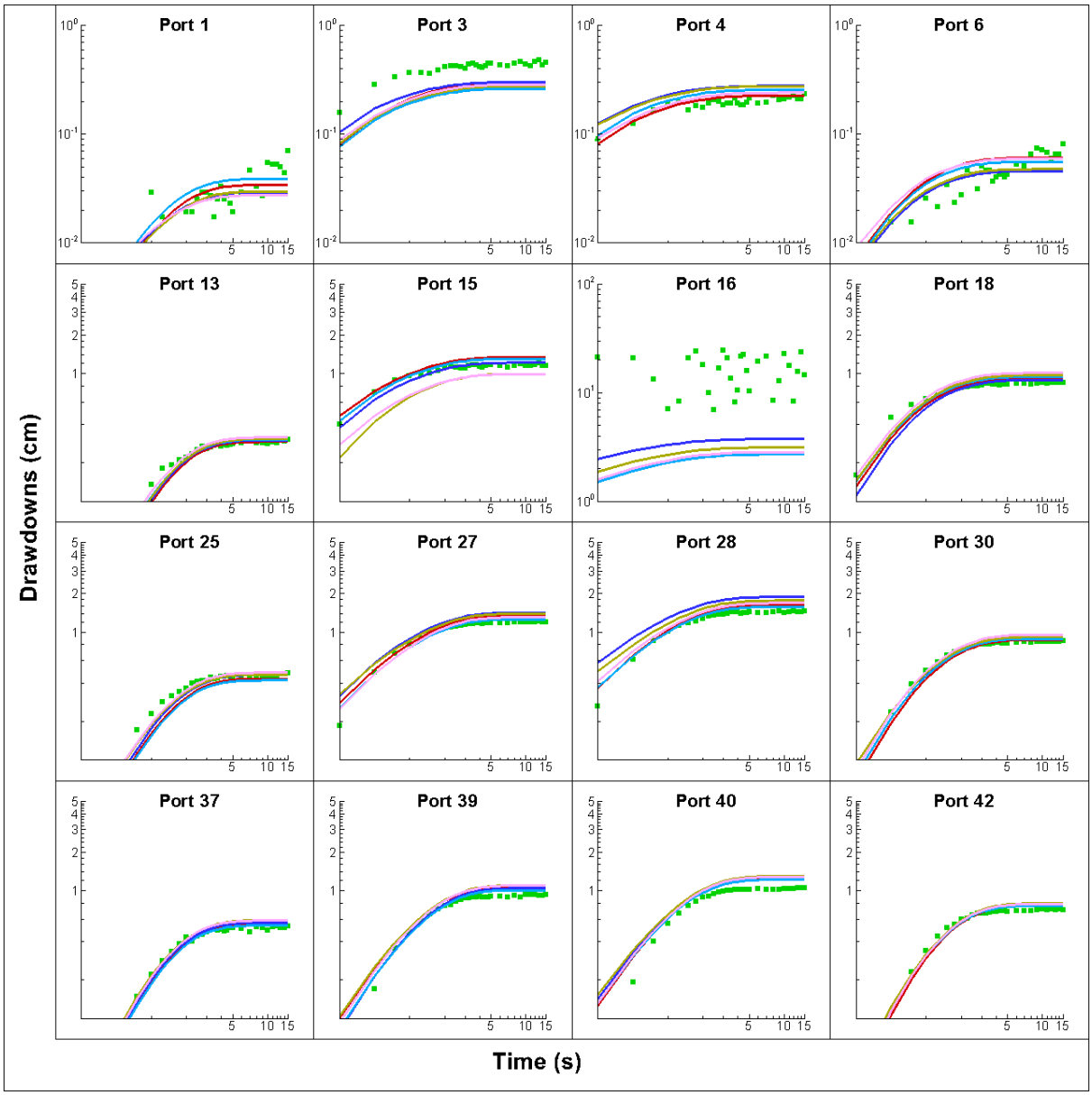

Fig. B26: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 16. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

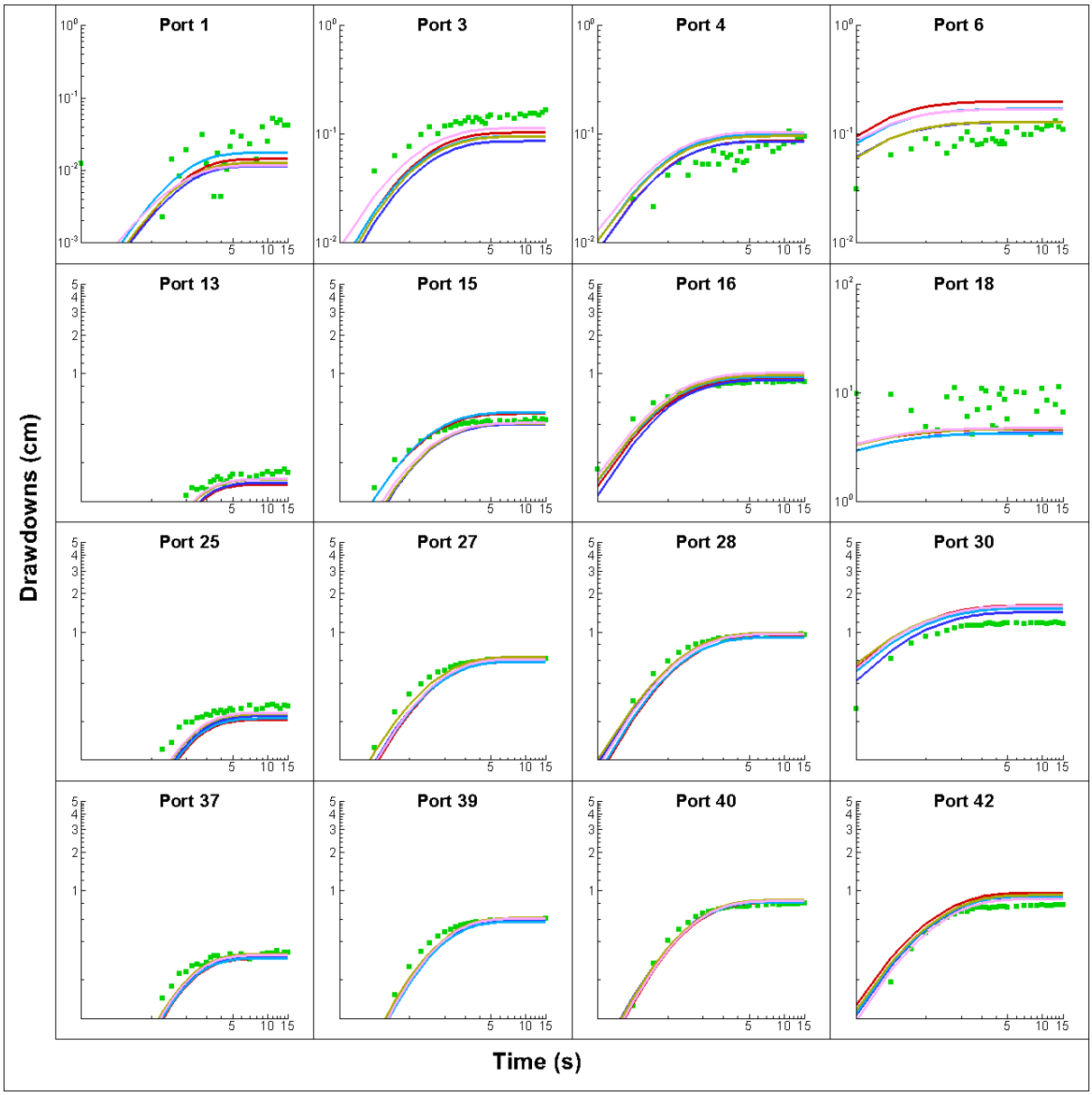

Fig. B27: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 18. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

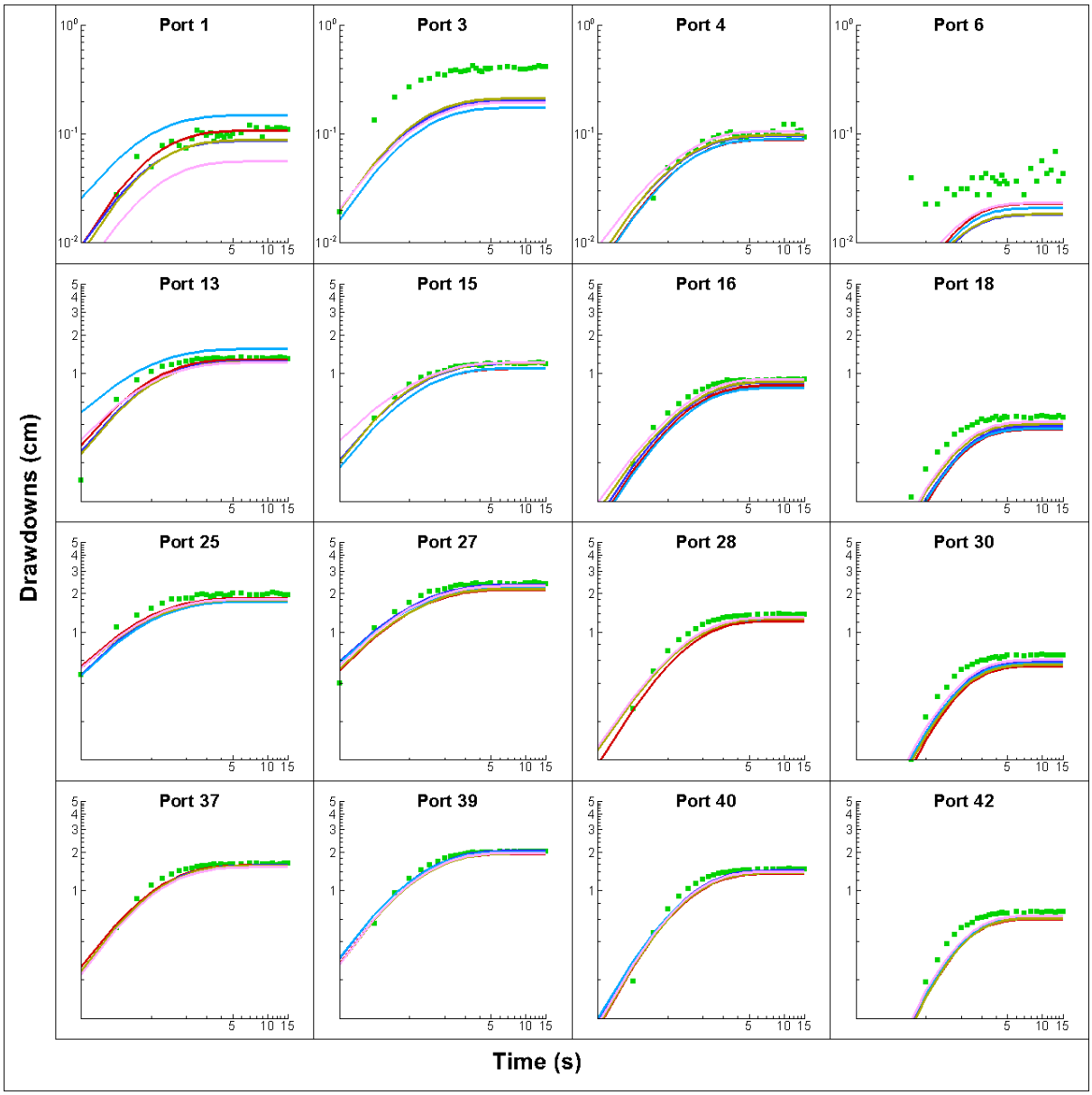

Fig. B28: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 20. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

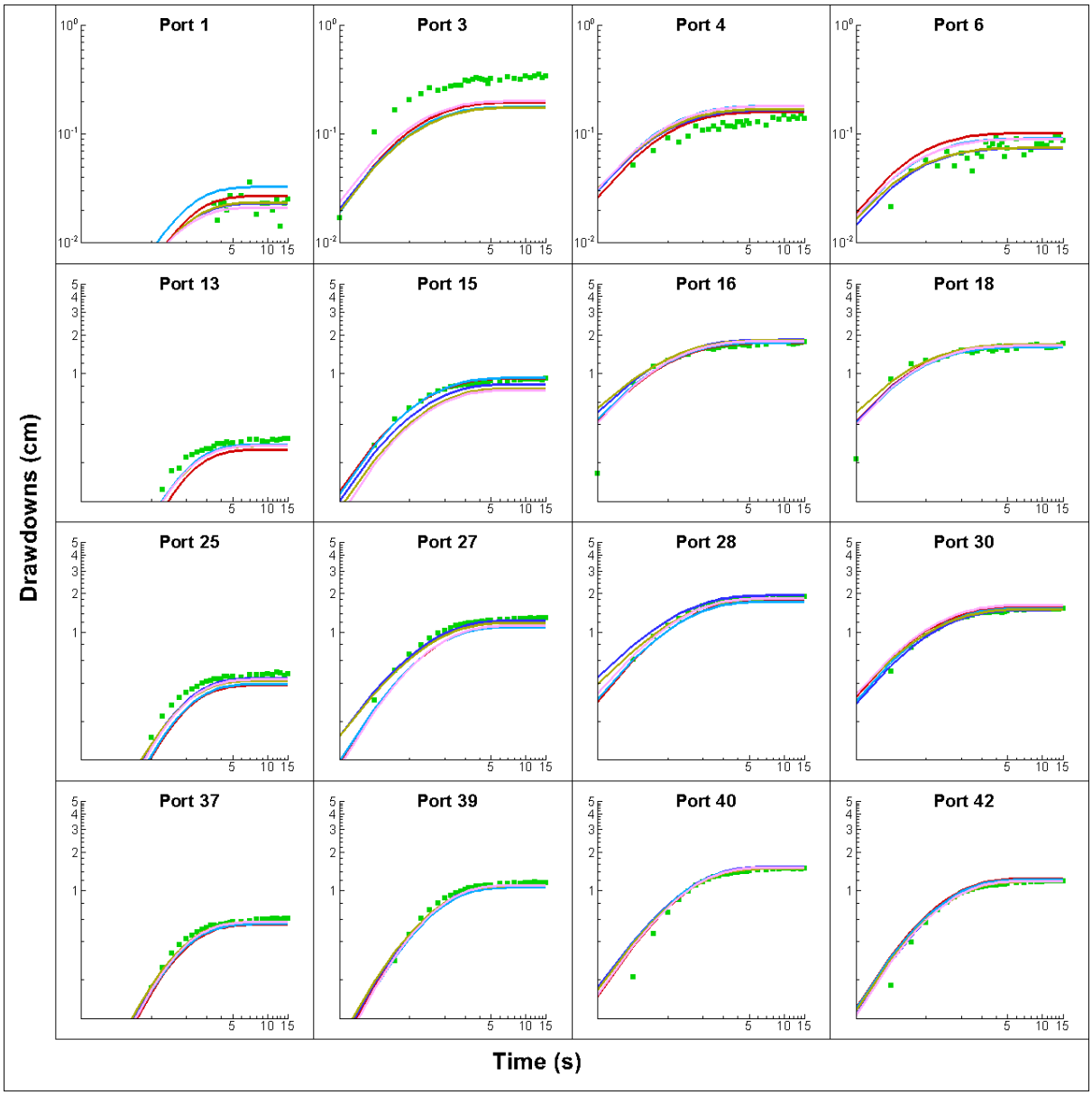

Fig. B29: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 23. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

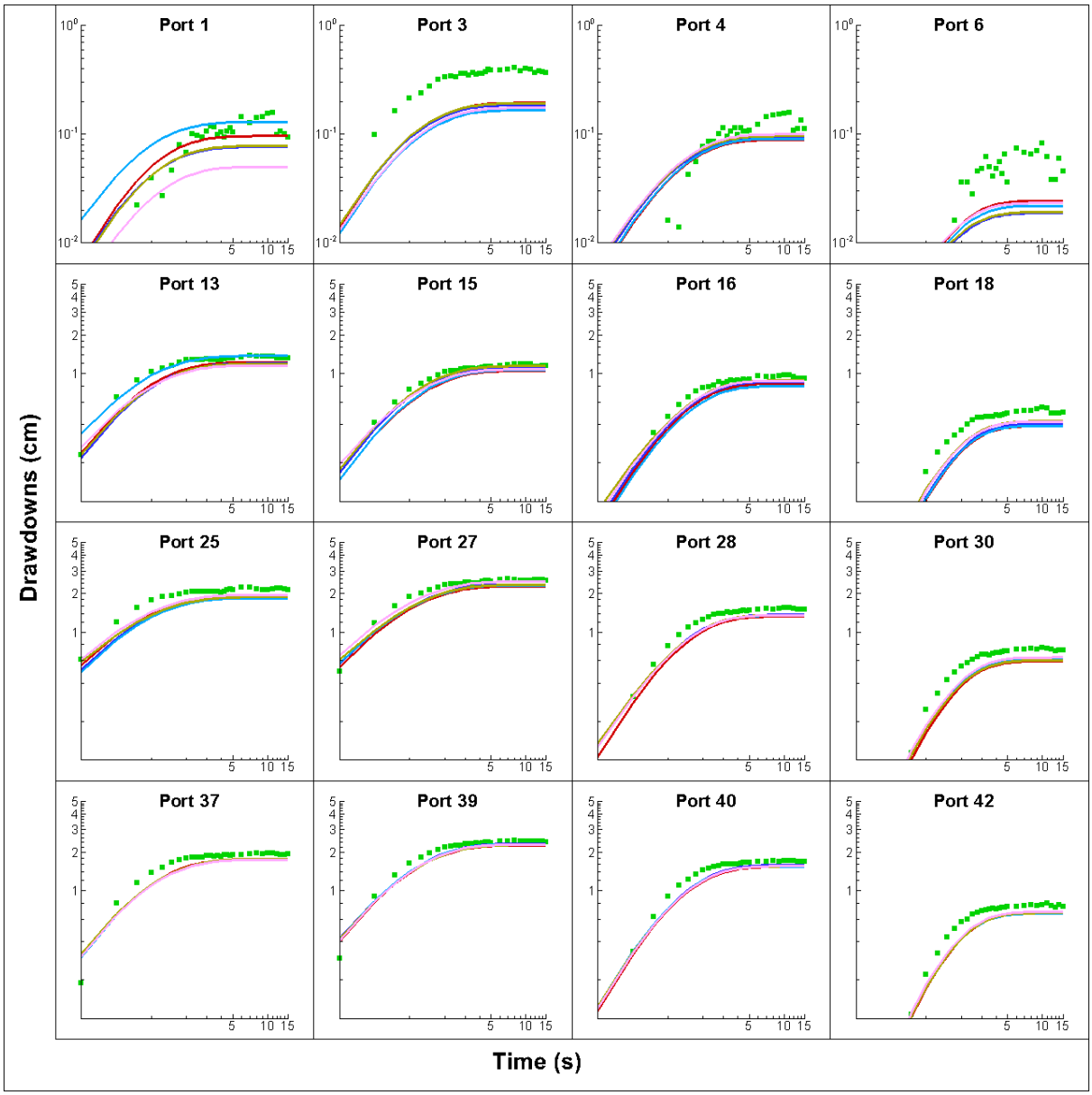

Fig. B30: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 26. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

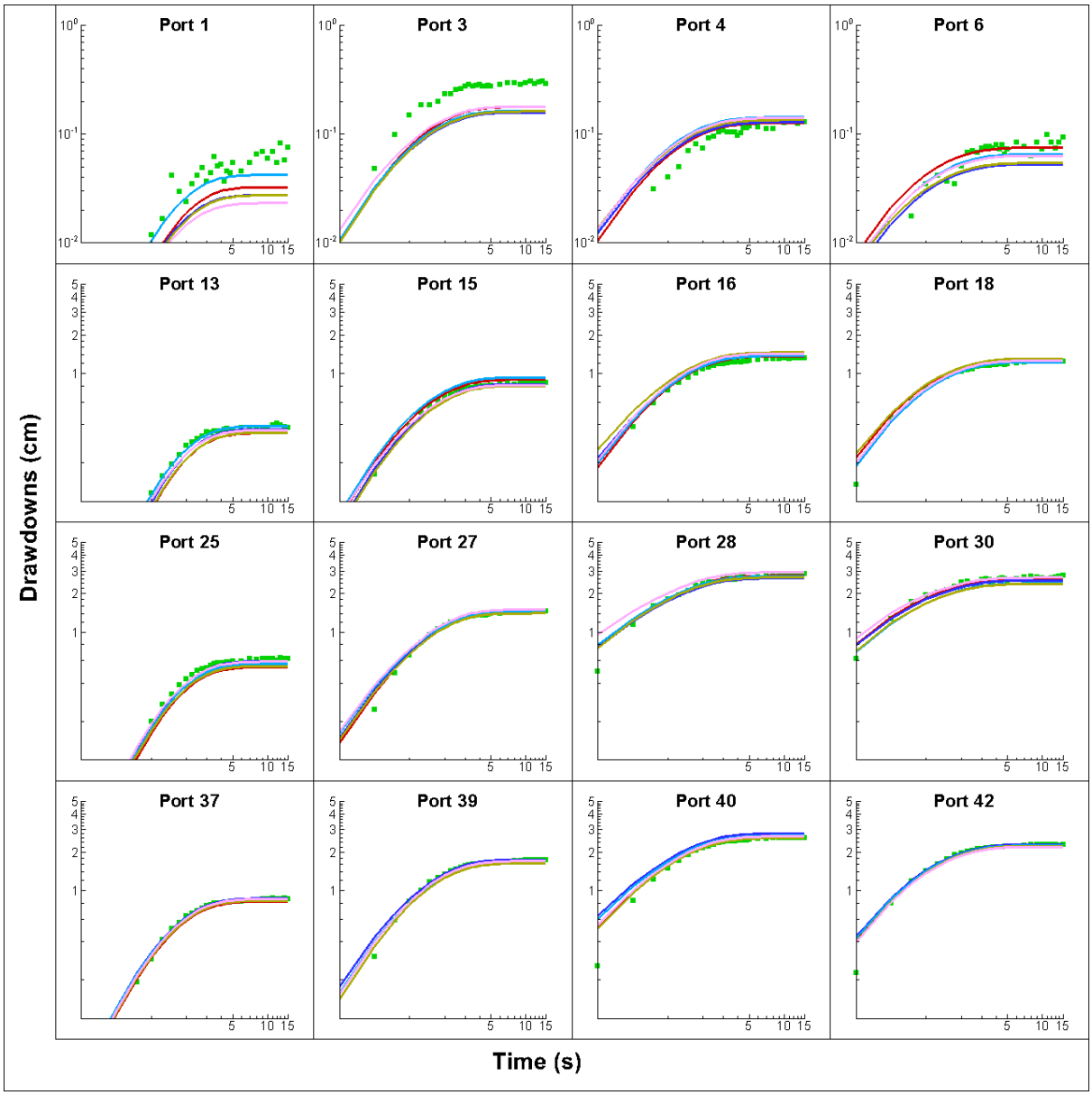

Fig. B31: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port29. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

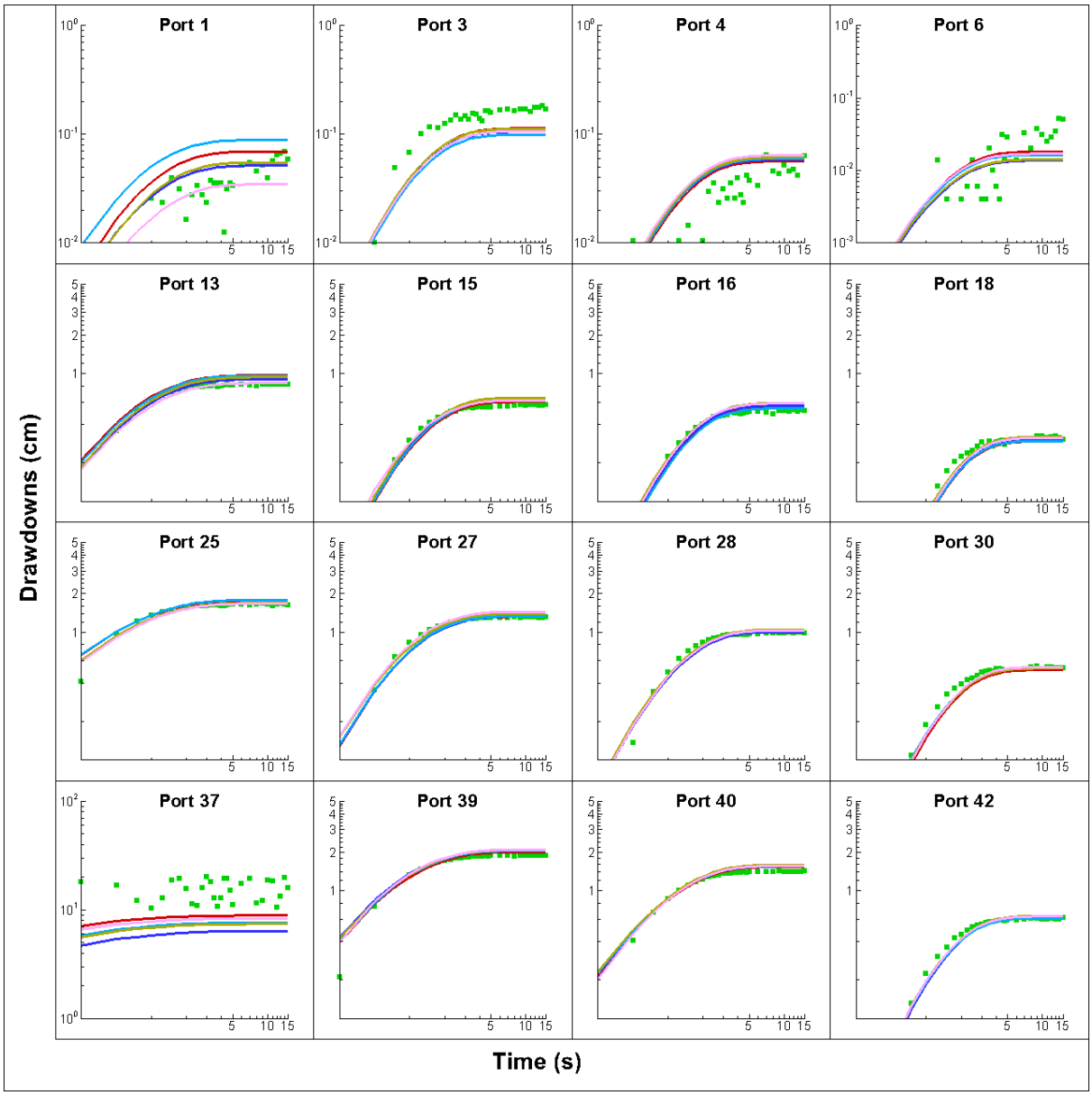

Fig. B32: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 37. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

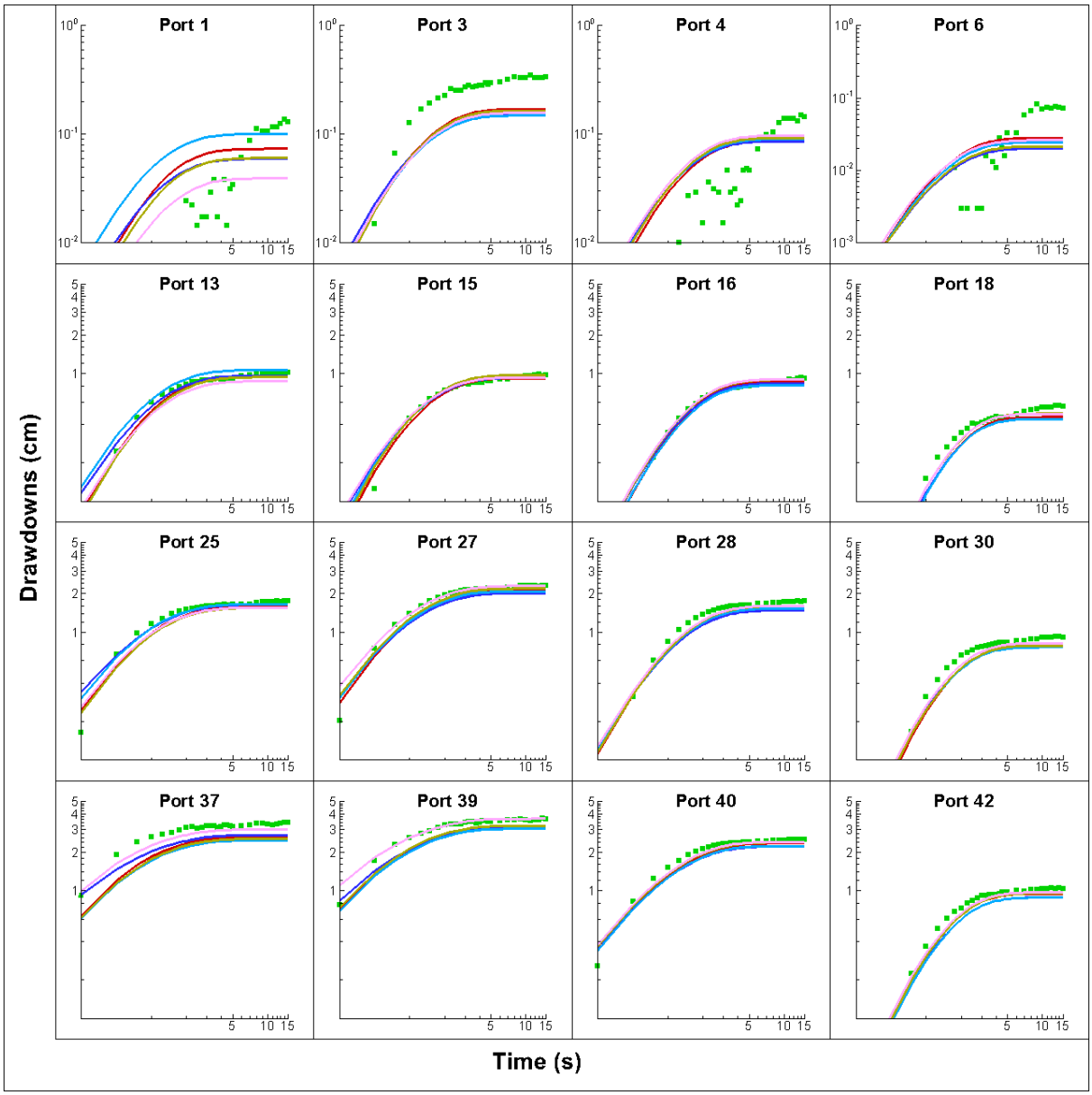

Fig. B33: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 38. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

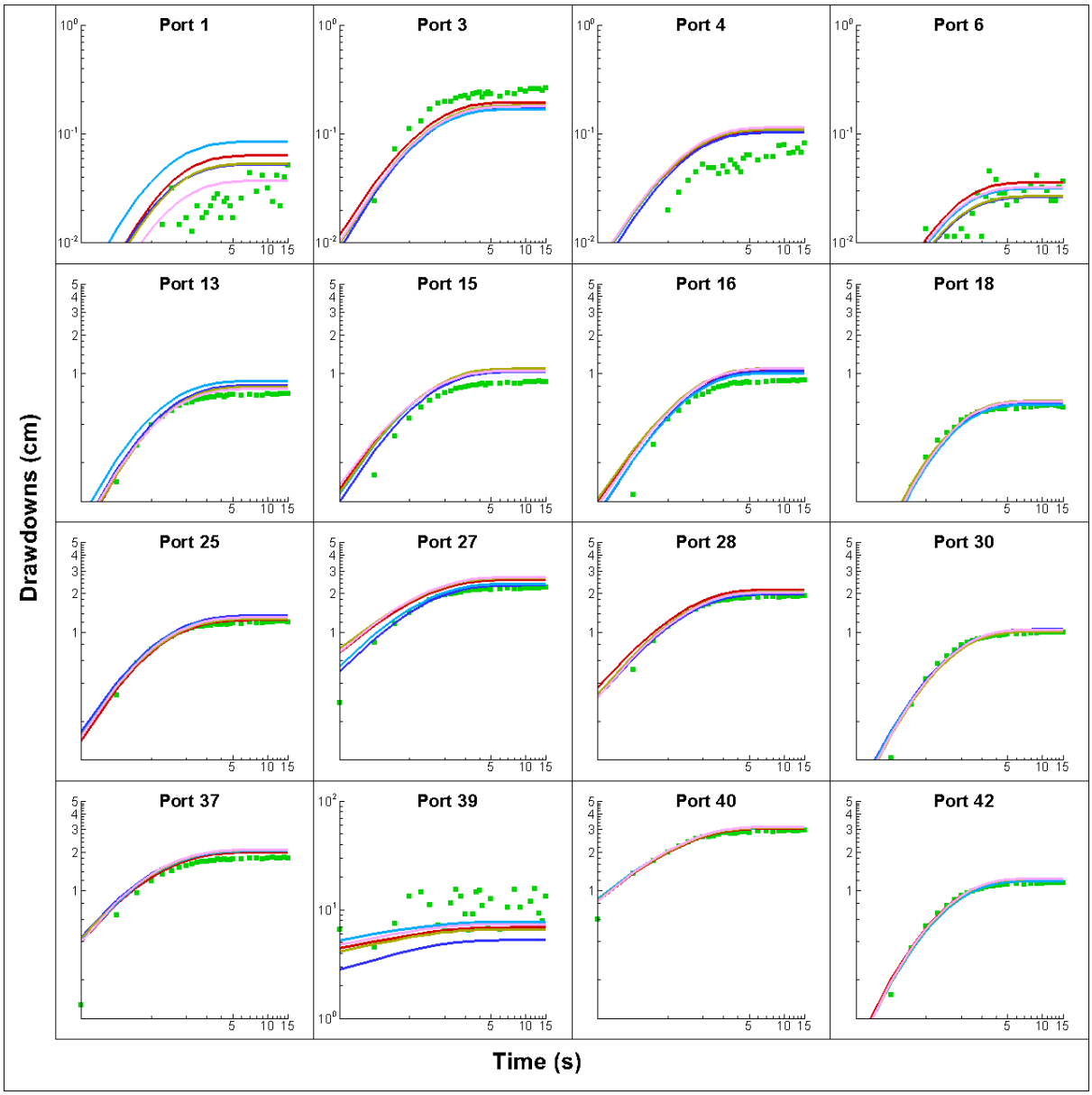

Fig. B34: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 39. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

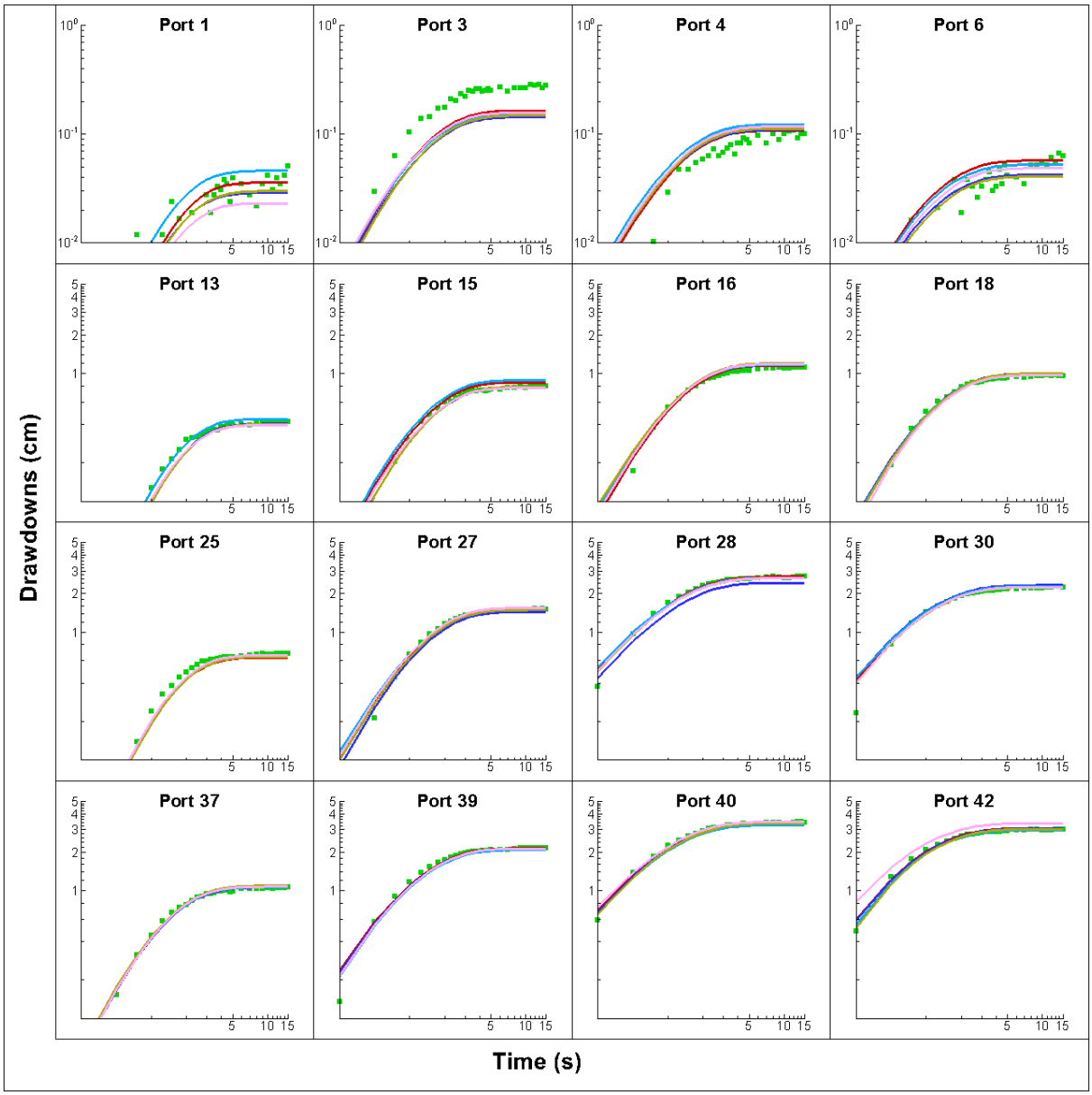

Fig. B35: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 41. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

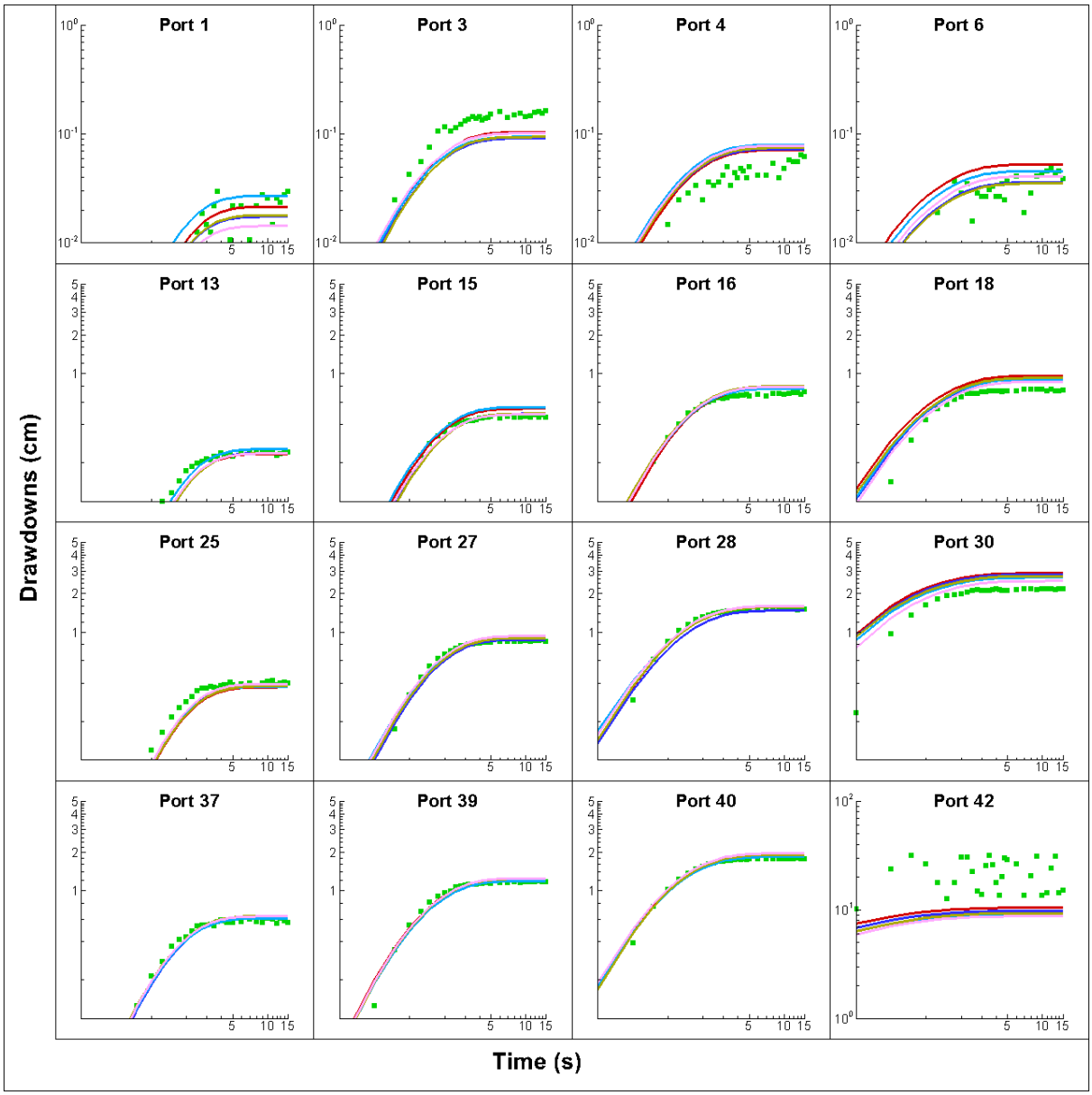

Fig. B36: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 42. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 8 pumping tests and 47 observation ports (Case 1).

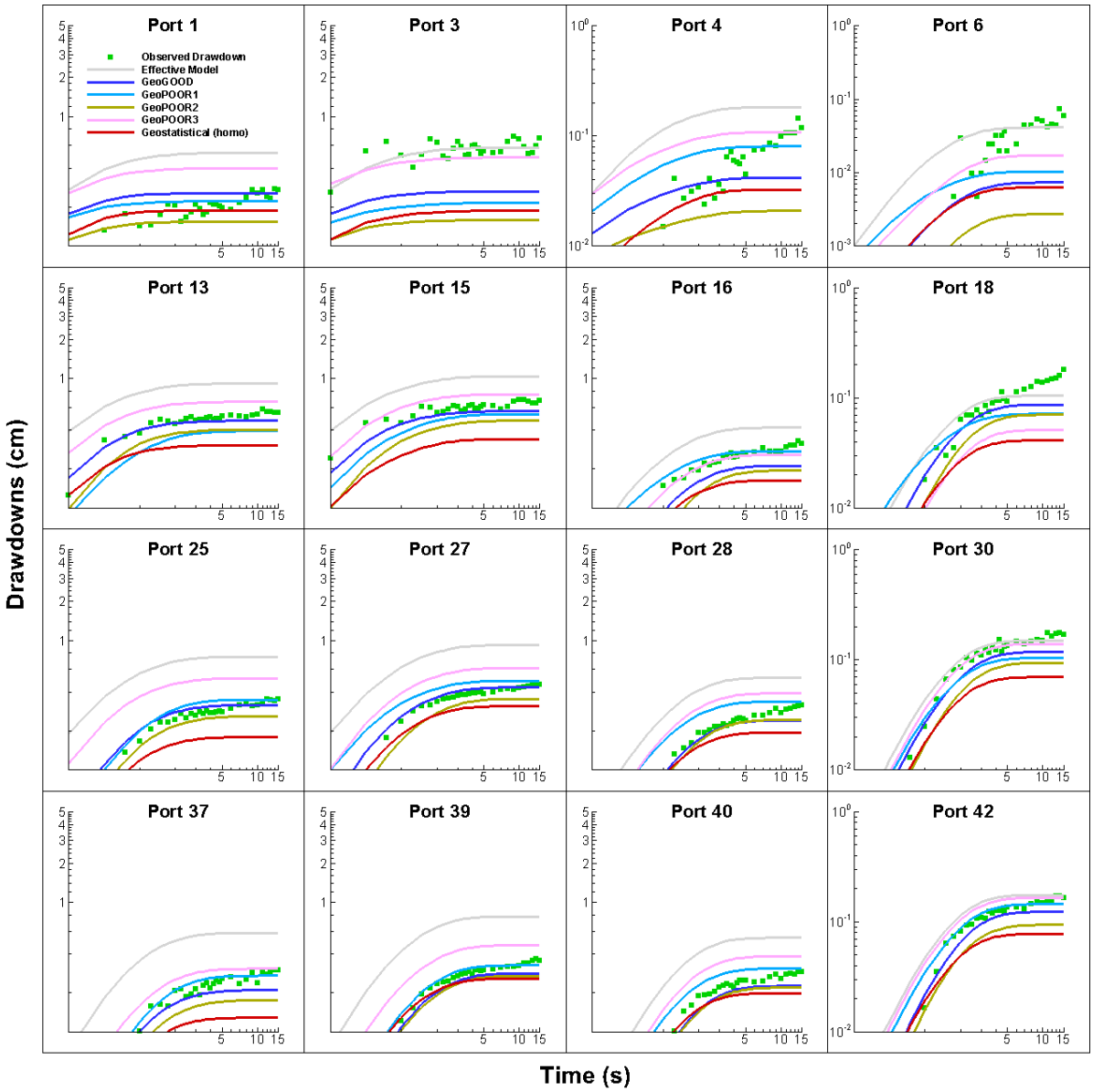

Fig. B37: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 8. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

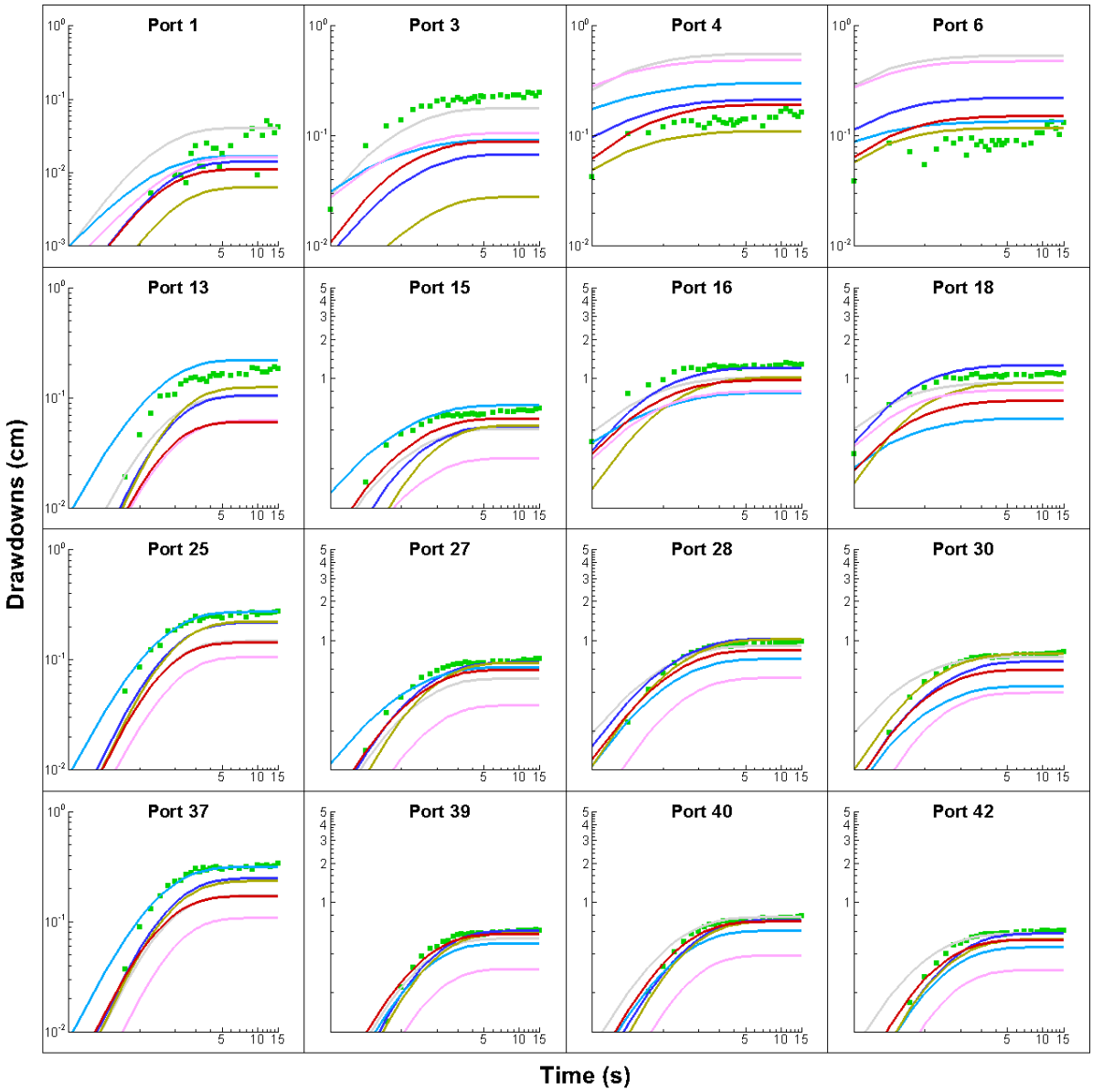

Fig. B38: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 11. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

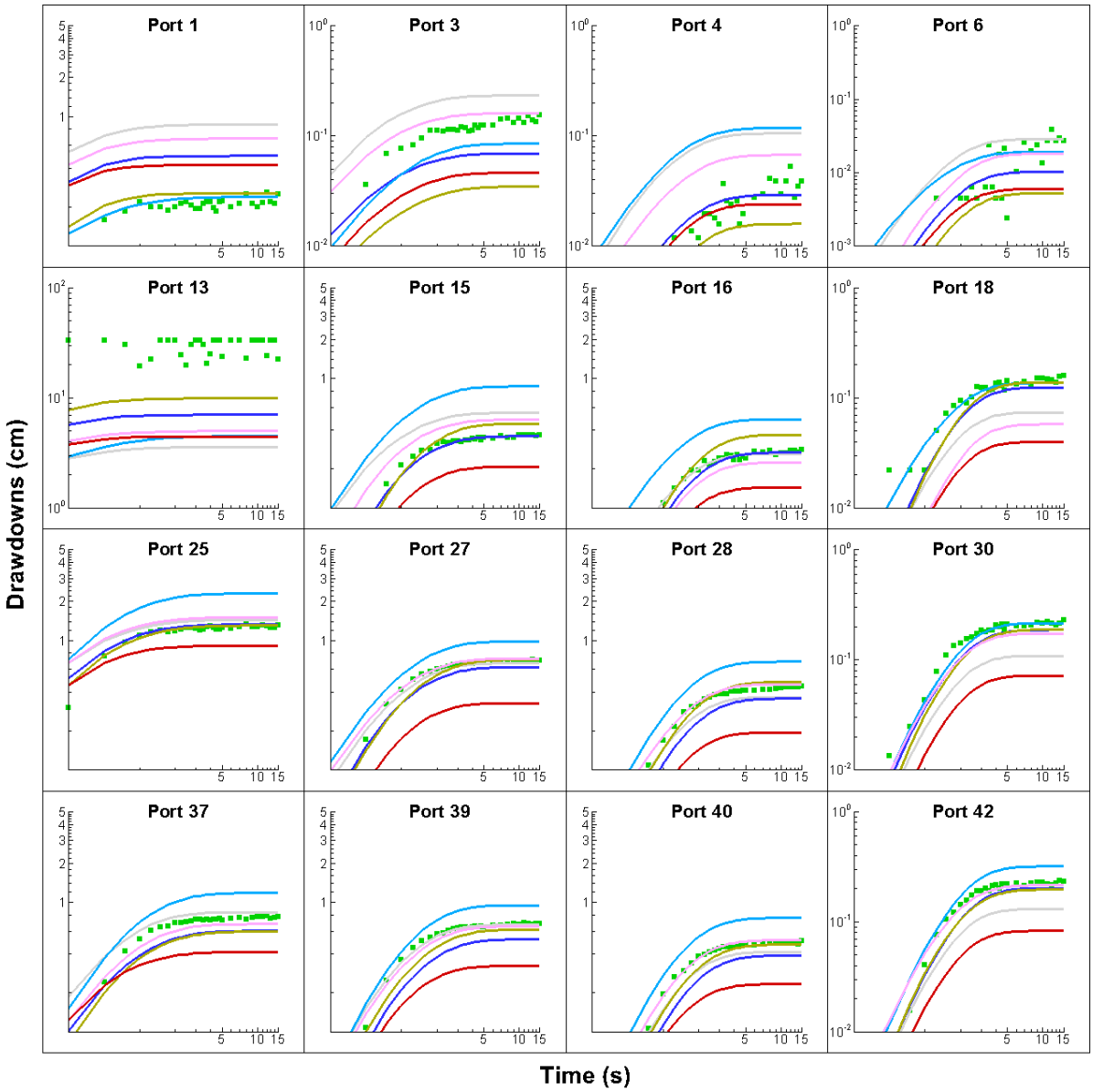

Fig. B39: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 13. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

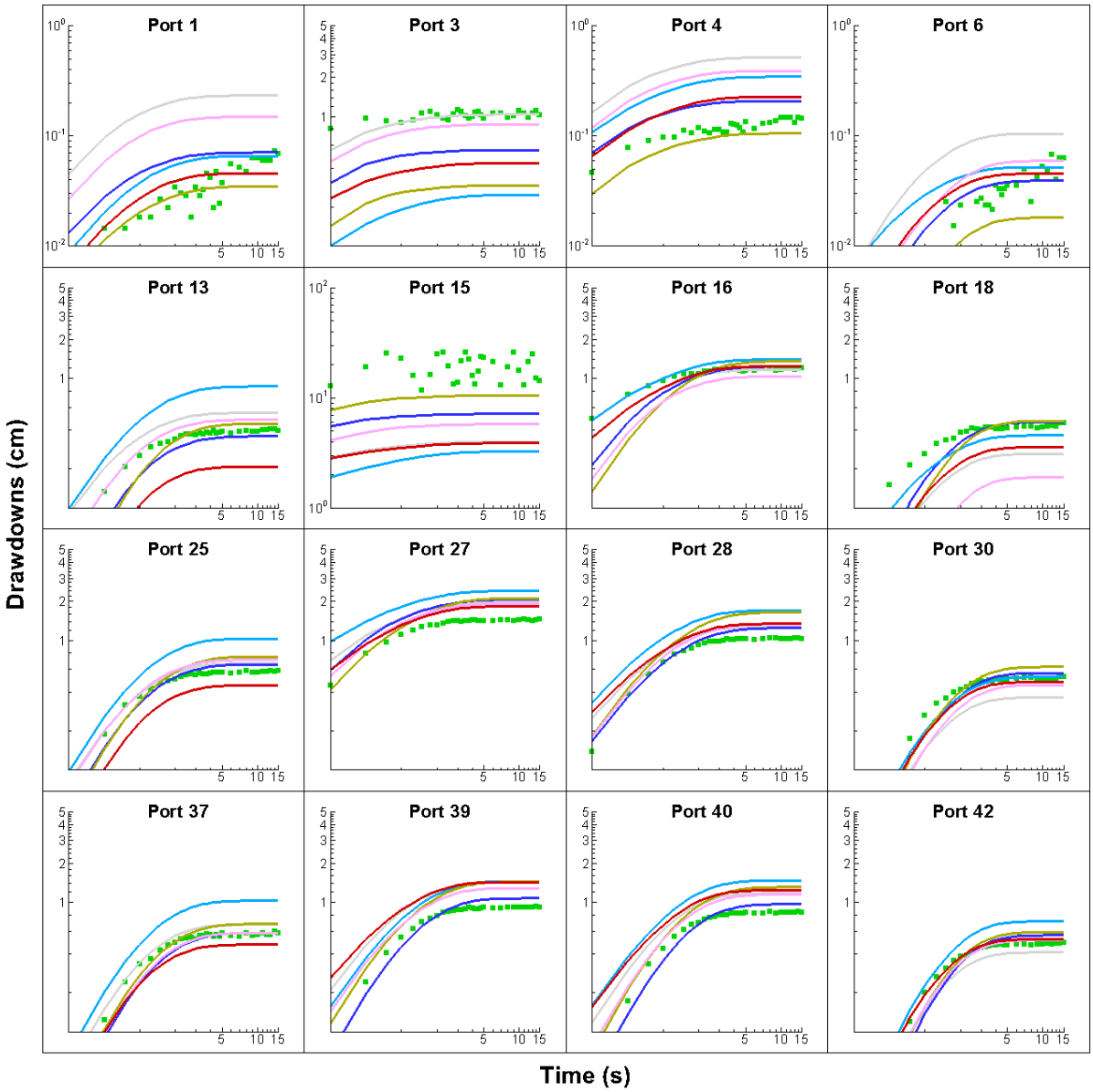

Fig. B40: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 15. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

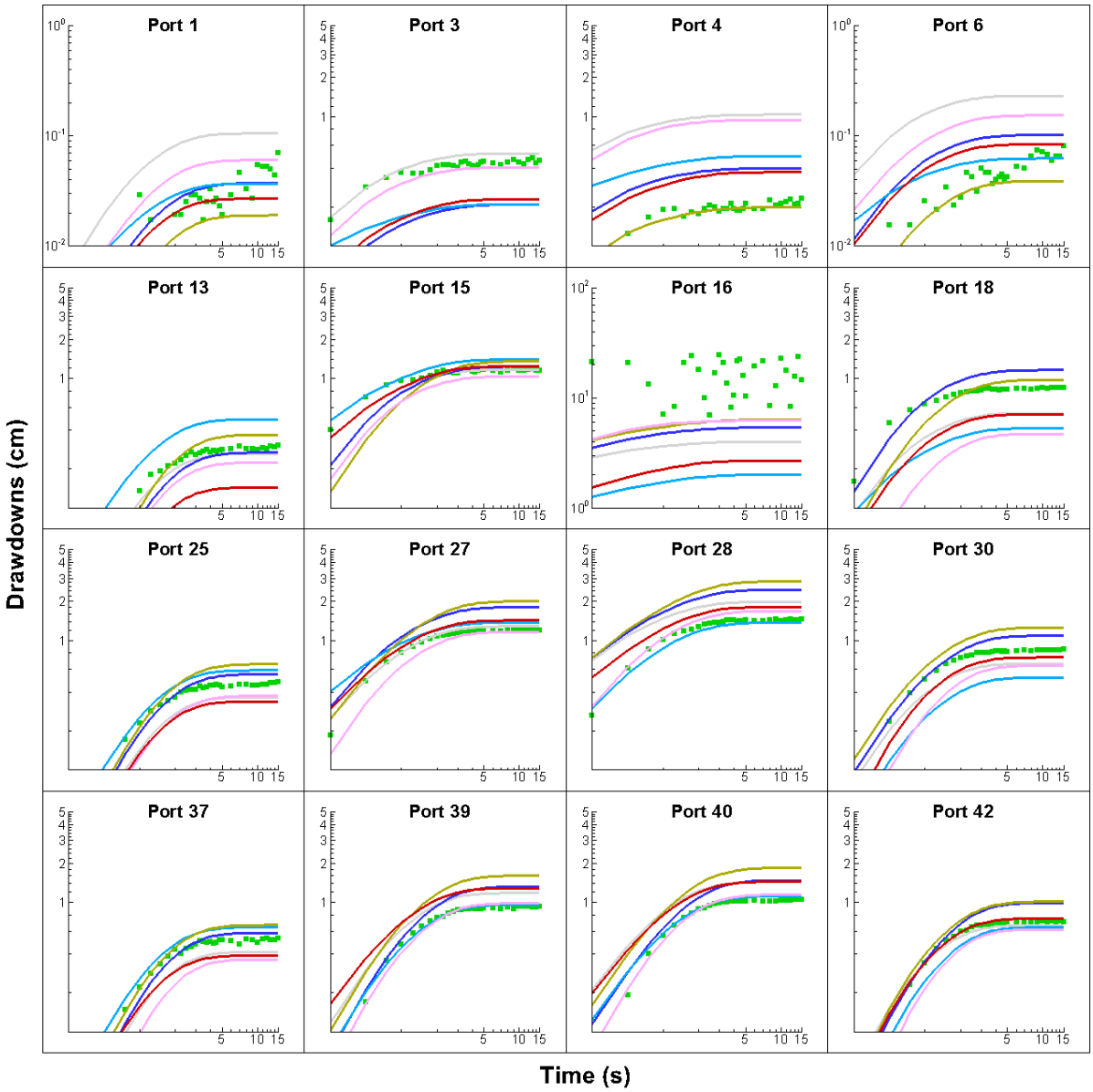

Fig. B41: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 16. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

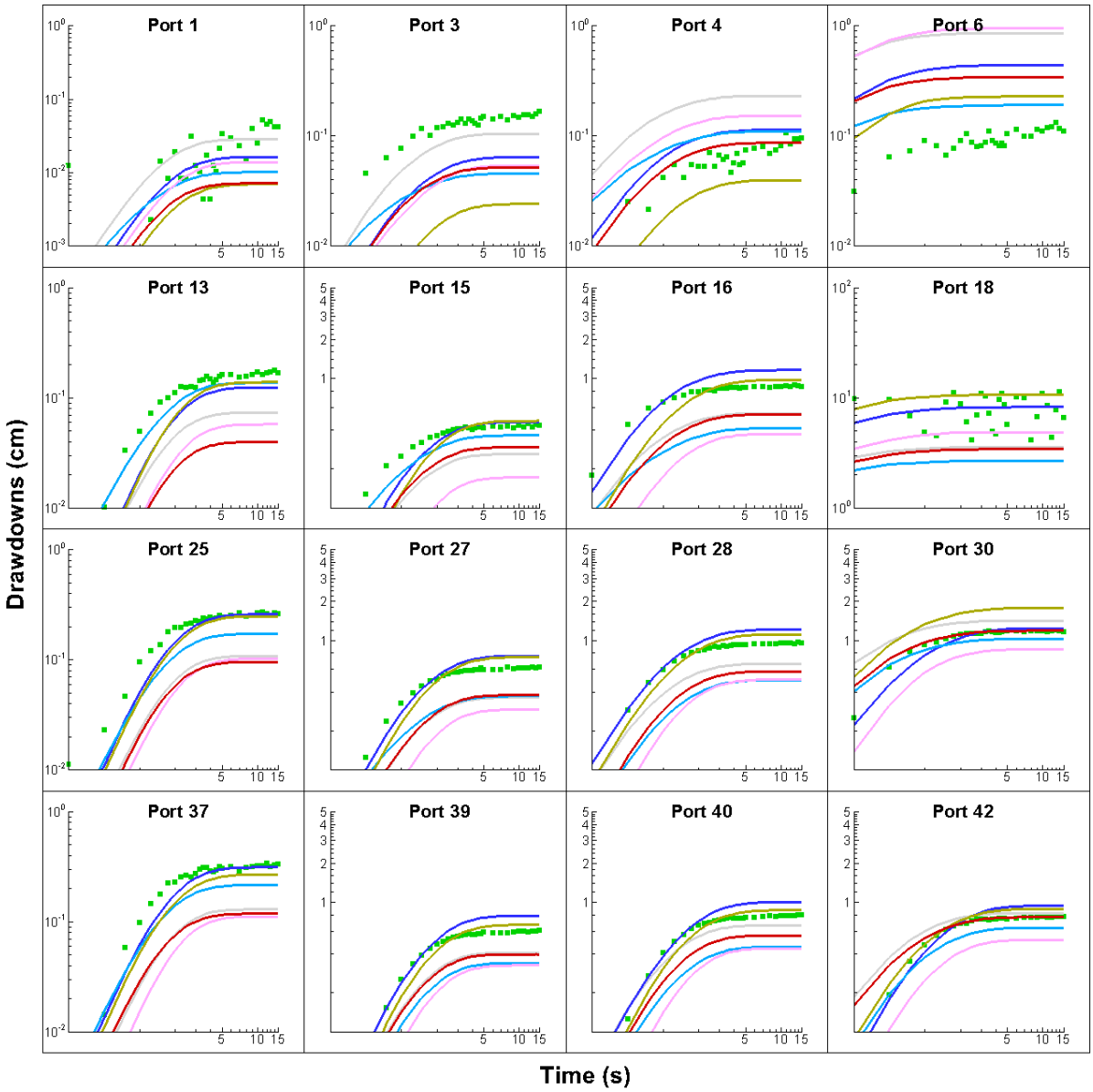

Fig. B42: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 18. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

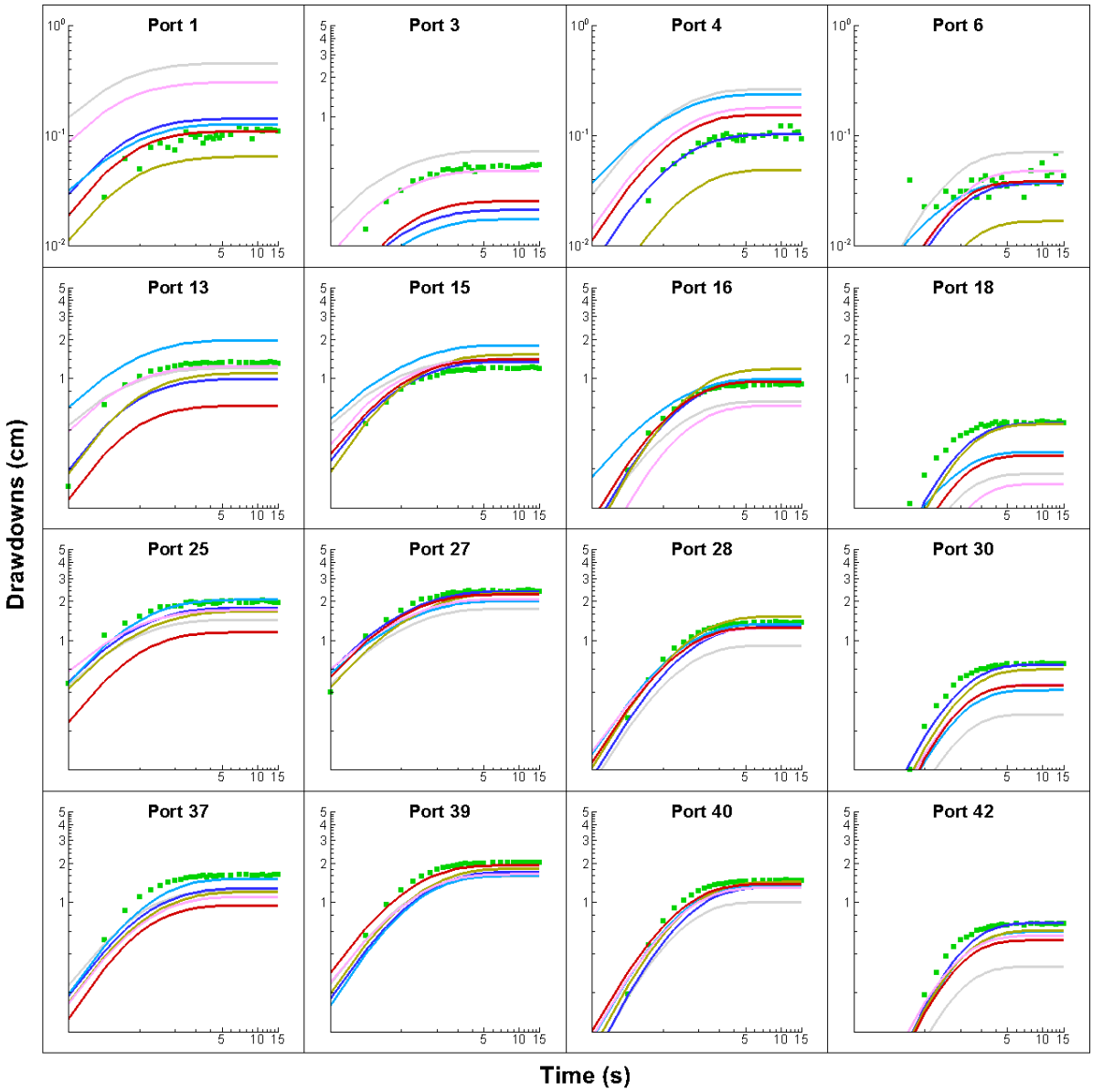

Fig. B43: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 20. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

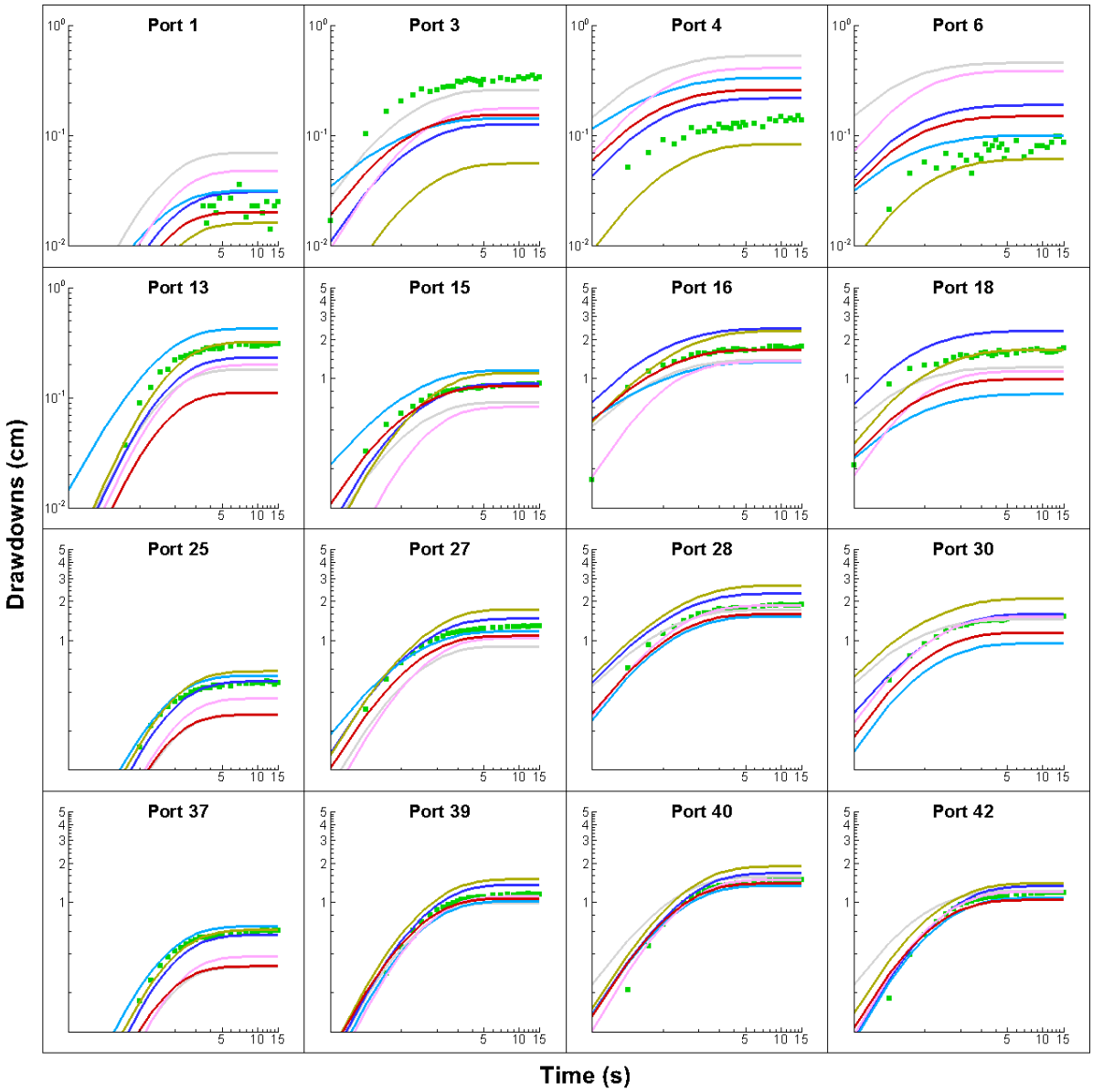

Fig. B44: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 23. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

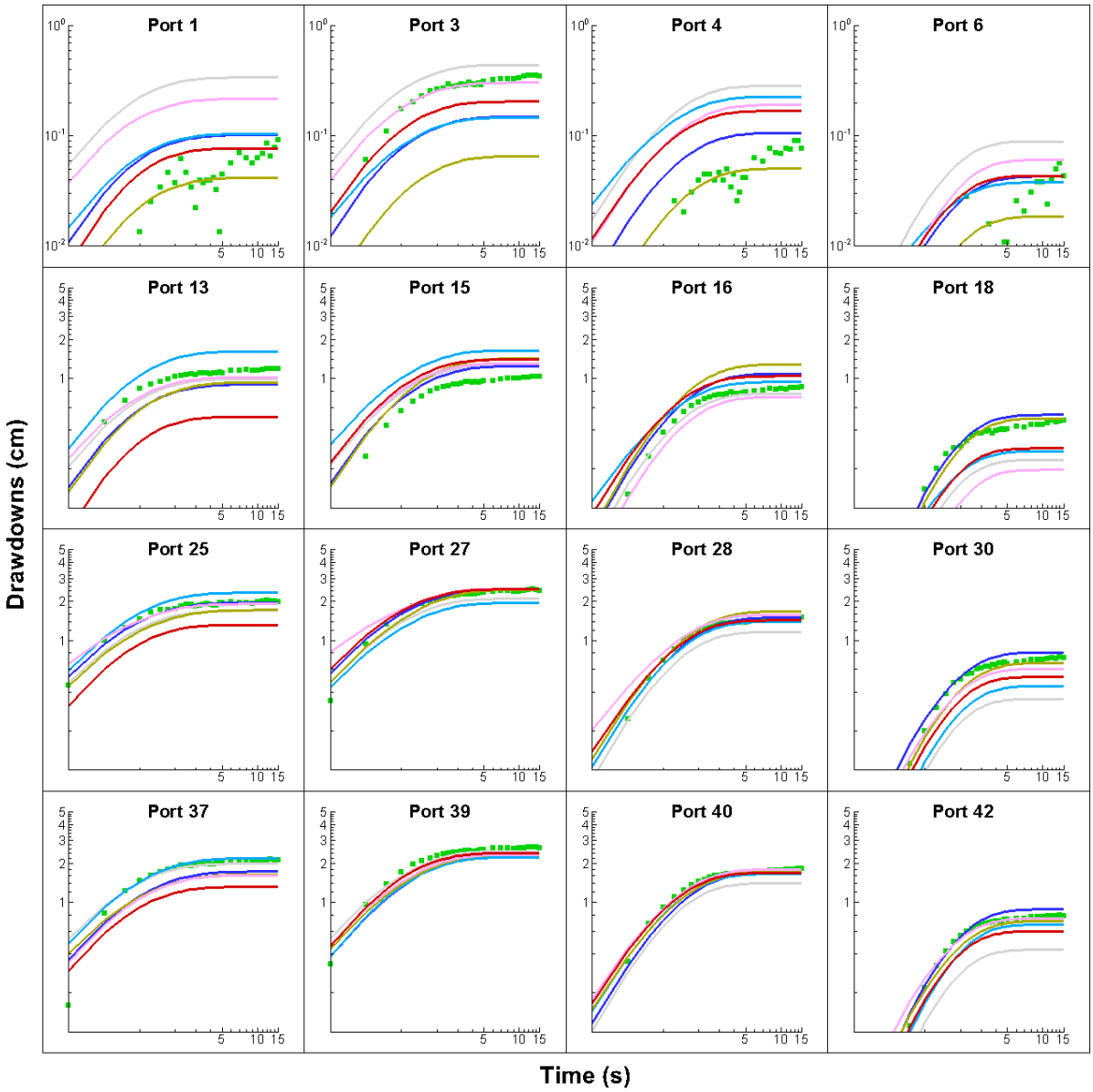

Fig. B45: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 32. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

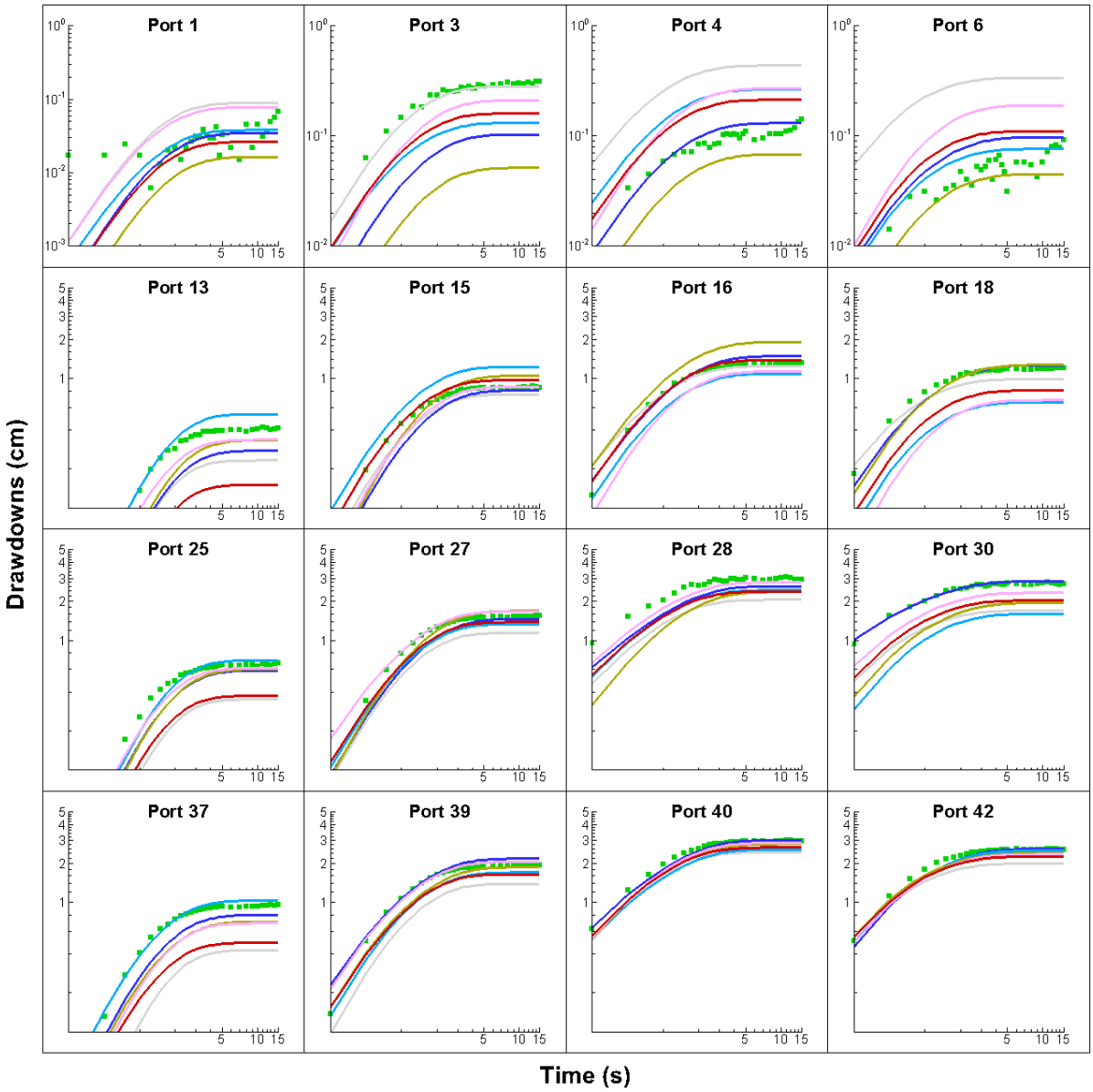

Fig. B46: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 35. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

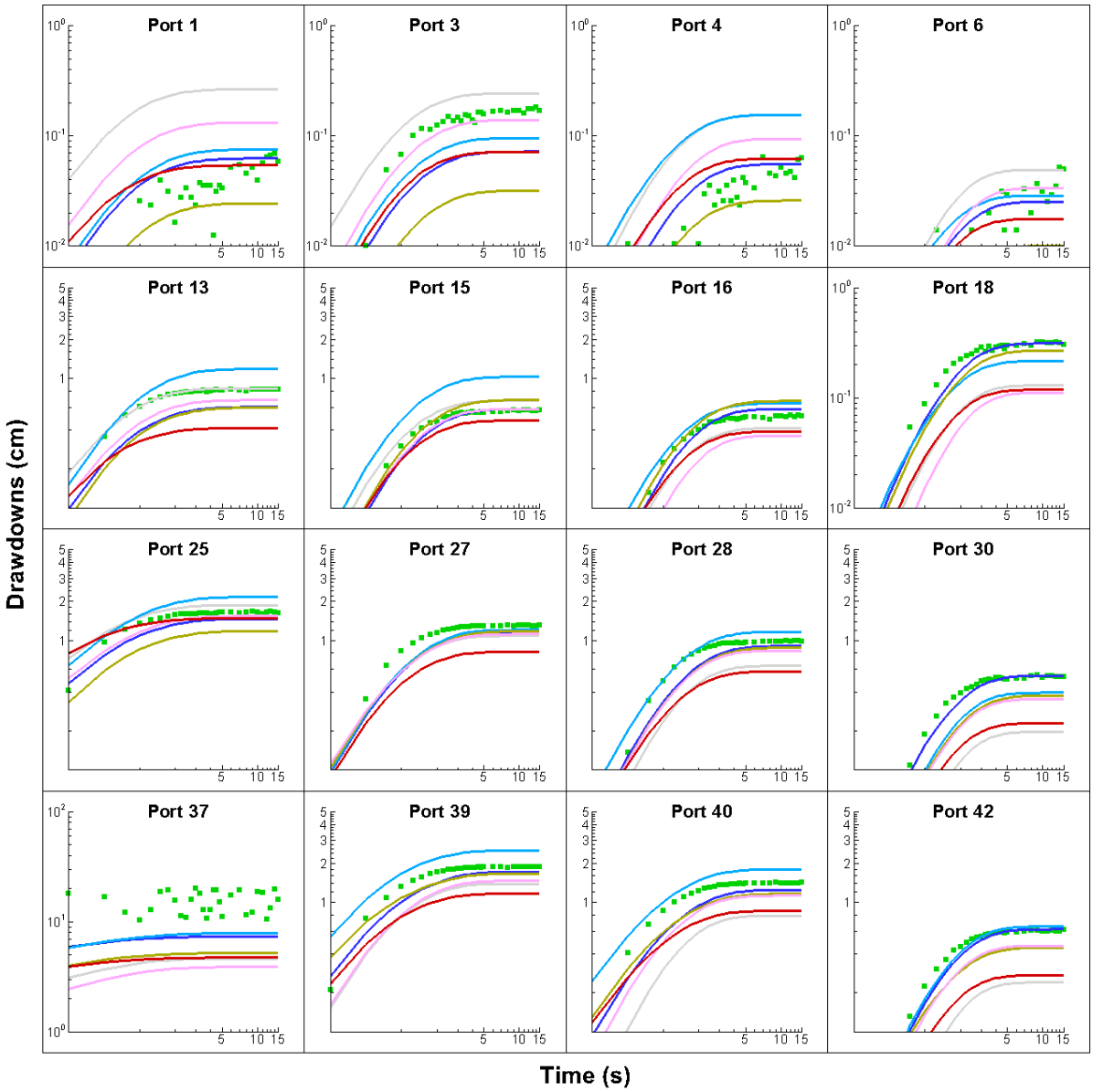

Fig. B47: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 37. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

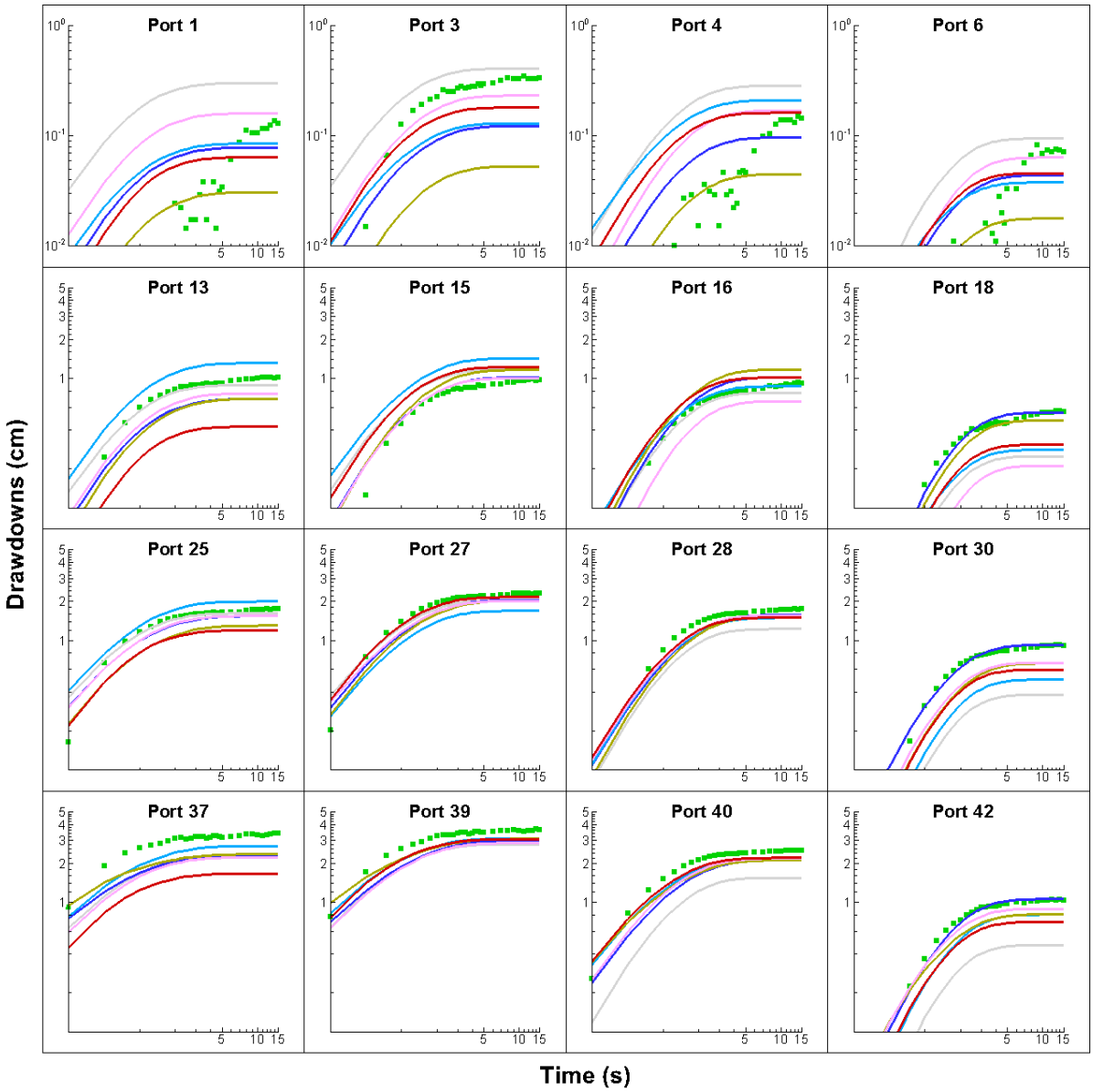

Fig. B48: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 38. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

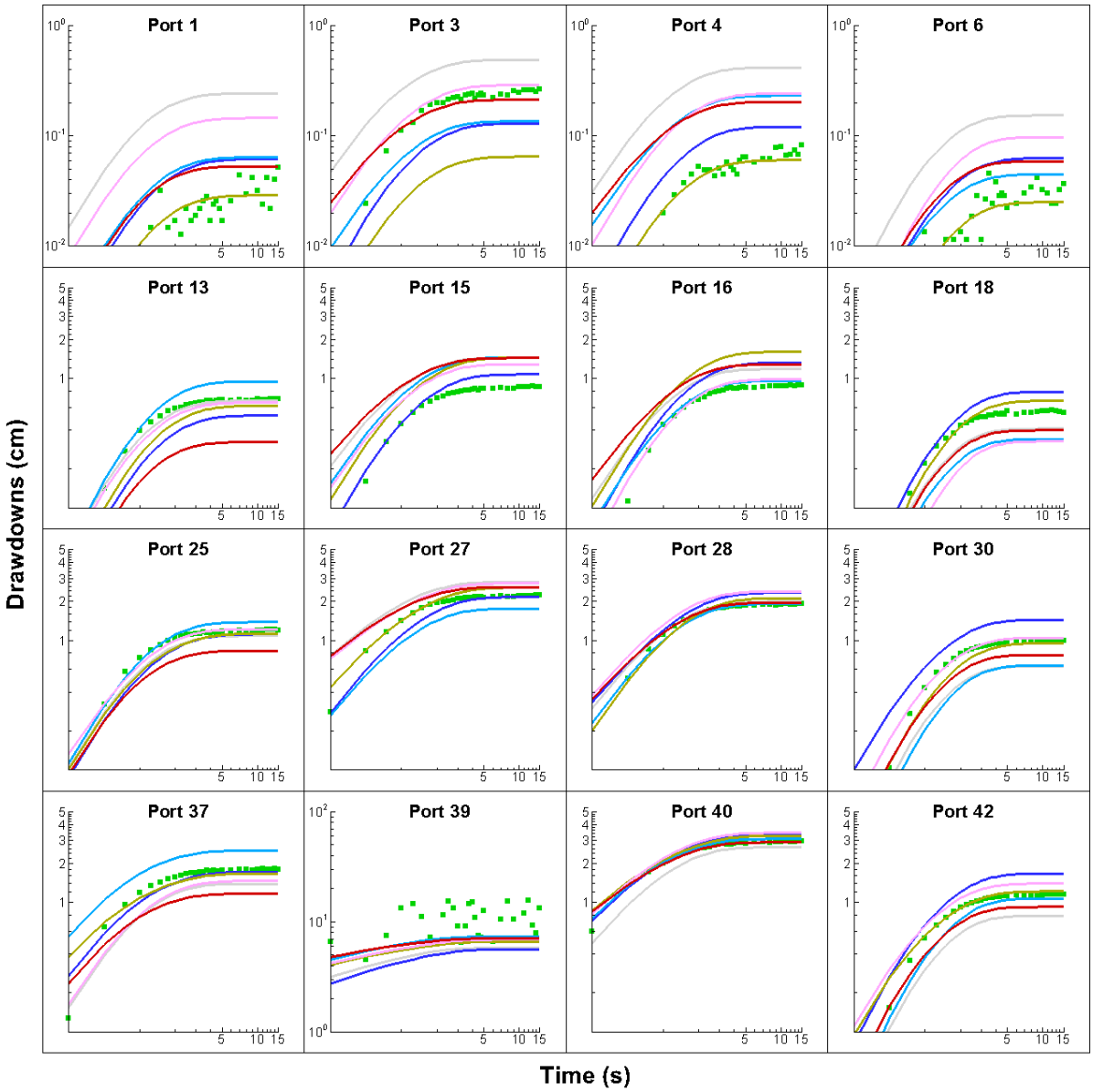

Fig. B49: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 39. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

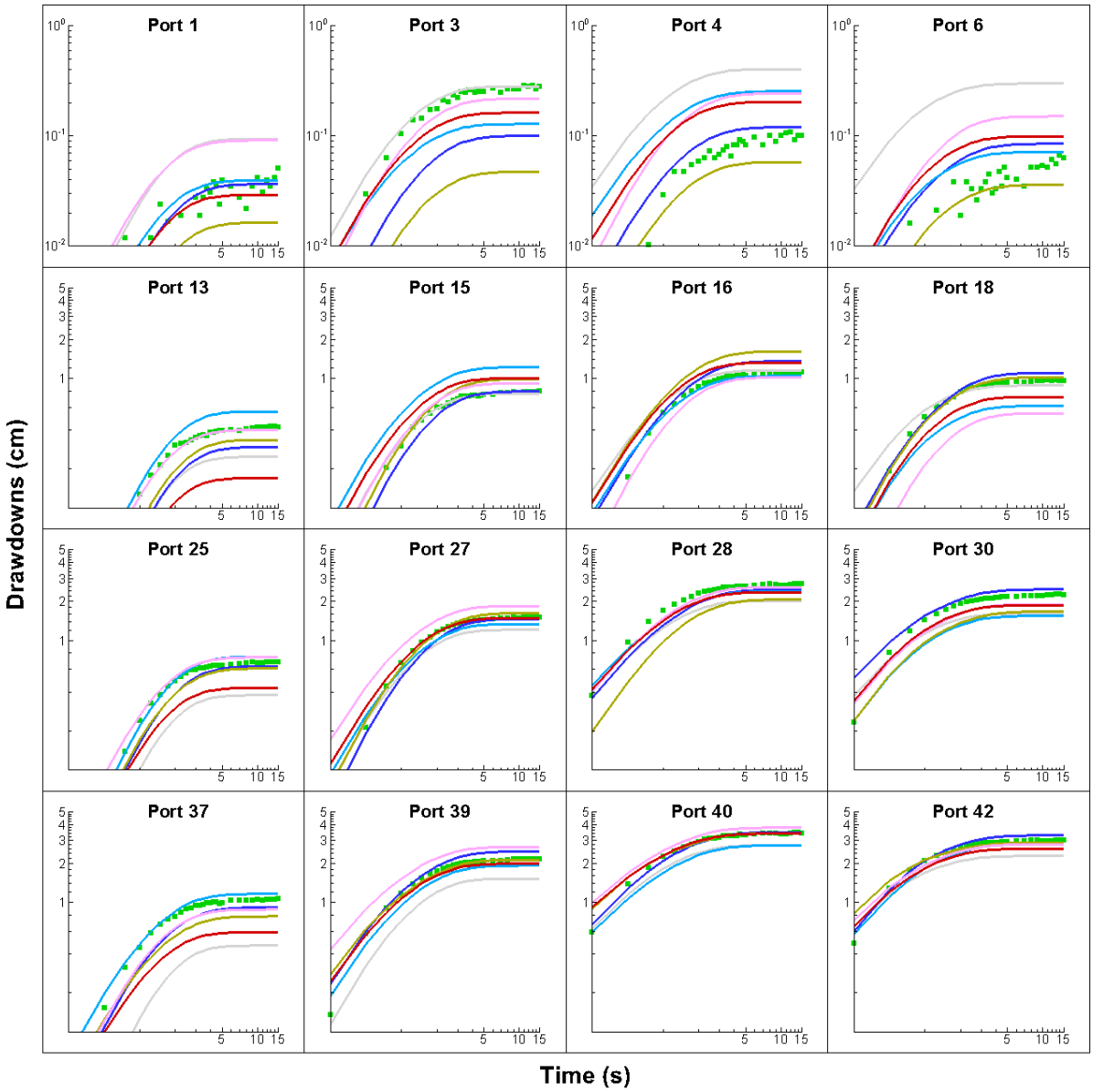

Fig. B50: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 41. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

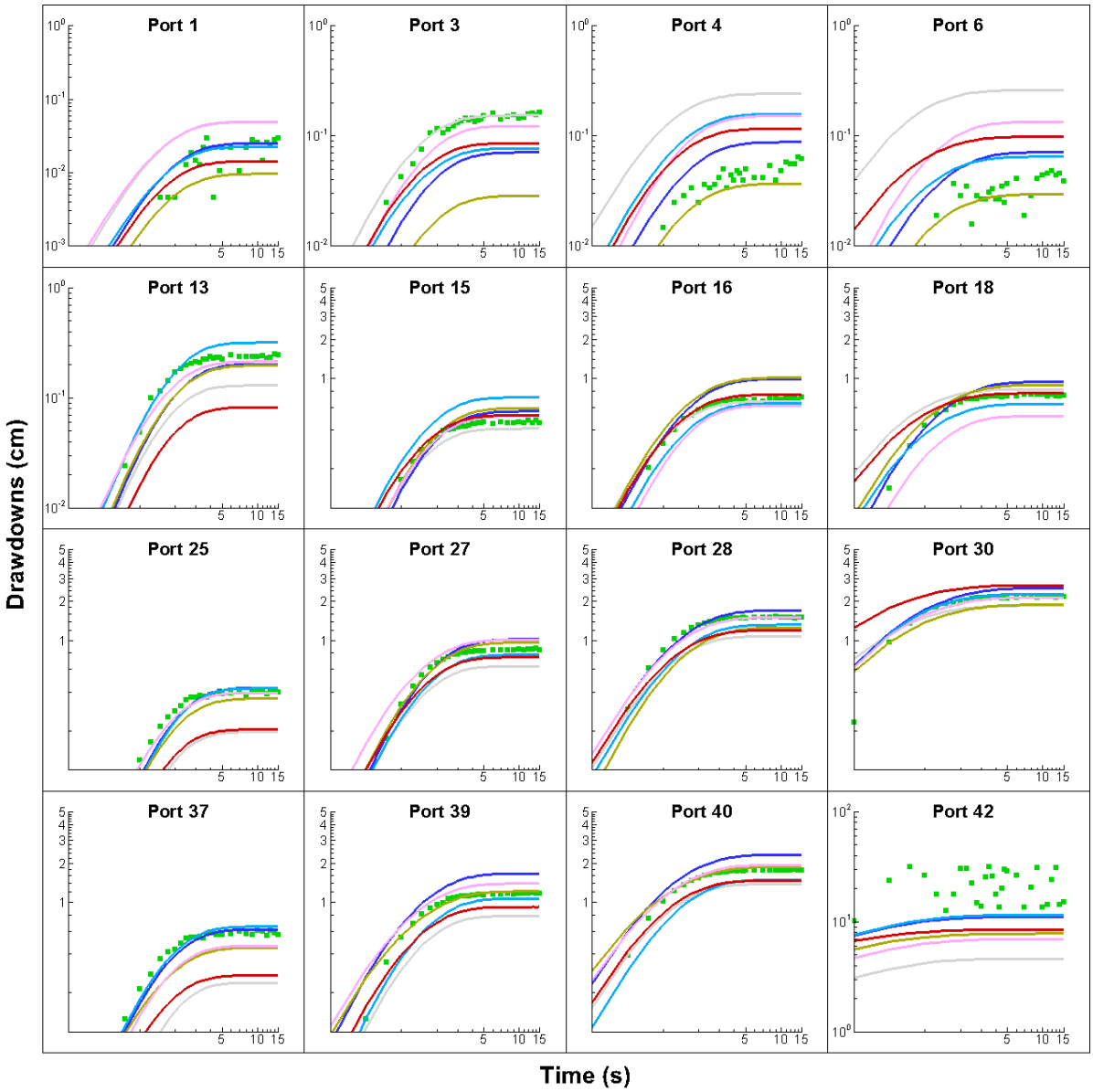

Fig. B51: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 42. Here, the *K* and *Ss* tomograms are obtained from different modeling approaches with 4 pumping tests and 15 observation ports (Case 2).

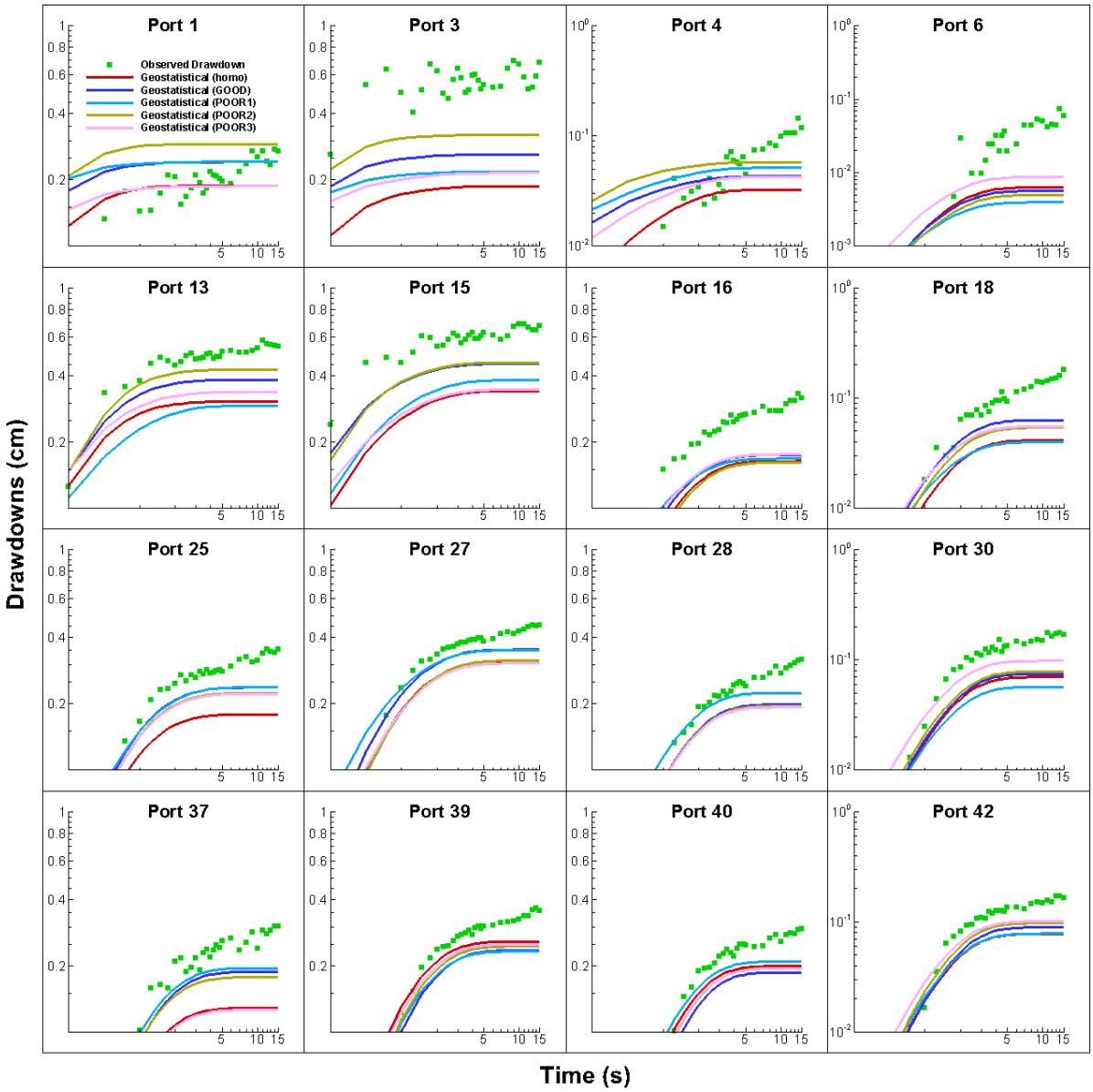

Fig. B52: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 8. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

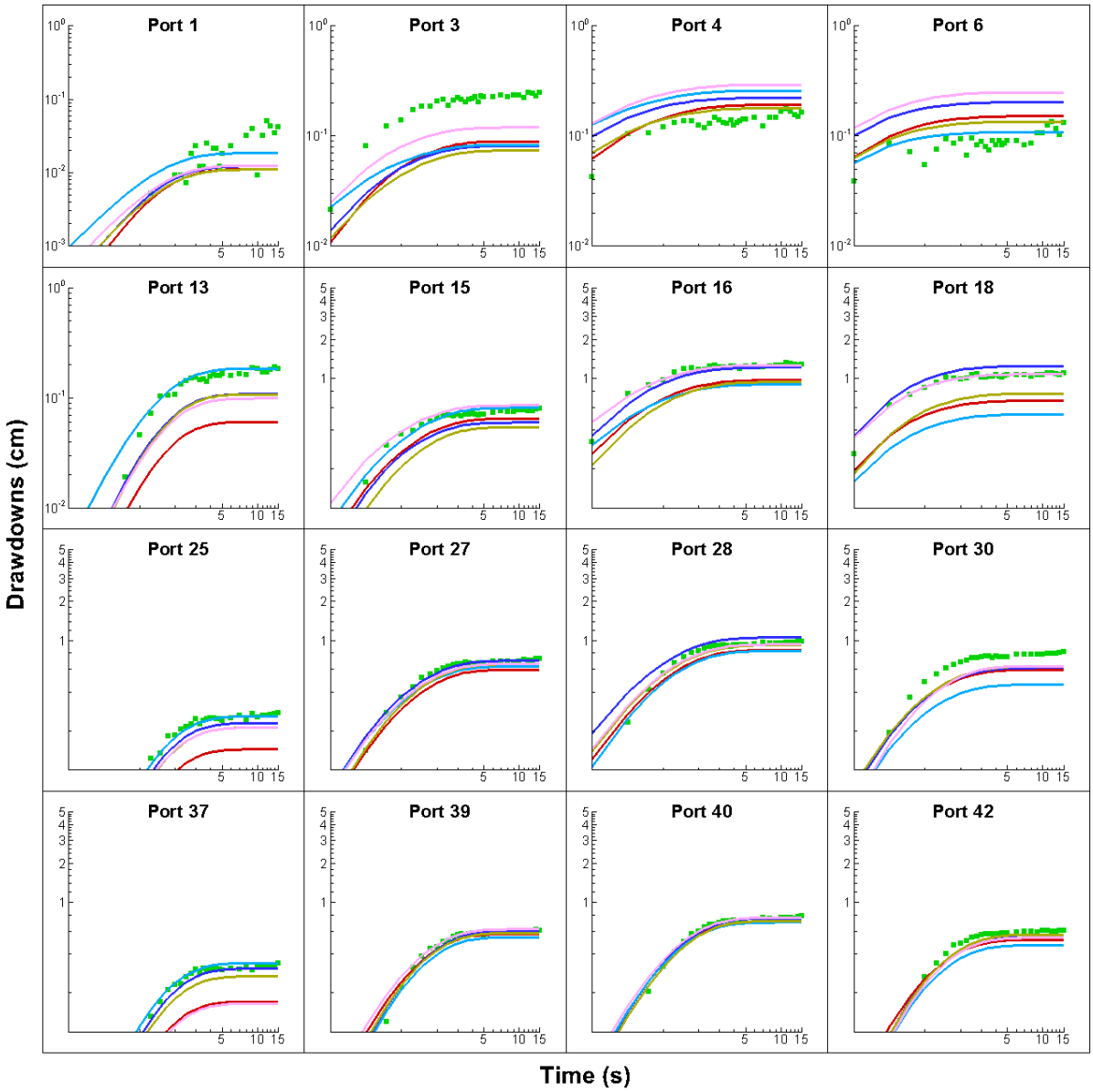

Fig. B53: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 11. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

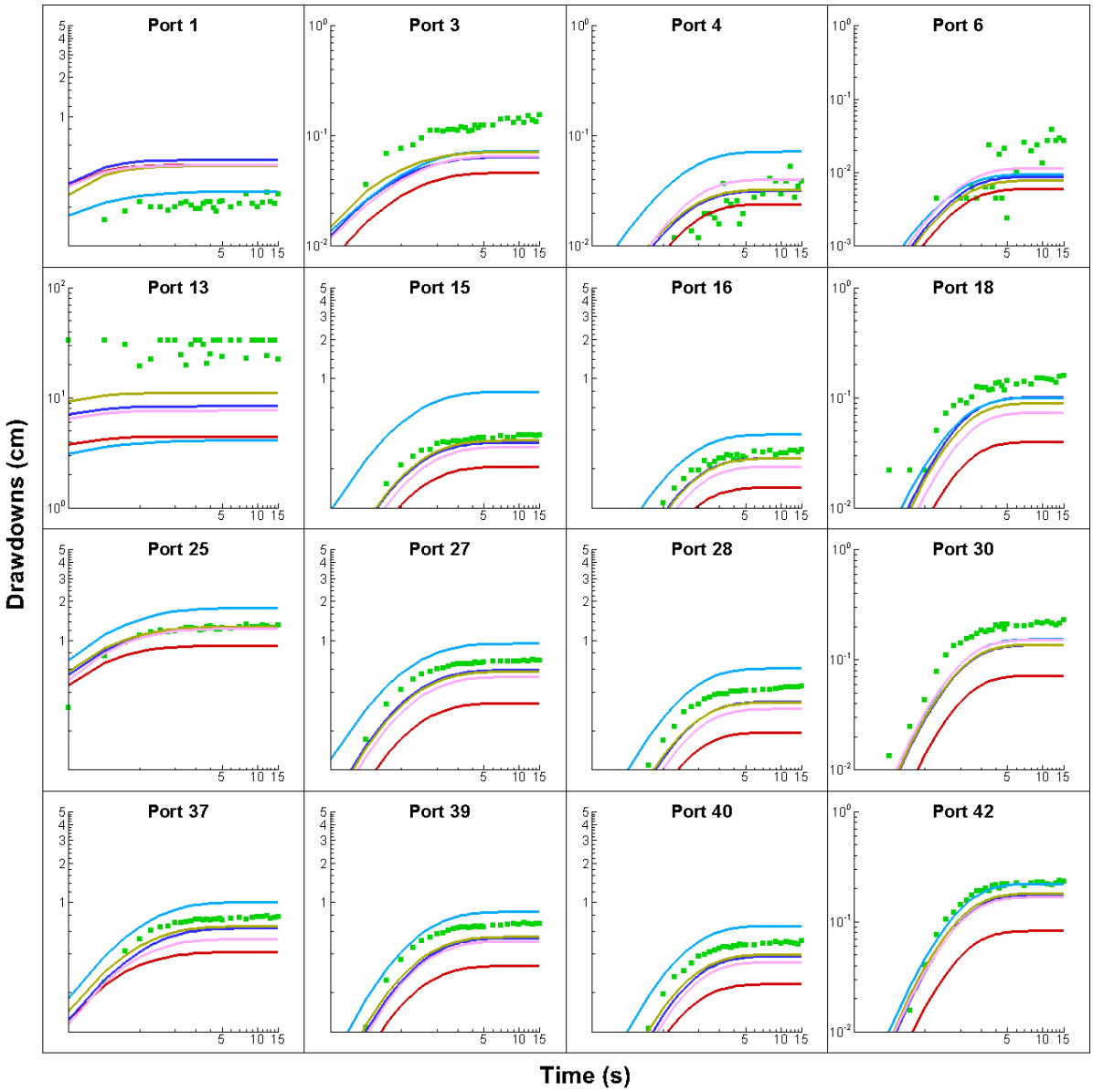

Fig. B54: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 13. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

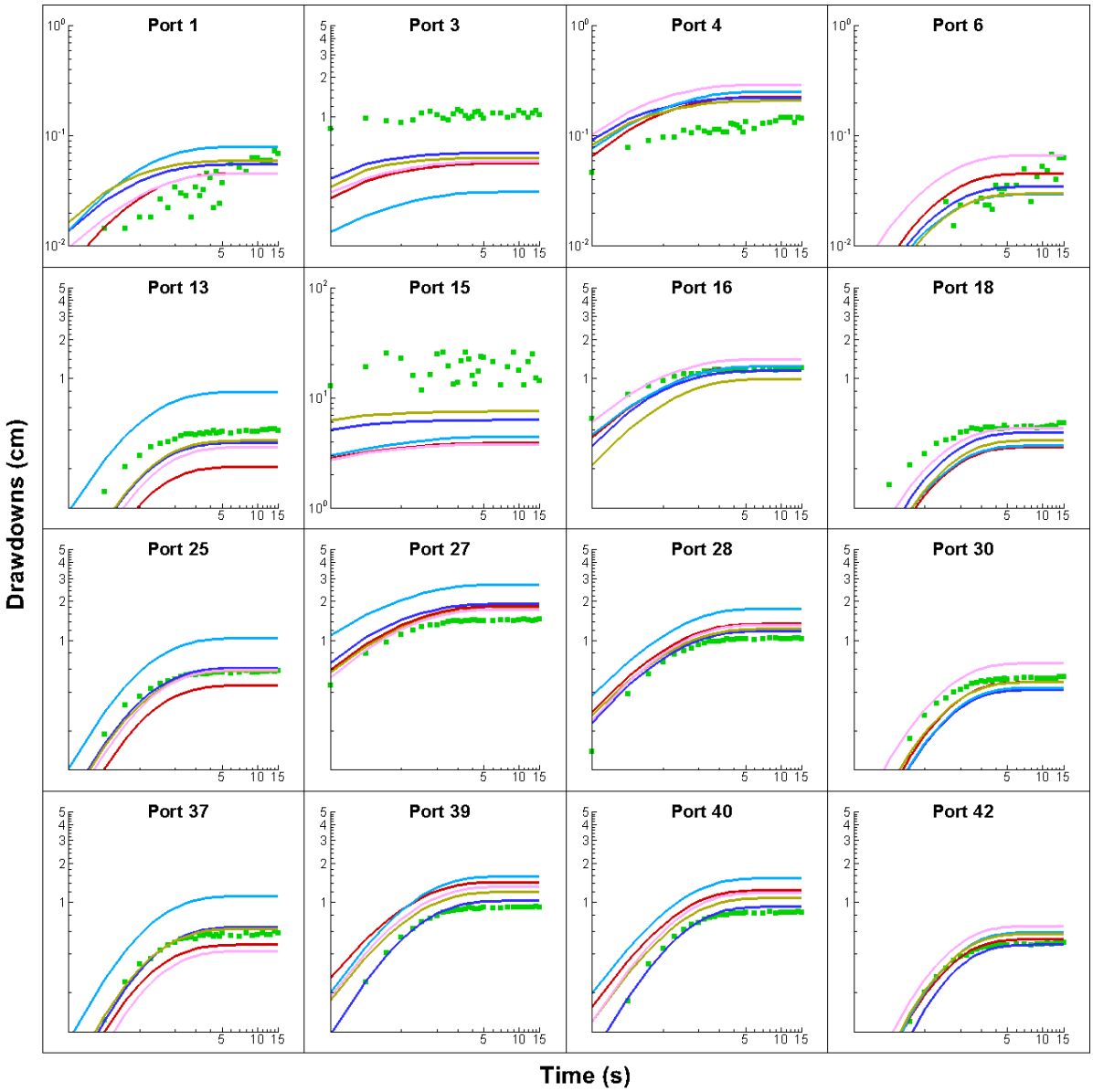

Fig. B55: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 15. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

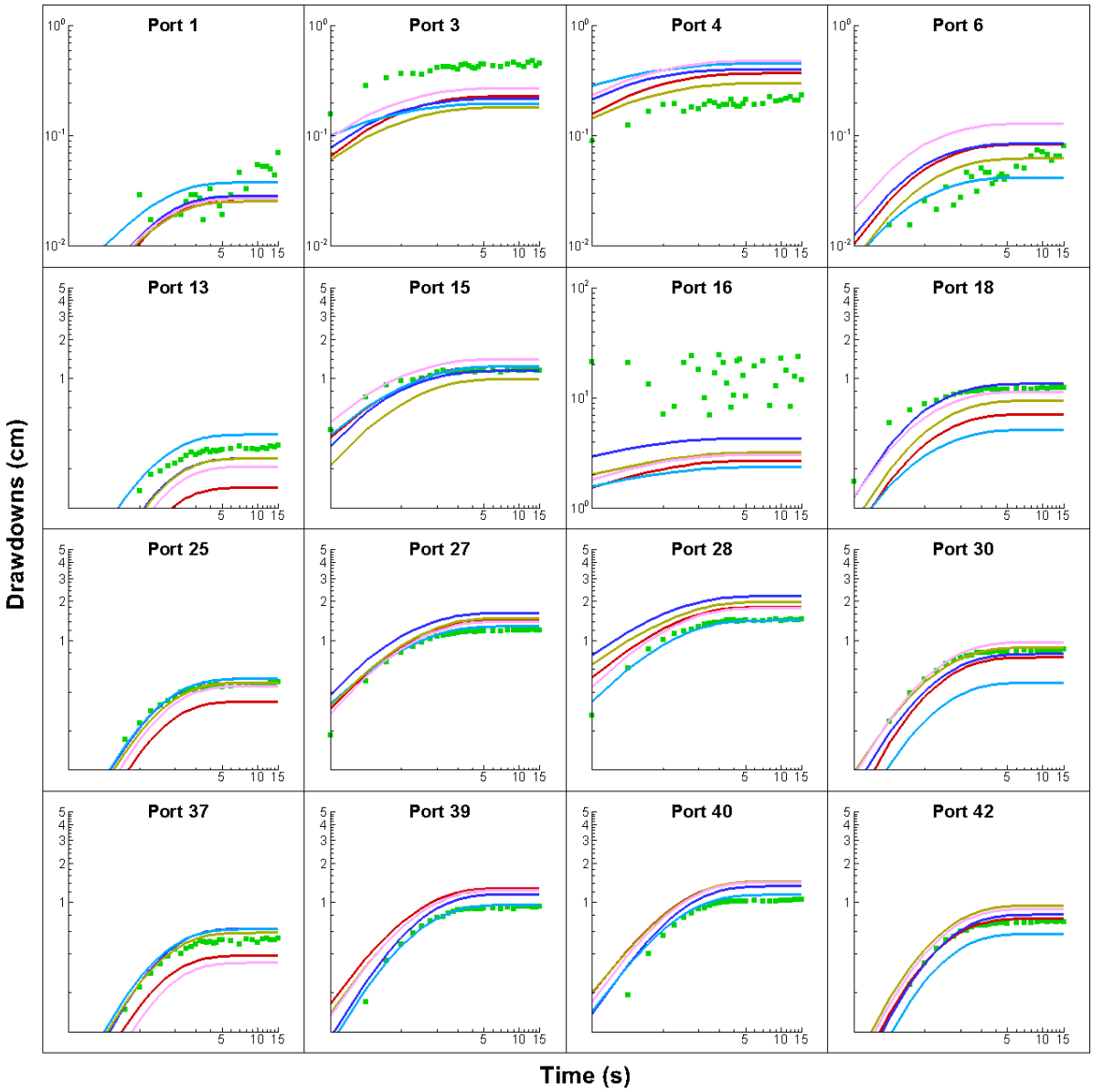

Fig. B56: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 16. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

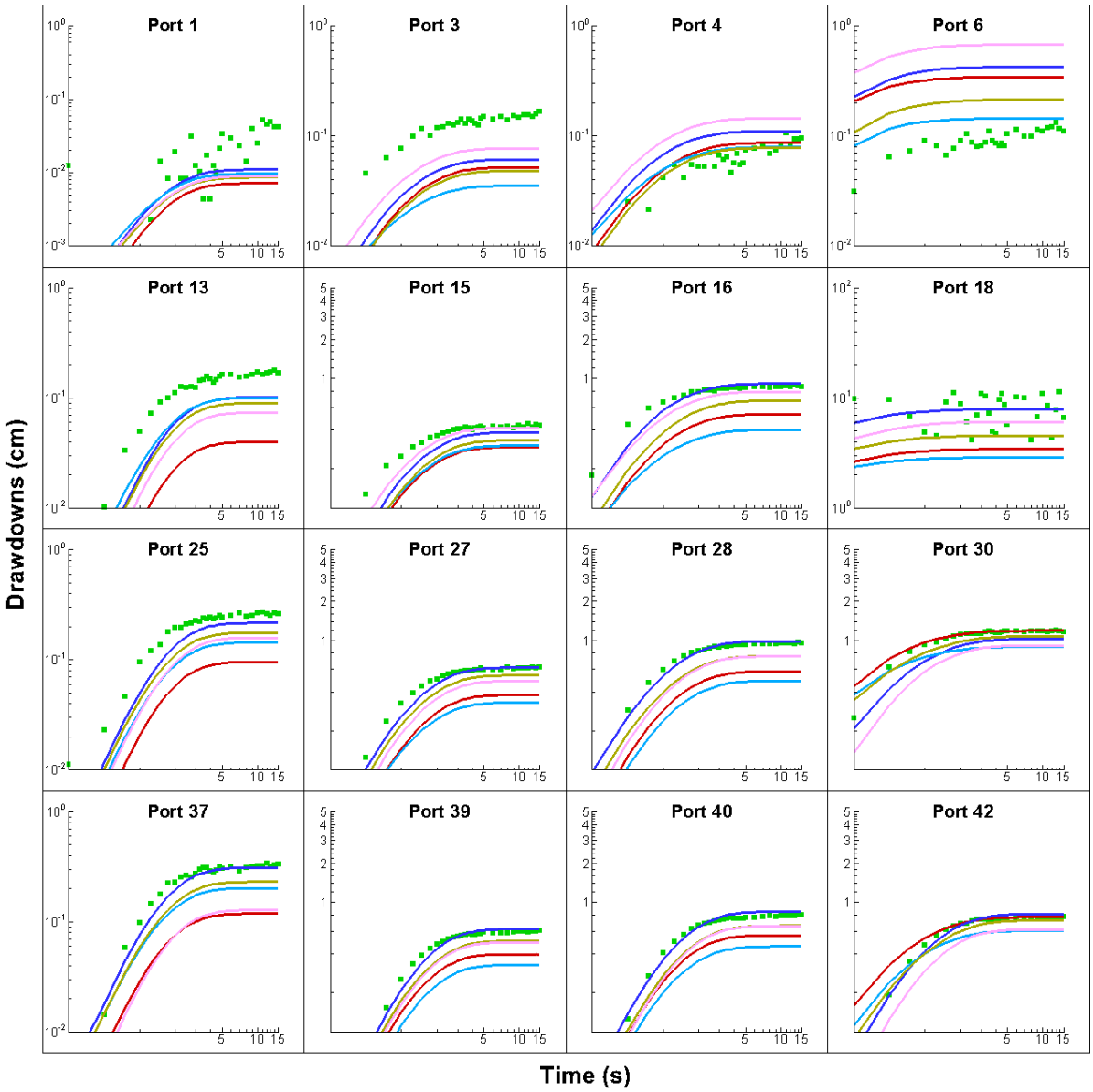

Fig. B57: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 18. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

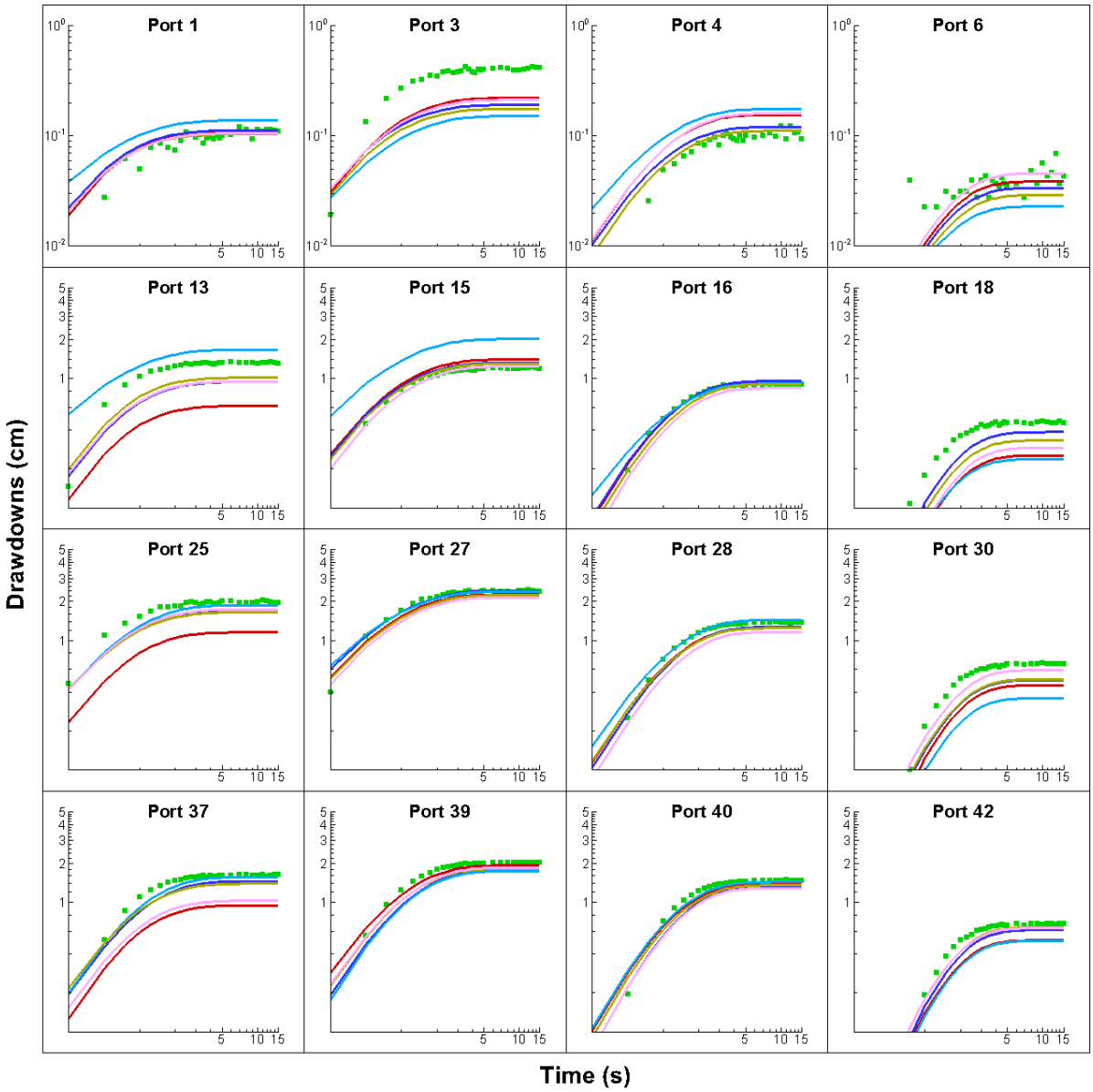

Fig. B58: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 20. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

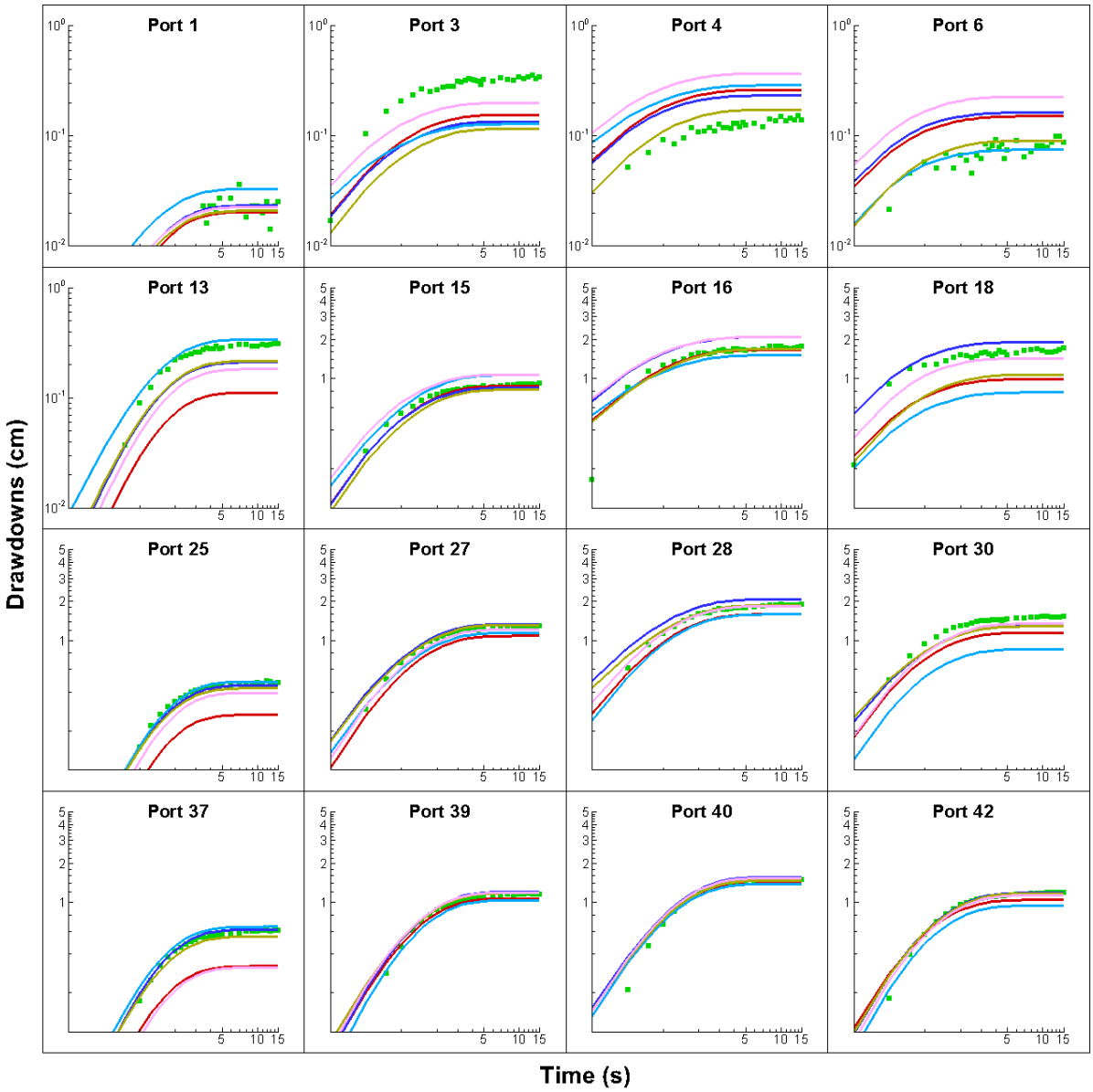

Fig. B59: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 23. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

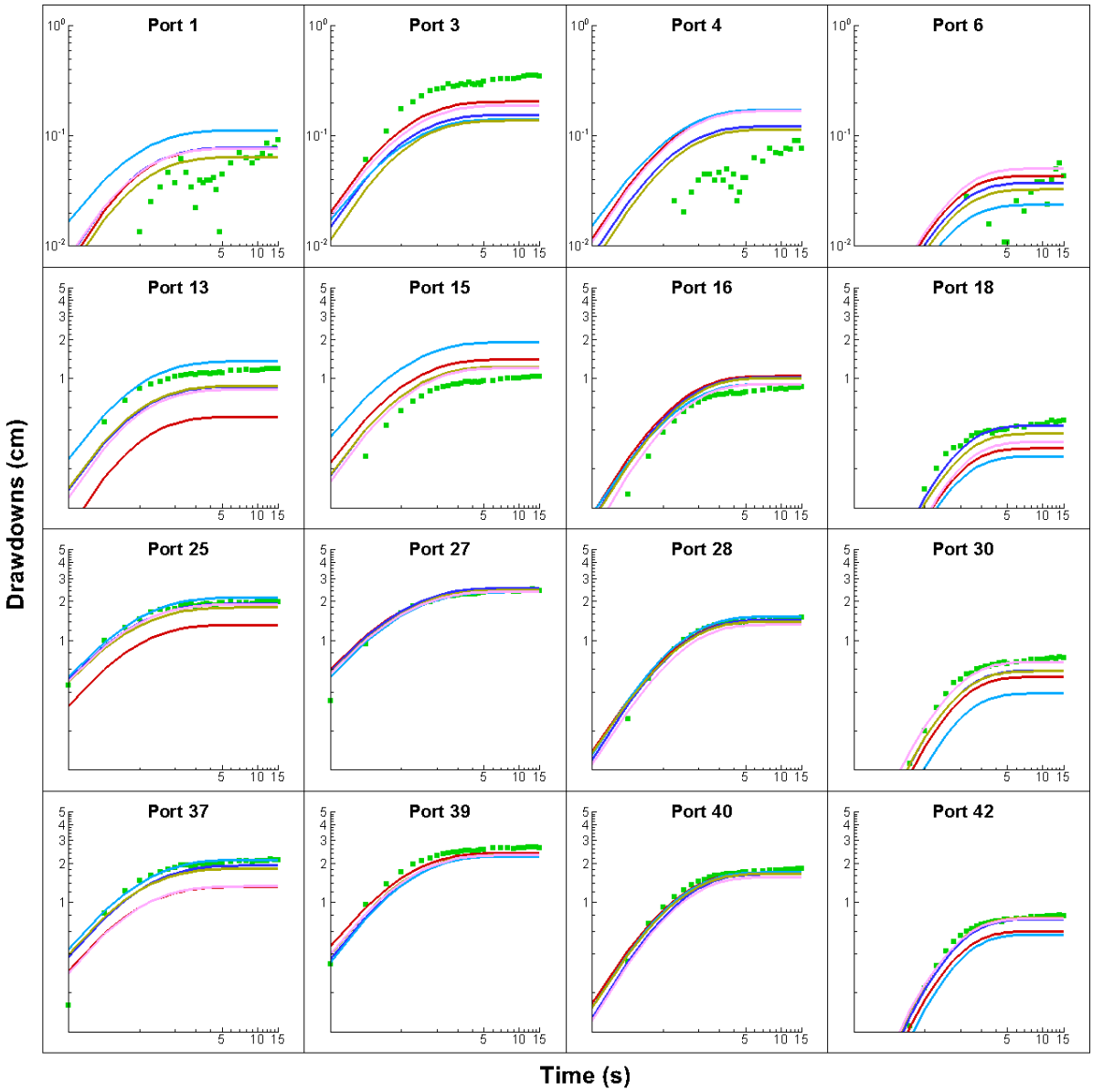

Fig. B60: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 32. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

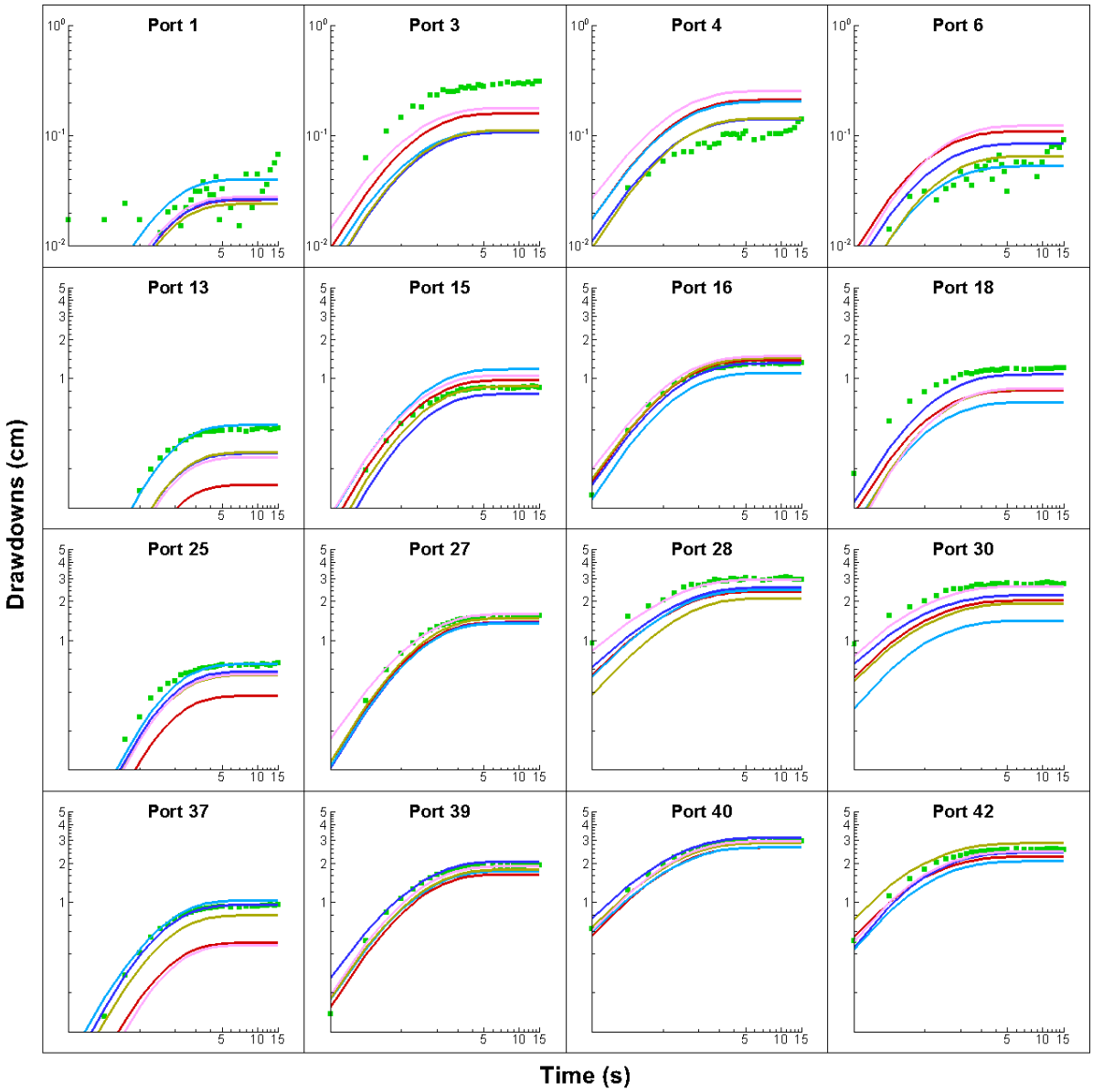

Fig. B61: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 35. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

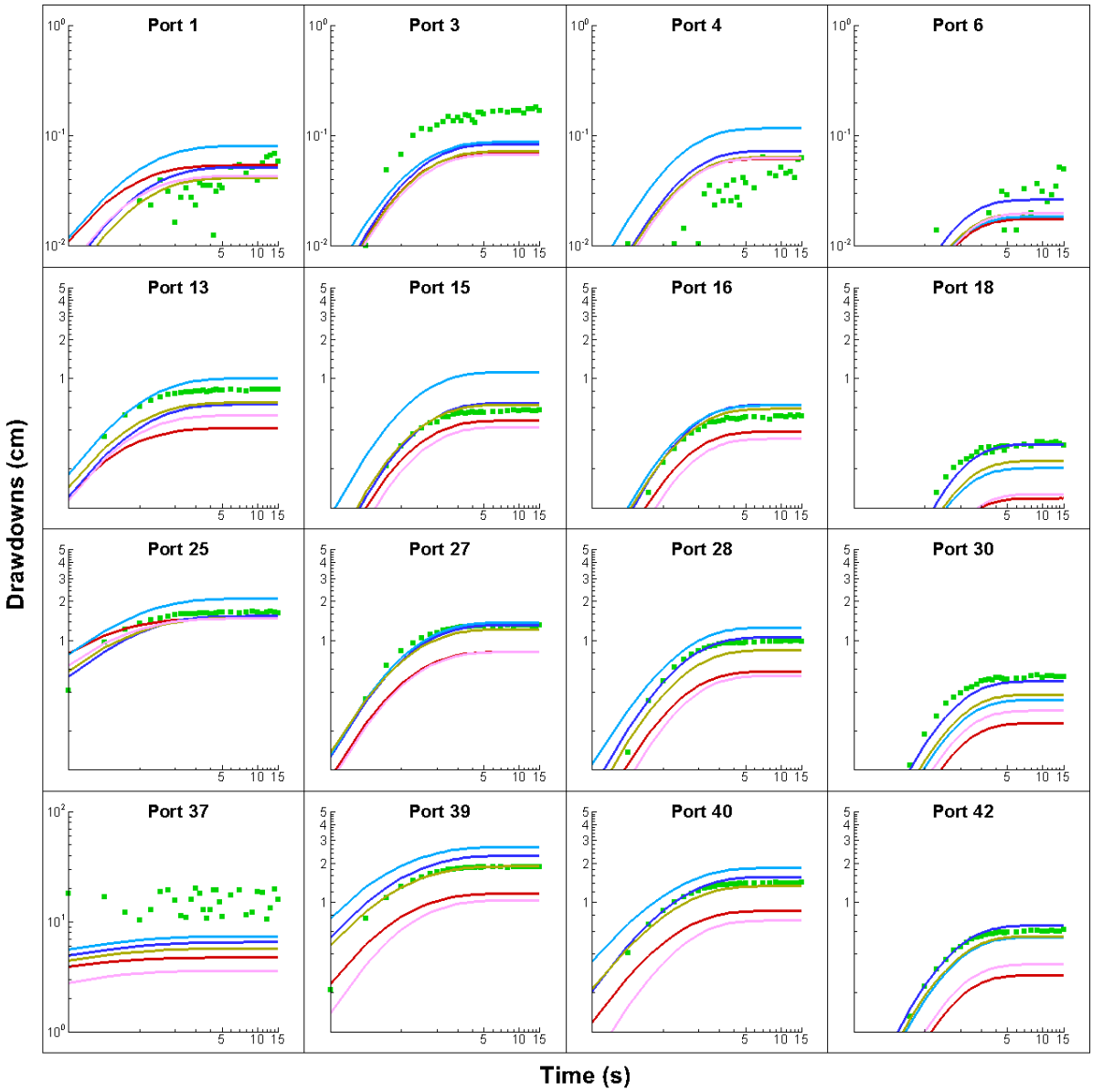

Fig. B62: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 37. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

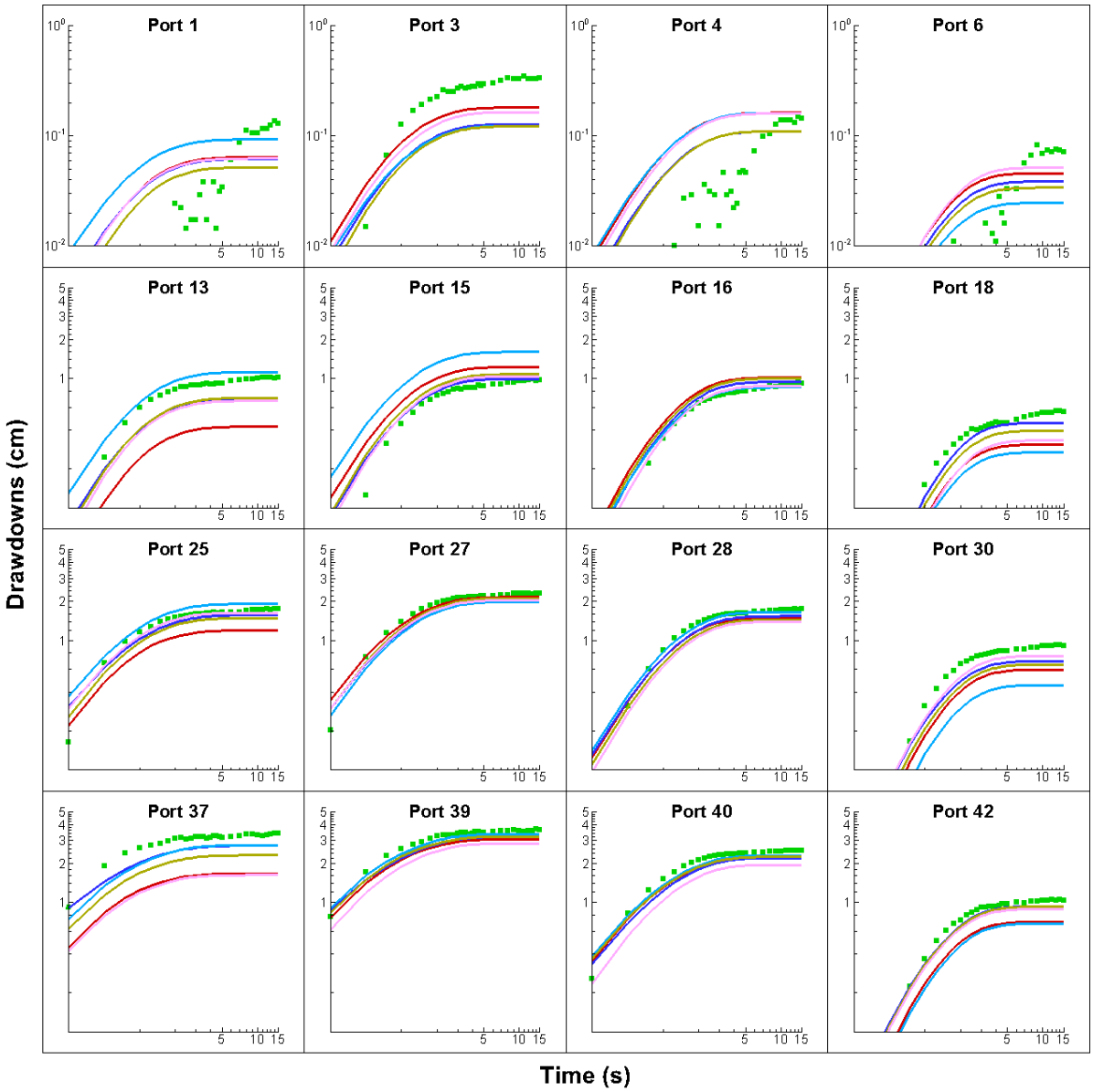

Fig. B63: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 38. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

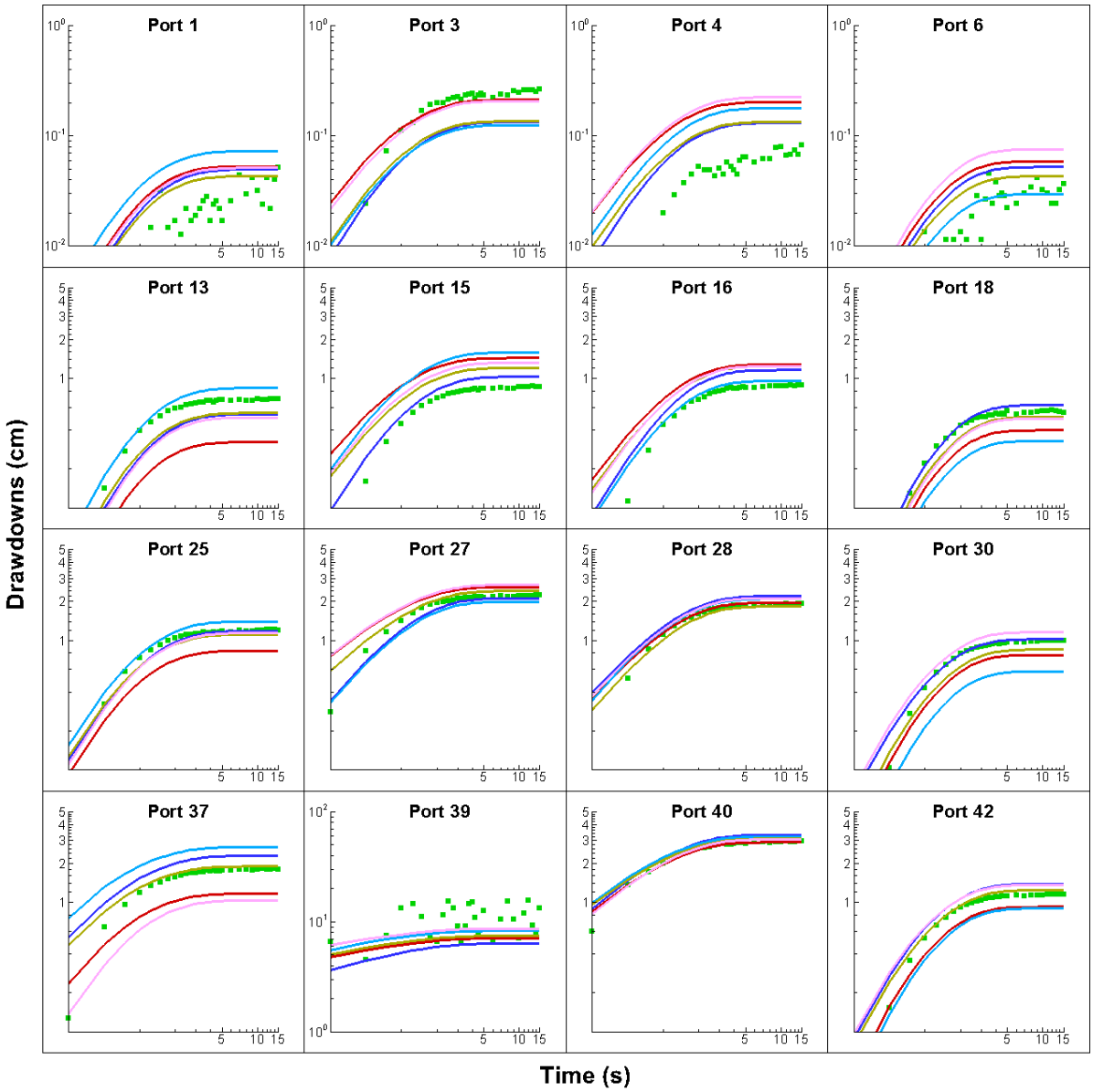

Fig. B64: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 39. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

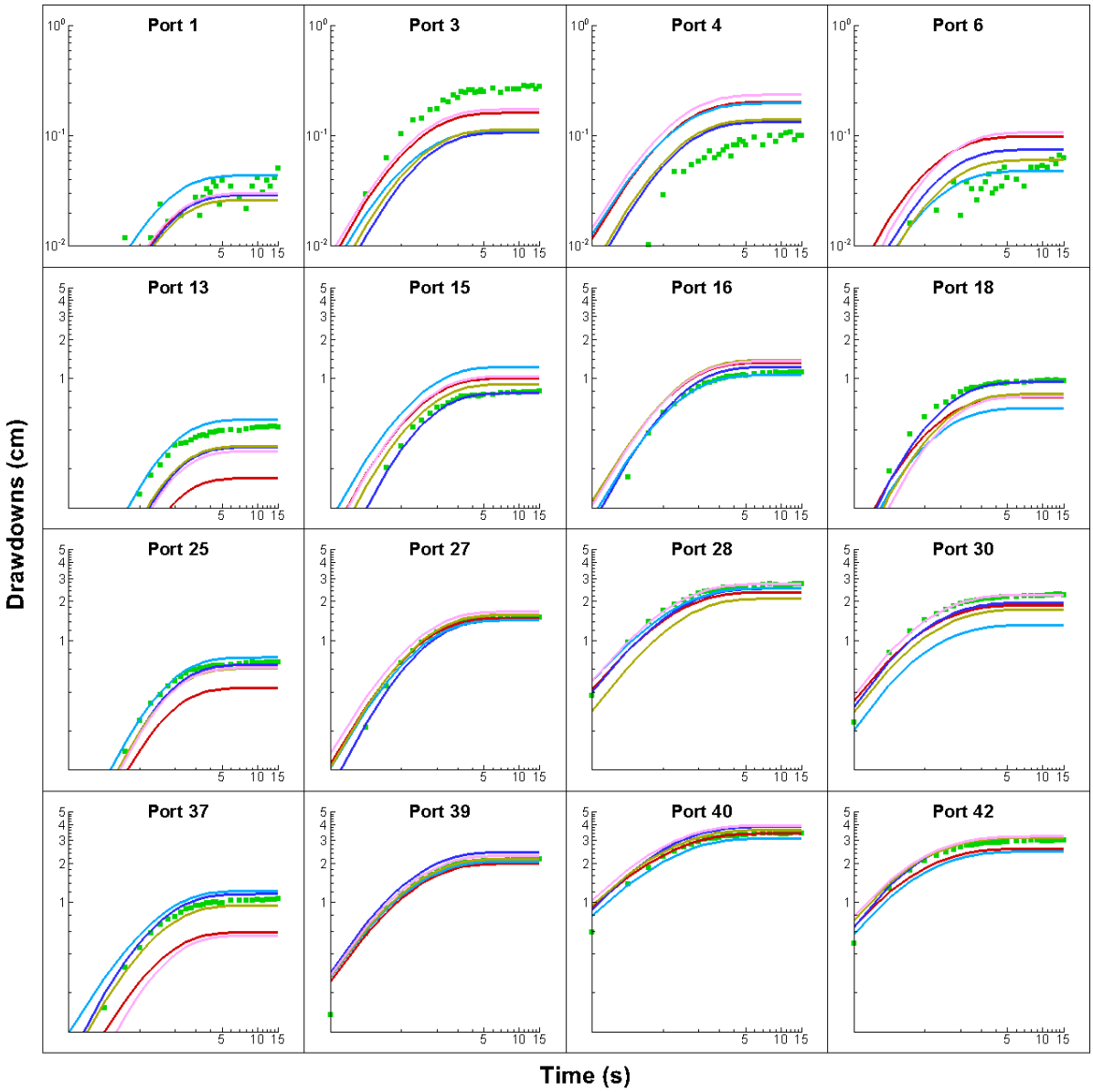

Fig. B65: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 41. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

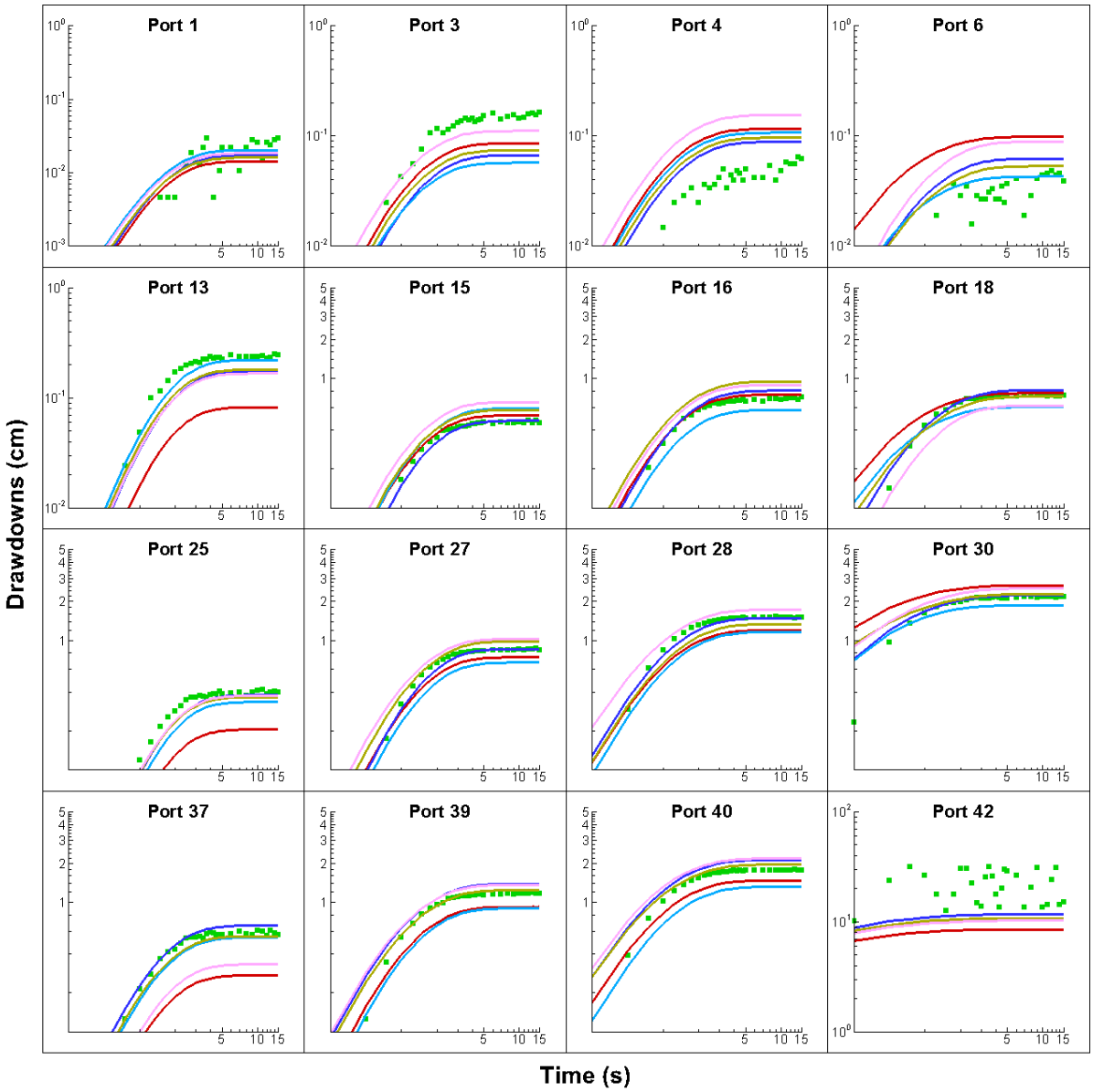

Fig. B66: Prediction of drawdown curves at 16 selected ports when conducting the pumping test at port 42. Here, the *K* and *Ss* tomograms are obtained from geostatistical models with different initial parameter fields through the simultaneous inversion of transient head data from 4 pumping tests and 15 observation ports (Case 2).

Table B1: K and S_s estimates as well as their corresponding 95% confidence intervals of the calibrated GOOD geology-based zonation model for Cases 1 and 2.

Zones	Case 1						Case 2					
	K (cm/s)			S_s (/cm)			K (cm/s)			S_s (/cm)		
	Value	*Lower*	*Upper*	*Value*	*Lower*	*Upper*	*Value*	*Lower*	*Upper*	*Value*	*Lower*	*Upper*
Zone1	2.49E-01	1.84E-01	3.36E-01	2.39E-05	2.20E-05	2.60E-05	3.09E-01	1.16E-01	8.25E-01	7.21E-05	5.52E-05	9.41E-05
Zone2	4.50E-02	2.88E-02	7.04E-02	5.75E-05	5.22E-05	6.32E-05	3.09E-02	6.80E-03	1.40E-01	3.26E-05	2.69E-05	3.95E-05
Zone3	1.02E-02	3.13E-03	3.35E-02	5.37E-05	4.96E-05	5.82E-05	4.52E-02	4.99E-03	4.10E-01	5.65E-05	3.98E-05	8.02E-05
Zone4	2.35E-01	1.97E-01	2.80E-01	5.03E-05	4.61E-05	5.49E-05	2.56E-01	1.55E-01	4.22E-01	3.95E-05	3.20E-05	4.86E-05
Zone5	3.08E-02	2.66E-02	3.58E-02	6.06E-05	5.52E-05	6.65E-05	2.18E-02	1.47E-02	3.22E-02	6.34E-05	5.16E-05	7.79E-05
Zone6	5.68E-03	4.73E-03	6.82E-03	4.68E-05	4.28E-05	5.12E-05	1.42E-02	9.14E-03	2.20E-02	3.41E-05	2.59E-05	4.48E-05
Zone7	1.62E-01	1.39E-01	1.88E-01	6.68E-05	5.95E-05	7.50E-05	9.90E-02	6.07E-02	1.61E-01	1.09E-04	9.10E-05	1.31E-04
Zone8	3.87E-02	3.40E-02	4.40E-02	5.48E-05	5.05E-05	5.95E-05	2.71E-02	2.27E-02	3.23E-02	9.05E-05	7.50E-05	1.09E-04
Zone9	1.82E-01	1.60E-01	2.08E-01	7.71E-05	7.09E-05	8.38E-05	4.24E-01	3.39E-01	5.32E-01	5.05E-05	4.22E-05	6.05E-05
Zone10	1.66E-02	1.54E-02	1.79E-02	5.21E-05	4.68E-05	5.80E-05	1.15E-02	8.63E-03	1.52E-02	1.05E-04	8.08E-05	1.37E-04
Zone11	1.12E-01	7.80E-02	1.60E-01	6.69E-05	5.91E-05	7.58E-05	1.62E-01	8.08E-02	3.26E-01	8.97E-05	6.68E-05	1.20E-04
Zone12	1.48E-01	9.80E-02	2.23E-01	1.11E-04	9.75E-05	1.25E-04	2.33E-01	1.11E-01	4.86E-01	7.95E-05	6.01E-05	1.05E-04
Zone13	2.74E-01	2.38E-01	3.15E-01	1.23E-04	1.12E-04	1.35E-04	2.88E-01	1.99E-01	4.17E-01	1.09E-04	8.71E-05	1.37E-04
Zone14	3.10E-02	2.73E-02	3.53E-02	8.73E-05	8.01E-05	9.50E-05	2.28E-02	1.28E-02	4.07E-02	9.58E-05	7.65E-05	1.20E-04
Zone15	1.03E-01	8.43E-02	1.25E-01	6.37E-05	5.80E-05	7.01E-05	3.55E-02	2.40E-02	5.25E-02	8.20E-05	6.55E-05	1.03E-04
Zone16	4.79E-02	4.21E-02	5.44E-02	9.30E-05	8.49E-05	1.02E-04	4.17E-02	2.06E-02	8.42E-02	8.43E-05	7.00E-05	1.02E-04
Zone17	4.11E-02	3.67E-02	4.61E-02	5.78E-05	5.17E-05	6.47E-05	2.92E-02	1.90E-02	4.47E-02	6.27E-05	5.06E-05	7.76E-05
Zone18	3.60E-01	3.22E-01	4.03E-01	4.71E-05	4.13E-05	5.37E-05	2.14E-01	1.24E-01	3.70E-01	6.49E-05	4.86E-05	8.68E-05

Table B2: K and S_s estimates as well as their corresponding 95% confidence intervals of the calibrated POOR1 geology-based zonation model for Cases 1 and 2.

Zones	Case 1						Case 2					
	K (cm/s)			S_s (/cm)			K (cm/s)			S_s (/cm)		
	Value	*Lower*	*Upper*	*Value*	*Lower*	*Upper*	*Value*	*Lower*	*Upper*	*Value*	*Lower*	*Upper*
Zone1	1.72E+00	1.50E+00	1.96E+00	2.79E-05	2.41E-05	3.22E-05	3.18E+00	4.88E-01	2.08E+01	5.17E-05	4.29E-05	6.23E-05
Zone2	5.03E-02	4.37E-02	5.78E-02	4.08E-05	3.80E-05	4.39E-05	2.98E-02	1.49E-02	5.95E-02	5.30E-05	4.53E-05	6.22E-05
Zone3	1.03E-01	8.84E-02	1.21E-01	5.36E-05	4.97E-05	5.78E-05	1.20E-01	7.64E-02	1.87E-01	3.51E-05	2.76E-05	4.47E-05
Zone4	6.53E-02	4.64E-02	9.19E-02	4.91E-05	4.46E-05	5.41E-05	8.64E-02	2.56E-02	2.92E-01	6.15E-05	4.99E-05	7.59E-05
Zone5	3.14E-02	2.62E-02	3.77E-02	7.21E-05	6.53E-05	7.95E-05	2.07E-02	1.15E-02	3.74E-02	6.56E-05	5.34E-05	8.06E-05
Zone6	1.80E-02	6.90E-03	4.71E-02	6.80E-05	5.97E-05	7.75E-05	1.77E-02	1.07E-05	2.96E+01	5.18E-05	4.18E-05	6.42E-05
Zone7	1.71E-02	1.53E-02	1.91E-02	6.13E-05	5.75E-05	6.54E-05	1.91E-02	1.22E-02	2.98E-02	4.22E-05	3.58E-05	4.97E-05
Zone8	3.22E-02	2.72E-02	3.81E-02	7.72E-05	7.06E-05	8.43E-05	4.34E-02	2.78E-02	6.77E-02	2.59E-05	2.00E-05	3.35E-05
Zone9	4.12E-01	3.74E-01	4.55E-01	7.07E-05	6.45E-05	7.75E-05	2.40E-01	1.84E-01	3.12E-01	9.79E-05	8.21E-05	1.17E-04
Zone10	4.93E-02	4.56E-02	5.34E-02	6.27E-05	5.81E-05	6.76E-05	4.91E-02	4.25E-02	5.69E-02	1.83E-04	1.59E-04	2.12E-04
Zone11	1.30E-01	1.08E-01	1.57E-01	1.26E-04	1.15E-04	1.38E-04	1.38E-01	7.76E-02	2.45E-01	8.04E-05	6.68E-05	9.67E-05
Zone12	9.67E-02	7.54E-02	1.24E-01	6.55E-05	5.78E-05	7.42E-05	5.50E-02	2.49E-02	1.21E-01	1.24E-04	1.00E-04	1.53E-04
Zone13	5.33E-02	4.21E-02	6.75E-02	6.79E-05	6.12E-05	7.53E-05	1.45E+00	9.59E-01	2.21E+00	6.46E-05	5.32E-05	7.85E-05
Zone14	1.16E-02	8.93E-03	1.51E-02	6.80E-05	5.95E-05	7.79E-05	2.77E-03	6.67E-04	1.15E-02	4.90E-05	4.14E-05	5.79E-05
Zone15	7.62E-02	6.39E-02	9.08E-02	8.54E-05	7.75E-05	9.41E-05	1.71E-01	9.43E-02	3.08E-01	1.06E-04	8.76E-05	1.29E-04
Zone16	3.66E-02	3.26E-02	4.10E-02	6.60E-05	5.88E-05	7.40E-05	1.10E-02	7.10E-03	1.69E-02	6.33E-05	5.42E-05	7.40E-05
Zone17	4.94E-02	4.46E-02	5.47E-02	6.66E-05	5.88E-05	7.54E-05	3.80E-02	1.87E-02	7.71E-02	6.29E-05	5.21E-05	7.61E-05
Zone18	2.81E-01	2.59E-01	3.05E-01	7.84E-05	7.09E-05	8.67E-05	2.31E-01	1.60E-01	3.34E-01	4.71E-05	3.83E-05	5.80E-05

Table B3: K and S_s estimates as well as their corresponding 95% confidence intervals of the calibrated POOR2 geology-based zonation model for Cases 1 and 2.

Zones	Case 1						Case 2					
	K (cm/s)			S_s (/cm)			K (cm/s)			S_s (/cm)		
	Value	*Lower*	*Upper*	*Value*	*Lower*	*Upper*	*Value*	*Lower*	*Upper*	*Value*	*Lower*	*Upper*
Zone1	1.67E-01	1.58E-01	1.77E-01	2.50E-05	2.04E-05	3.07E-05	1.79E-01	1.58E-01	2.03E-01	2.94E-05	1.61E-05	5.37E-05
Zone2	3.65E-02	3.46E-02	3.84E-02	7.72E-05	7.13E-05	8.36E-05	3.90E-02	3.38E-02	4.50E-02	6.22E-05	3.97E-05	9.73E-05
Zone3	2.41E-01	2.20E-01	2.63E-01	2.89E-05	2.29E-05	3.65E-05	3.40E-01	2.87E-01	4.02E-01	1.55E-04	1.08E-04	2.22E-04
Zone4	3.58E-02	3.35E-02	3.83E-02	2.59E-04	2.29E-04	2.93E-04	1.91E-02	1.46E-02	2.51E-02	8.43E-05	3.62E-05	1.97E-04
Zone5	3.67E-01	3.04E-01	4.43E-01	9.58E-05	6.60E-05	1.39E-04	3.42E-01	8.45E-02	1.38E+00	9.68E-05	1.01E-05	9.26E-04

Table B4: K and S_s estimates as well as their corresponding 95% confidence intervals of the calibrated POOR3 geology-based zonation model for Cases 1 and 2.

Zones	Case 1						Case 2					
	K (cm/s)			S_s (/cm)			K (cm/s)			S_s (/cm)		
	Value	*Lower*	*Upper*	*Value*	*Lower*	*Upper*	*Value*	*Lower*	*Upper*	*Value*	*Lower*	*Upper*
Zone1	3.92E-01	3.11E-01	4.96E-01	3.97E-05	3.40E-05	4.64E-05	1.04E+00	8.10E-01	1.33E+00	3.13E-05	1.82E-05	5.38E-05
Zone2	5.36E-02	2.22E-02	1.30E-01	6.07E-05	5.53E-05	6.66E-05	4.63E-02	1.77E-02	1.21E-01	3.90E-05	2.35E-05	6.47E-05
Zone3	9.62E-02	8.29E-02	1.12E-01	4.07E-05	3.70E-05	4.48E-05	1.61E-01	1.30E-01	1.98E-01	1.14E-04	9.14E-05	1.42E-04
Zone4	2.19E-01	1.65E-01	2.90E-01	6.23E-05	5.71E-05	6.81E-05	7.04E-02	3.86E-02	1.28E-01	4.23E-05	2.99E-05	5.96E-05
Zone5	4.12E-02	3.28E-02	5.17E-02	6.79E-05	5.70E-05	8.10E-05	4.63E-02	2.77E-02	7.72E-02	1.21E-04	8.21E-05	1.80E-04
Zone6	1.86E-02	1.68E-02	2.05E-02	7.07E-05	6.06E-05	8.25E-05	3.02E-02	2.27E-02	4.01E-02	5.41E-05	3.81E-05	7.67E-05
Zone7	4.64E-02	3.11E-02	6.93E-02	1.42E-04	1.27E-04	1.59E-04	6.10E-02	1.58E-02	2.36E-01	5.77E-05	3.49E-05	9.56E-05
Zone8	1.63E-01	9.63E-02	2.76E-01	5.85E-05	4.84E-05	7.08E-05	9.85E-02	1.52E-03	6.38E+00	6.78E-05	3.82E-05	1.20E-04
Zone9	3.32E-01	2.79E-01	3.95E-01	4.88E-05	4.29E-05	5.57E-05	2.45E-01	1.61E-01	3.72E-01	2.81E-05	1.85E-05	4.26E-05
Zone10	1.43E-01	1.14E-01	1.78E-01	6.92E-05	5.97E-05	8.02E-05	2.55E-01	1.49E-01	4.34E-01	7.99E-05	5.08E-05	1.26E-04
Zone11	6.06E-02	5.12E-02	7.17E-02	6.48E-05	5.64E-05	7.44E-05	6.19E-02	2.85E-02	1.34E-01	6.87E-05	4.44E-05	1.06E-04
Zone12	3.58E-02	3.00E-02	4.28E-02	6.98E-05	6.18E-05	7.88E-05	1.88E-02	1.17E-02	3.04E-02	1.79E-04	1.16E-04	2.78E-04
Zone13	1.65E-01	9.27E-02	2.94E-01	7.63E-05	6.42E-05	9.07E-05	6.66E-02	1.52E-03	2.92E+00	7.75E-05	4.43E-05	1.36E-04
Zone14	4.02E-02	3.36E-02	4.81E-02	8.56E-05	7.53E-05	9.73E-05	5.15E-02	2.11E-02	1.26E-01	1.40E-04	1.04E-04	1.89E-04
Zone15	2.80E-01	2.06E-01	3.81E-01	7.46E-05	6.23E-05	8.93E-05	3.78E-02	1.49E-02	9.59E-02	6.98E-05	4.35E-05	1.12E-04
Zone16	2.69E-02	2.43E-02	2.98E-02	7.02E-05	6.26E-05	7.88E-05	1.01E-01	1.11E-02	9.10E-01	6.63E-05	4.39E-05	1.00E-04
Zone17	1.14E+00	2.87E-01	4.50E+00	4.99E-05	4.25E-05	5.84E-05	1.93E-01	1.66E-03	2.24E+01	3.87E-05	2.36E-05	6.35E-05
Zone18	3.07E-01	2.56E-01	3.69E-01	3.97E-05	3.28E-05	4.81E-05	1.23E-01	6.61E-02	2.31E-01	5.82E-05	3.72E-05	9.11E-05

Table B5: L_1 norms (cm) of model calibration of 8 pumping tests for Case 1.

		PORT 2	PORT 5	PORT 14	PORT 17	PORT 32	PORT 35	PORT 44	PORT 47
Effective Parameter Model		0.213	0.257	0.209	0.181	0.233	0.255	0.388	0.262
Geology-based Zonation Model	GOOD	0.036	0.055	0.114	0.113	0.104	0.112	0.160	0.119
	POOR1	0.043	0.079	0.103	0.172	0.136	0.118	0.127	0.120
	POOR2	0.032	0.041	0.091	0.155	0.133	0.194	0.226	0.165
	POOR3	0.041	0.066	0.137	0.167	0.144	0.154	0.208	0.163
Geostatistical Model with Homogeneous Initial Fields		0.030	0.030	0.045	0.048	0.057	0.056	0.049	0.053
Geostatistical Model with Heterogeneous Initial Fields	GOOD	0.030	0.031	0.050	0.049	0.061	0.061	0.057	0.063
	POOR1	0.032	0.029	0.053	0.049	0.065	0.062	0.052	0.063
	POOR2	0.030	0.032	0.048	0.052	0.063	0.062	0.051	0.060
	POOR3	0.031	0.029	0.052	0.055	0.066	0.061	0.053	0.055
Min		0.029							
Max		0.388							

Table B6: L_2 norms (cm^2) of model calibration of 8 pumping tests for Case 1.

		PORT 2	PORT 5	PORT 14	PORT 17	PORT 32	PORT 35	PORT 44	PORT 47
Effective Parameter Model		0.077	0.114	0.092	0.060	0.101	0.118	0.304	0.126
Geology-based Zonation Model	GOOD	0.002	0.005	0.037	0.028	0.027	0.025	0.059	0.030
	POOR1	0.003	0.011	0.026	0.065	0.041	0.028	0.032	0.028
	POOR2	0.002	0.003	0.026	0.061	0.037	0.096	0.126	0.064
	POOR3	0.004	0.007	0.046	0.048	0.038	0.050	0.118	0.056
Geostatistical Model with Homogeneous Initial Fields		0.001	0.001	0.005	0.005	0.007	0.006	0.005	0.007
Geostatistical Model with Heterogeneous Initial Fields	GOOD	0.001	0.001	0.005	0.005	0.008	0.008	0.007	0.009
	POOR1	0.002	0.001	0.006	0.005	0.008	0.007	0.005	0.008
	POOR2	0.001	0.001	0.005	0.006	0.008	0.007	0.005	0.008
	POOR3	0.002	0.001	0.006	0.006	0.009	0.007	0.005	0.007
Min		0.001							
Max		0.304							

Table B7: L_1 norms (cm) of model validation of 16 pumping tests for Case 1.

		PORT 8	PORT 11	PORT 13	PORT 15	PORT 16	PORT 18	PORT 20	PORT 23	PORT 26	PORT 29	PORT 37	PORT 38	PORT 39	PORT 40	PORT 41	PORT 42
Effective Parameter Model		0.301	0.129	0.141	0.246	0.193	0.136	0.265	0.218	0.242	0.252	0.207	0.284	0.261	0.267	0.233	0.188
Geology-based Zonation Model	GOOD	0.064	0.072	0.076	0.105	0.116	0.096	0.154	0.120	0.125	0.110	0.128	0.135	0.183	0.128	0.116	0.153
	POOR1	0.077	0.139	0.138	0.187	0.193	0.151	0.140	0.207	0.168	0.123	0.116	0.146	0.209	0.125	0.132	0.179
	POOR2	0.059	0.107	0.053	0.167	0.209	0.114	0.157	0.175	0.145	0.239	0.133	0.165	0.193	0.182	0.167	0.151
	POOR3	0.160	0.150	0.111	0.165	0.133	0.146	0.193	0.188	0.150	0.173	0.121	0.155	0.242	0.165	0.158	0.132
Geostatistical Model with Homogeneous Initial Fields		0.045	0.048	0.105	0.082	0.097	0.085	0.086	0.082	0.099	0.077	0.088	0.076	0.174	0.097	0.071	0.140
Geostatistical Model with Heterogeneous Initial Fields	GOOD	0.046	0.040	0.096	0.080	0.102	0.069	0.072	0.065	0.089	0.088	0.078	0.082	0.187	0.114	0.072	0.117
	POOR1	0.054	0.044	0.139	0.097	0.086	0.069	0.091	0.084	0.098	0.087	0.088	0.095	0.189	0.112	0.079	0.124
	POOR2	0.046	0.045	0.086	0.096	0.120	0.079	0.079	0.078	0.096	0.079	0.095	0.074	0.186	0.107	0.067	0.112
	POOR3	0.045	0.052	0.097	0.095	0.126	0.086	0.071	0.083	0.096	0.091	0.076	0.065	0.184	0.098	0.072	0.122
Min		0.040															
Max		0.301															

Table B8: L_2 norms (cm^2) of model validation of 16 pumping tests for Case 1.

		PORT 8	PORT 11	PORT 13	PORT 15	PORT 16	PORT 18	PORT 20	PORT 23	PORT 26	PORT 29	PORT 37	PORT 38	PORT 39	PORT 40	PORT 41	PORT 42
Effective Parameter Model		0.154	0.046	0.053	0.143	0.076	0.038	0.116	0.093	0.091	0.112	0.073	0.134	0.126	0.119	0.089	0.074
Geology-based Zonation Model	GOOD	0.010	0.011	0.017	0.034	0.038	0.030	0.045	0.032	0.029	0.023	0.058	0.038	0.095	0.043	0.028	0.081
	POOR1	0.012	0.036	0.040	0.063	0.084	0.066	0.034	0.112	0.049	0.051	0.031	0.042	0.148	0.026	0.037	0.151
	POOR2	0.010	0.023	0.006	0.079	0.102	0.037	0.040	0.084	0.037	0.127	0.040	0.060	0.103	0.085	0.090	0.095
	POOR3	0.042	0.035	0.038	0.076	0.048	0.065	0.064	0.105	0.040	0.086	0.029	0.062	0.144	0.051	0.054	0.042
Geostatistical Model with Homogeneous Initial Fields		0.007	0.004	0.034	0.016	0.028	0.032	0.014	0.020	0.016	0.017	0.022	0.016	0.071	0.020	0.011	0.085
Geostatistical Model with Heterogeneous Initial Fields	GOOD	0.007	0.003	0.024	0.017	0.022	0.019	0.010	0.009	0.014	0.017	0.016	0.015	0.086	0.034	0.013	0.059
	POOR1	0.008	0.004	0.056	0.023	0.024	0.018	0.016	0.022	0.017	0.021	0.024	0.024	0.105	0.030	0.015	0.067
	POOR2	0.007	0.004	0.017	0.023	0.030	0.023	0.011	0.021	0.015	0.021	0.024	0.017	0.086	0.028	0.011	0.047
	POOR3	0.005	0.005	0.034	0.023	0.045	0.032	0.009	0.016	0.016	0.018	0.015	0.008	0.078	0.022	0.014	0.053
Min		0.003															
Max		0.154															

Table B9: L_1 norms (cm) of model calibration of 4 pumping tests for Case 2.

		PORT 26	PORT 29	PORT 44	PORT 47
Effective Parameter Model		0.283	0.315	0.380	0.289
Geology-based Zonation Model	GOOD	0.071	0.070	0.103	0.073
	POOR1	0.078	0.114	0.102	0.092
	POOR2	0.122	0.265	0.175	0.152
	POOR3	0.170	0.140	0.153	0.174
Geostatistical Model with Homogeneous Initial Fields		0.043	0.039	0.041	0.046
Geostatistical Model with Heterogeneous Initial Fields	GOOD	0.044	0.044	0.052	0.054
	POOR1	0.056	0.043	0.047	0.055
	POOR2	0.045	0.042	0.045	0.059
	POOR2	0.048	0.042	0.045	0.055
Min		0.039			
Max		0.380			

Table B10: L_2 norms (cm^2) of model calibration of 4 pumping tests for Case 2.

		PORT 26	PORT 29	PORT 44	PORT 47
Effective Parameter Model		0.111	0.166	0.241	0.150
Geology-based Zonation Model	GOOD	0.010	0.011	0.025	0.017
	POOR1	0.012	0.024	0.020	0.023
	POOR2	0.034	0.187	0.076	0.063
	POOR3	0.049	0.039	0.033	0.050
Geostatistical Model with Homogeneous Initial Fields		0.003	0.003	0.004	0.006
Geostatistical Model with Heterogeneous Initial Fields	GOOD	0.004	0.004	0.008	0.008
	POOR1	0.006	0.004	0.006	0.008
	POOR2	0.004	0.005	0.005	0.008
	POOR2	0.004	0.004	0.005	0.009
Min		0.003			
Max		0.241			

Table B11: L_1 norms (cm) of model validation of 16 pumping tests for Case 2.

		PORT 8	PORT 11	PORT 13	PORT 15	PORT 16	PORT 18	PORT 20	PORT 23	PORT 34	PORT 35	PORT 37	PORT 38	PORT 39	PORT 40	PORT 41	PORT 42
Effective Parameter Model		0.288	0.123	0.132	0.230	0.182	0.135	0.273	0.209	0.233	0.266	0.215	0.289	0.254	0.268	0.238	0.190
Geology-based Zonation Model	GOOD	0.074	0.092	0.073	0.142	0.197	0.156	0.136	0.181	0.129	0.116	0.096	0.141	0.217	0.168	0.111	0.188
	POOR1	0.103	0.120	0.241	0.303	0.159	0.151	0.123	0.151	0.153	0.197	0.136	0.160	0.204	0.160	0.146	0.178
	POOR2	0.070	0.091	0.062	0.205	0.257	0.136	0.151	0.217	0.140	0.187	0.129	0.162	0.187	0.183	0.167	0.157
	POOR3	0.153	0.193	0.113	0.212	0.167	0.207	0.194	0.174	0.158	0.150	0.165	0.179	0.234	0.186	0.175	0.137
Geostatistical Model with Homogeneous Initial Fields		0.063	0.076	0.140	0.191	0.176	0.123	0.153	0.101	0.156	0.169	0.217	0.166	0.216	0.153	0.101	0.187
Geostatistical Model with Heterogeneous Initial Fields	GOOD	0.067	0.092	0.071	0.129	0.168	0.120	0.128	0.142	0.114	0.119	0.099	0.093	0.236	0.169	0.113	0.188
	POOR1	0.070	0.085	0.183	0.335	0.133	0.160	0.104	0.120	0.136	0.184	0.167	0.123	0.265	0.171	0.119	0.184
	POOR2	0.078	0.072	0.076	0.163	0.179	0.118	0.139	0.097	0.115	0.138	0.089	0.120	0.209	0.145	0.116	0.216
	POOR3	0.061	0.073	0.097	0.180	0.181	0.153	0.131	0.113	0.120	0.109	0.245	0.168	0.230	0.183	0.135	0.212
Min		0.061															
Max		0.335															

Table B12: L_2 norms (cm^2) of model validation of 16 pumping tests for Case 2.

		PORT 8	PORT 11	PORT 13	PORT 15	PORT 16	PORT 18	PORT 20	PORT 23	PORT 34	PORT 35	PORT 37	PORT 38	PORT 39	PORT 40	PORT 41	PORT 42
Effective Parameter Model		0.144	0.043	0.048	0.131	0.069	0.037	0.117	0.084	0.096	0.130	0.079	0.142	0.115	0.124	0.094	0.074
Geology-based Zonation Model	GOOD	0.020	0.024	0.017	0.093	0.098	0.126	0.047	0.105	0.035	0.029	0.022	0.051	0.099	0.055	0.026	0.097
	POOR1	0.038	0.031	0.150	0.153	0.067	0.053	0.031	0.050	0.059	0.082	0.041	0.050	0.122	0.050	0.051	0.209
	POOR2	0.024	0.027	0.020	0.145	0.141	0.128	0.072	0.101	0.038	0.091	0.039	0.058	0.092	0.081	0.087	0.093
	POOR3	0.050	0.065	0.051	0.115	0.078	0.074	0.070	0.074	0.043	0.040	0.058	0.071	0.102	0.072	0.057	0.053
Geostatistical Model with Homogeneous Initial Fields		0.011	0.012	0.039	0.084	0.064	0.028	0.065	0.022	0.059	0.054	0.102	0.087	0.104	0.045	0.021	0.173
Geostatistical Model with Heterogeneous Initial Fields	GOOD	0.017	0.025	0.019	0.073	0.069	0.098	0.062	0.058	0.028	0.031	0.021	0.019	0.126	0.073	0.034	0.153
	POOR1	0.020	0.020	0.092	0.211	0.054	0.060	0.024	0.041	0.043	0.085	0.061	0.033	0.207	0.061	0.035	0.227
	POOR2	0.030	0.013	0.034	0.081	0.064	0.041	0.097	0.019	0.033	0.047	0.034	0.041	0.110	0.048	0.029	0.236
	POOR3	0.011	0.014	0.059	0.065	0.066	0.103	0.041	0.035	0.033	0.025	0.138	0.097	0.102	0.070	0.039	0.168
Min		0.011															
Max		0.236															

Appendix C: Supplementary Information for Study II

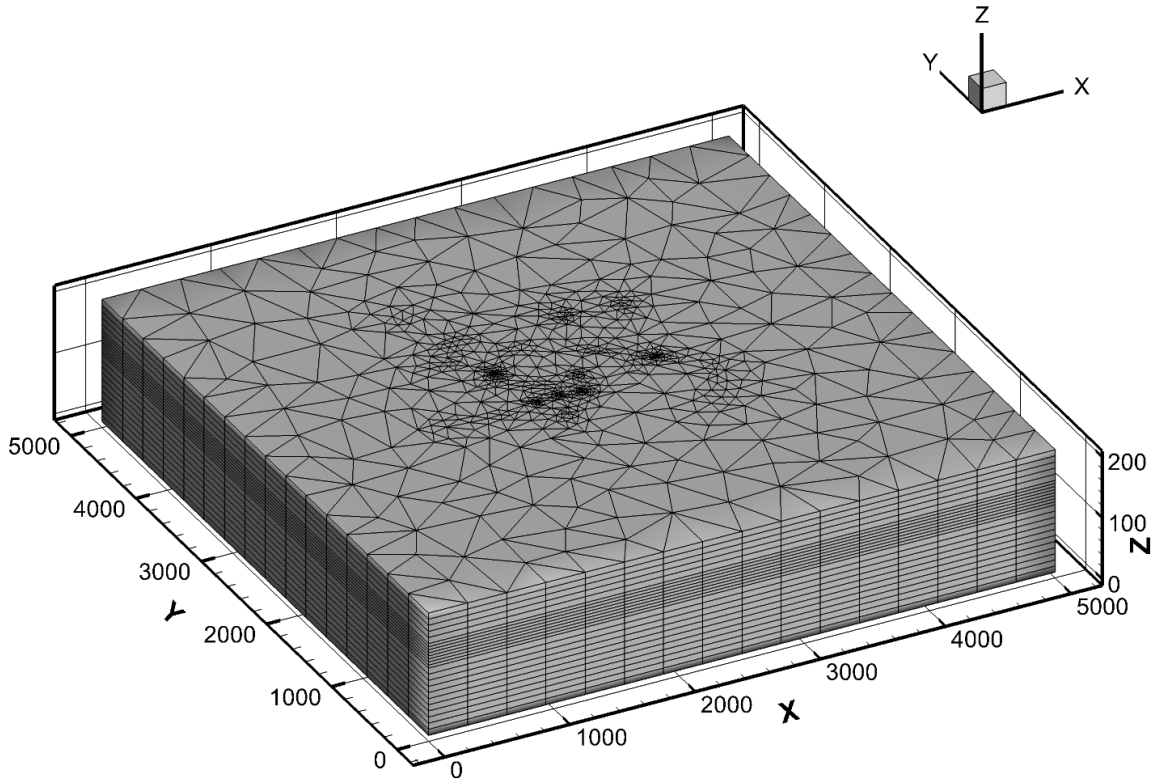

Fig. C1: Finite element mesh used for simulations.

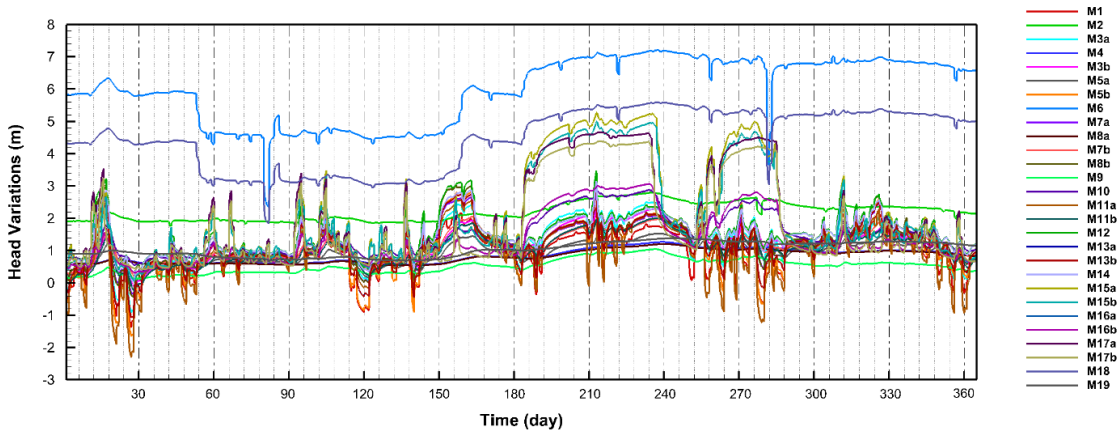

Fig. C2: Generated synthetic drawdown variations (municipal well data) at all 28 monitoring locations during the year of 2013 for model calibration (0 – 120 days) and validation (180 – 365 days).

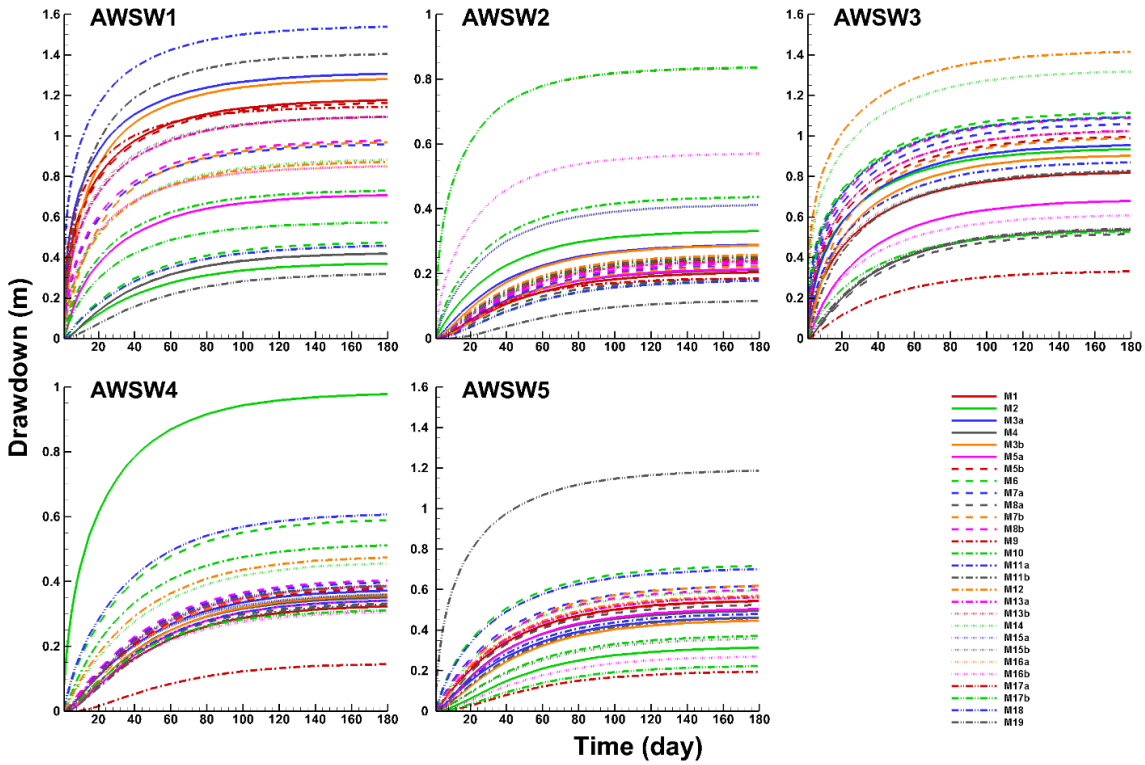

Fig. C3: Generated independent pumping test data from five additional water-supply wells for model validation.

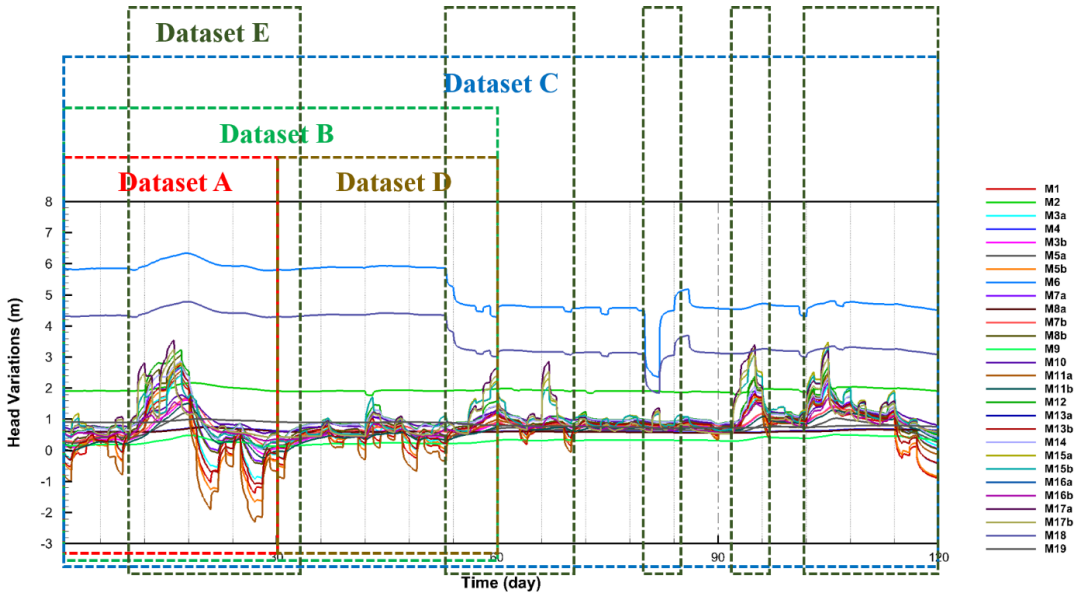

Fig. C4: Selected datasets for model calibration. Dataset A (0-30 days), Dataset B (0-60 days), Dataset C (0-120 days), Dataset D (30-60 days), Dataset E (0-120 days with selected large drawdown variation periods).

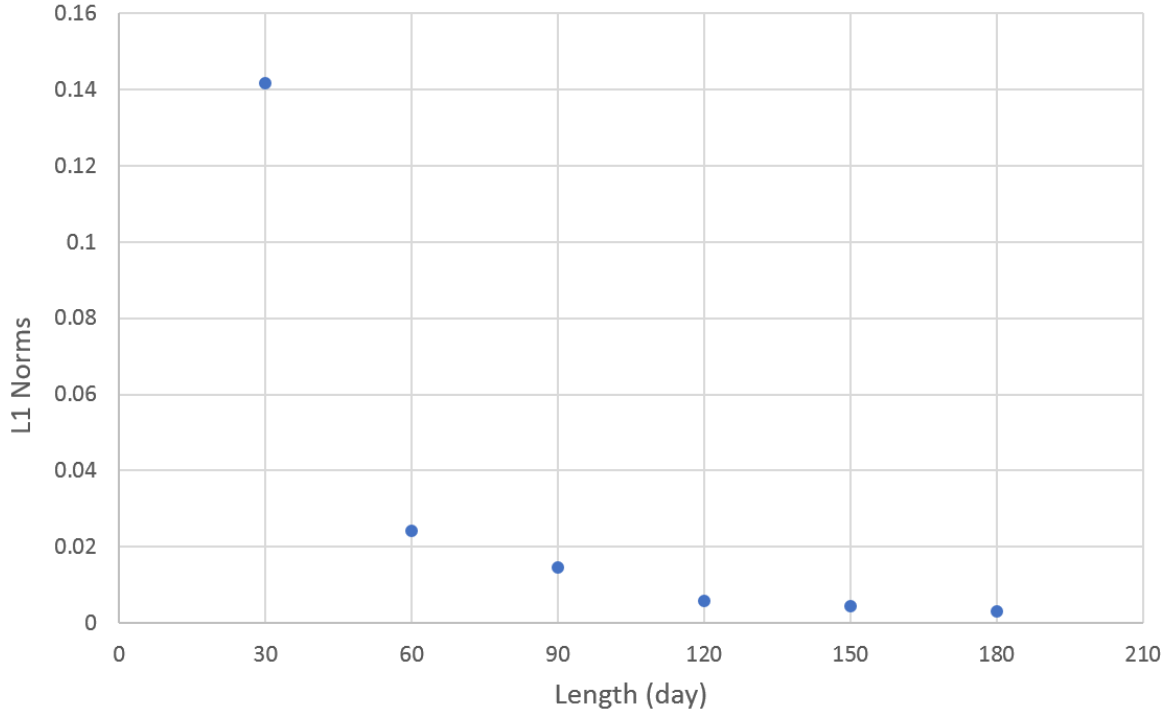

Fig. C5: The calculated mean absolute error (L_1) between simulated and observed drawdowns when sets of prior pumping/injection records with different lengths were incorporated for simulations.

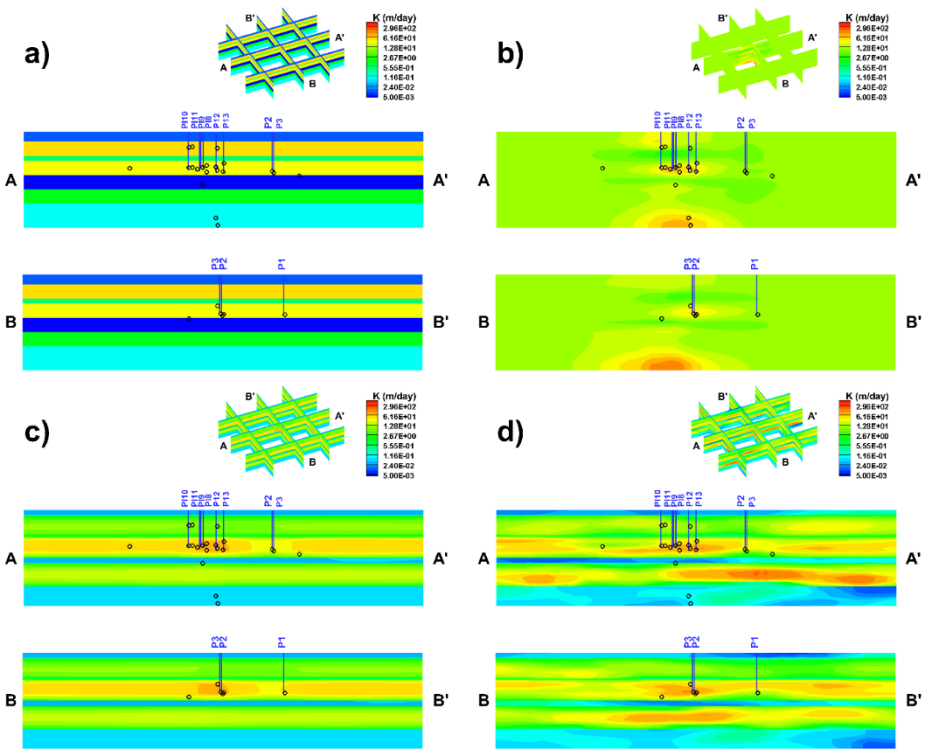

Fig. C6: Estimated K tomograms from three model cases through the interpretation of **Dataset B** as well as the "true" K field. a) Case 2, b) Case 3a, c) Case 3b, and d) "true" K field. In each contour map, small black circles represent the location of monitoring well screens.

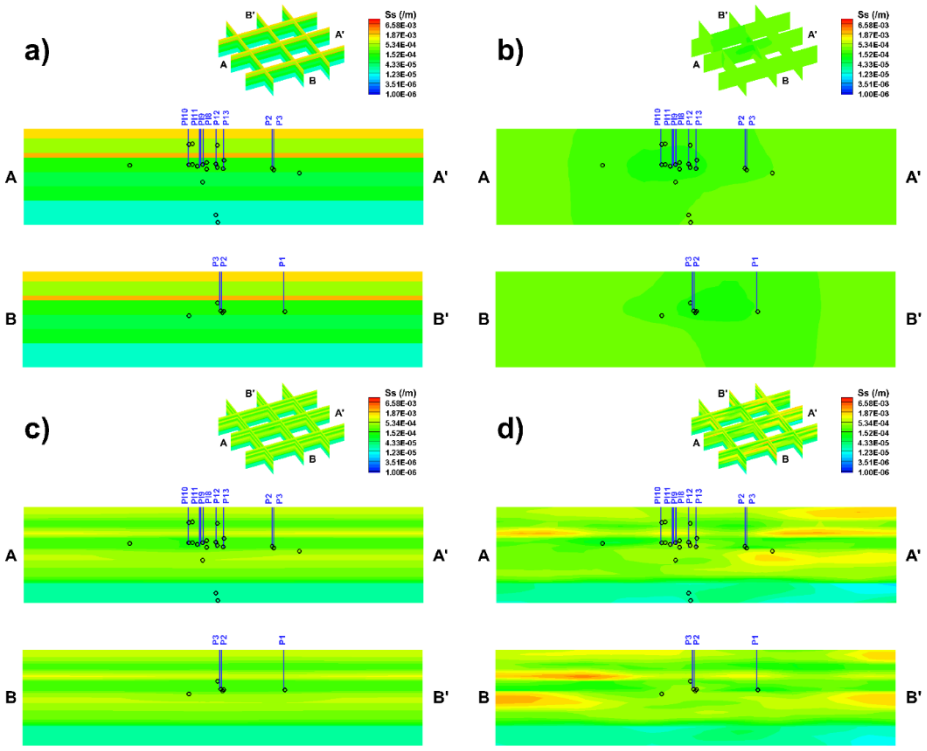

Fig. C7: Estimated S_s tomograms from three model cases through the interpretation of **Dataset B** as well as the "true" S_s field. a) Case 2, b) Case 3a, c) Case 3b, and d) "true" S_s field. In each contour map, small black circles represent the location of monitoring well screens.

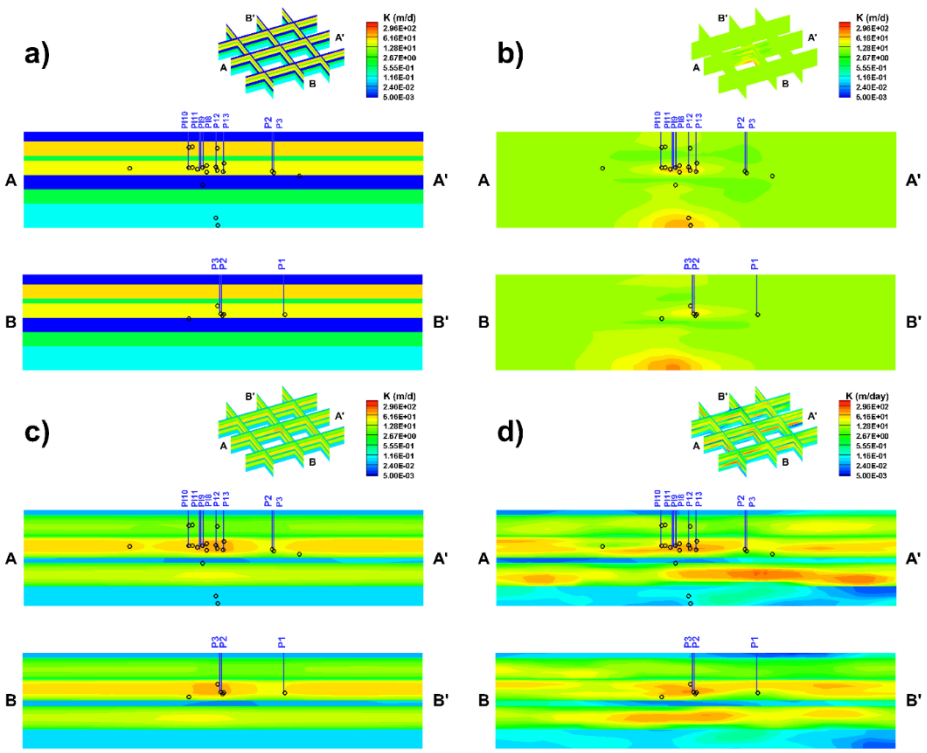

Fig. C8: Estimated *K* tomograms from three model cases through the interpretation of **Dataset C** as well as the "true" *K* field. a) Case 2, b) Case 3a, c) Case 3b, and d) "true" *K* field. In each contour map, small black circles represent the location of monitoring well screens.

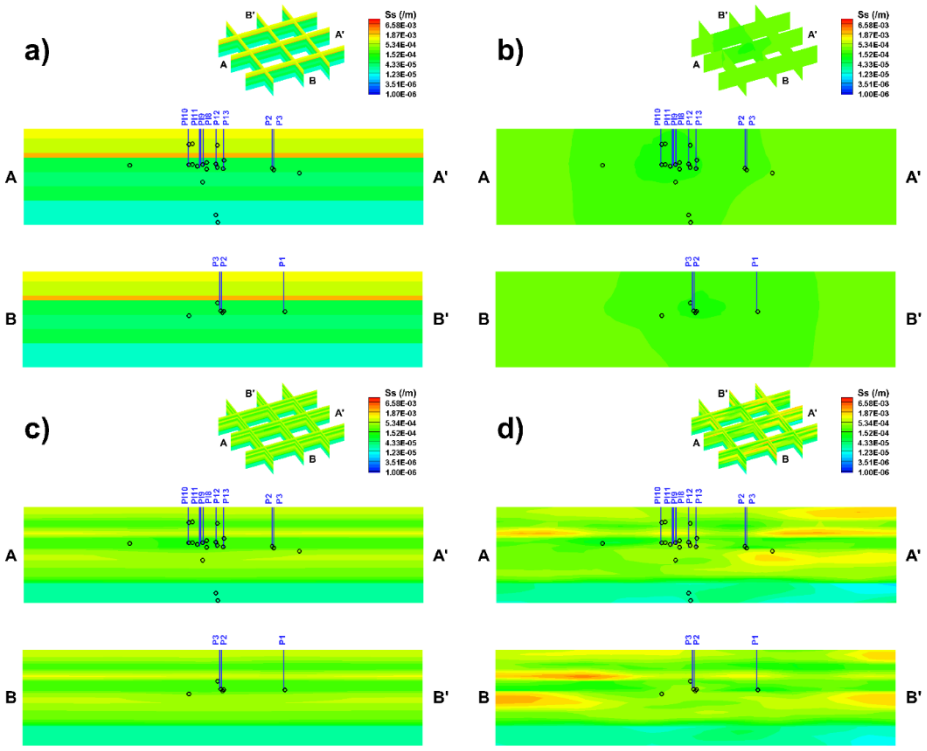

Fig. C9: Estimated S_s tomograms from three model cases through the interpretation of **Dataset C** as well as the "true" S_s field. a) Case 2, b) Case 3a, c) Case 3b, and d) "true" S_s field. In each contour map, small black circles represent the location of monitoring well screens.

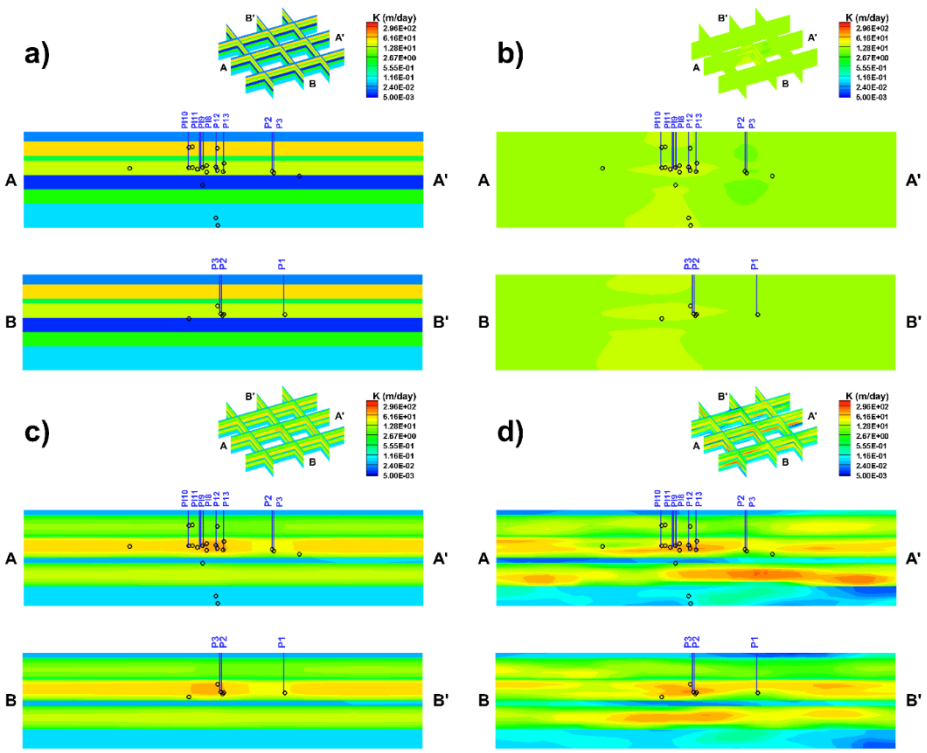

Fig. C10: Estimated *K* tomograms from three model cases through the interpretation of **Dataset D** as well as the "true" *K* field. a) Case 2, b) Case 3a, c) Case 3b, and d) "true" *K* field. In each contour map, small black circles represent the location of monitoring well screens.

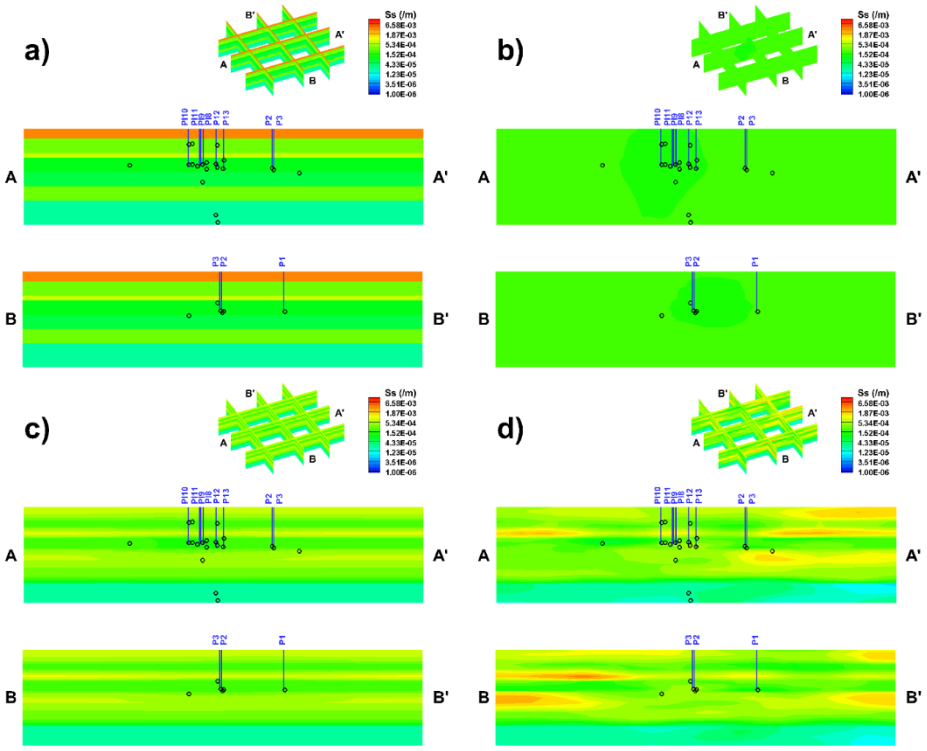

Fig. C11: Estimated S_s tomograms from three model cases through the interpretation of **Dataset D** as well as the "true" S_s field. a) Case 2, b) Case 3a, c) Case 3b, and d) "true" S_s field. In each contour map, small black circles represent the location of monitoring well screens.

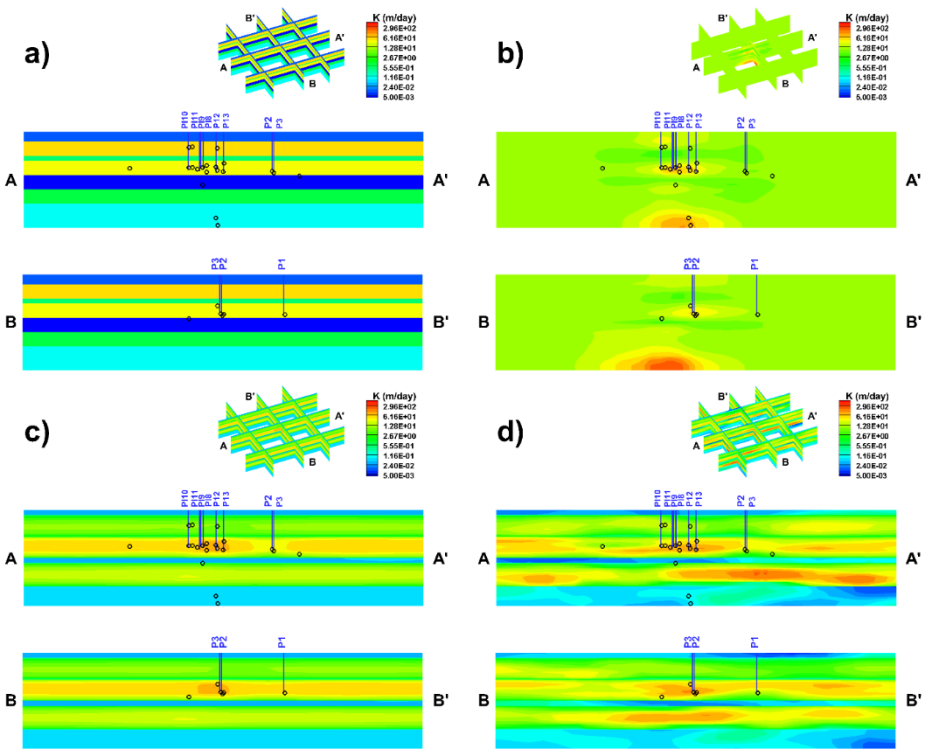

Fig. C12: Estimated K tomograms from three model cases through the interpretation of **Dataset E** as well as the "true" K field. a) Case 2, b) Case 3a, c) Case 3b, and d) "true" K field. In each contour map, small black circles represent the location of monitoring well screens.

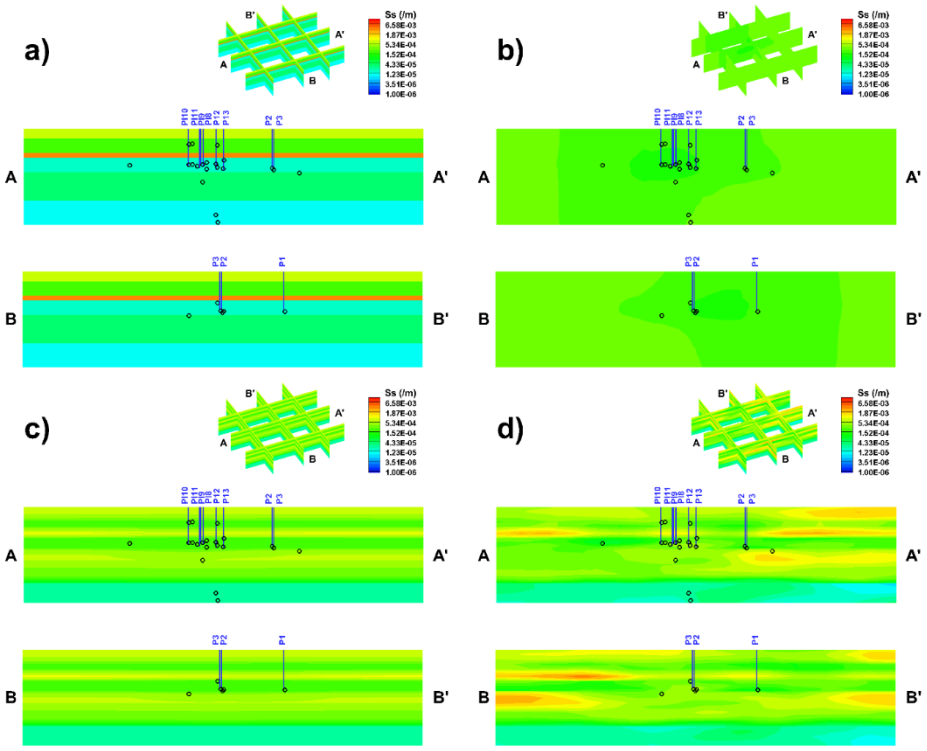

Fig. C13: Estimated S_s tomograms from three model cases through the interpretation of **Dataset E** as well as the "true" S_s field. a) Case 2, b) Case 3a, c) Case 3b, and d) "true" S_s field. In each contour map, small black circles represent the location of monitoring well screens.

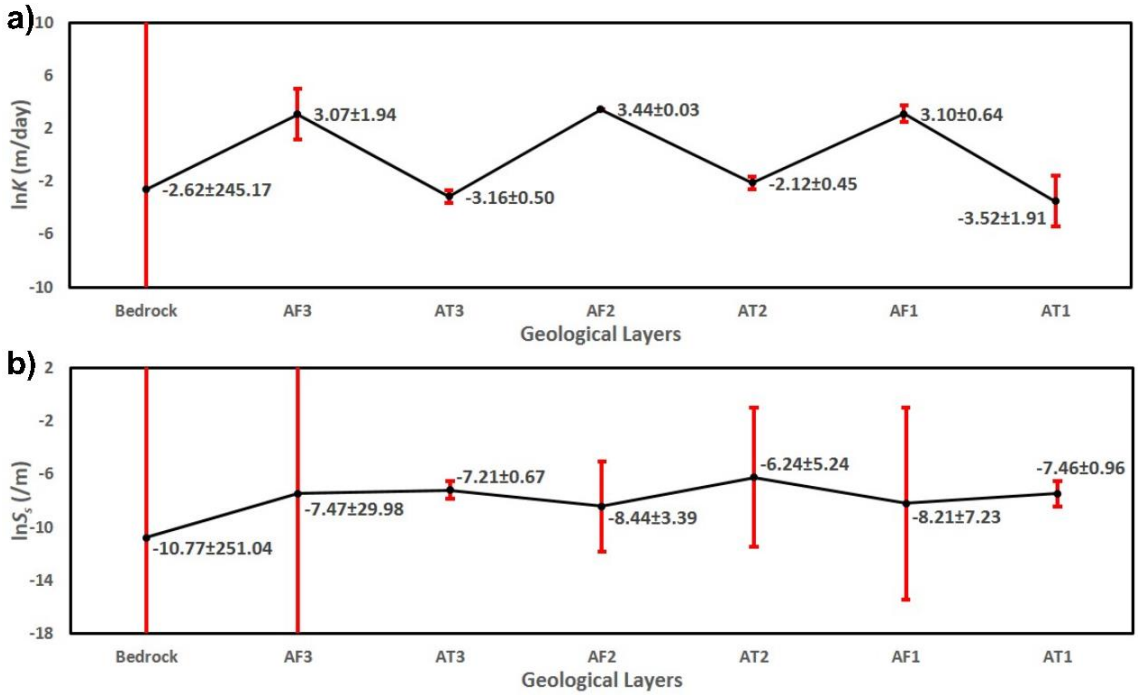

Fig. C14: Estimated ln$K$ and lnS_s values as well as their 95% confidence intervals from the geological model (Case 2) for **Dataset A**.

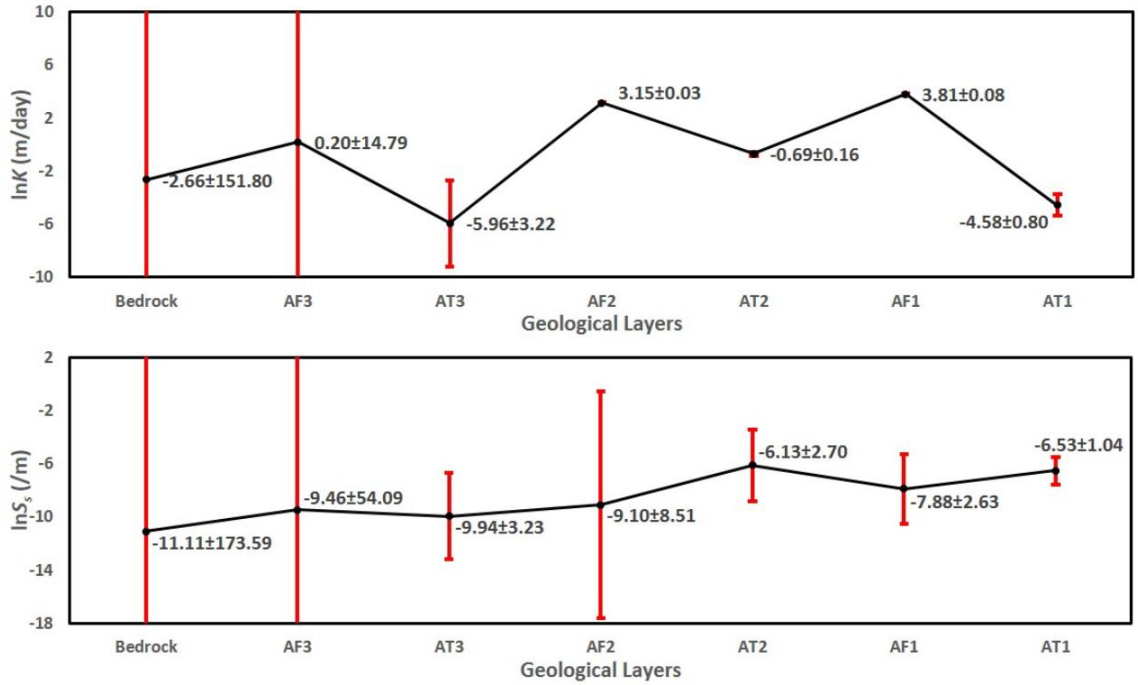

Fig. C15: Estimated ln$K$ and lnS_s values as well as their 95% confidence intervals from the geological model (Case 2) for **Dataset B**.

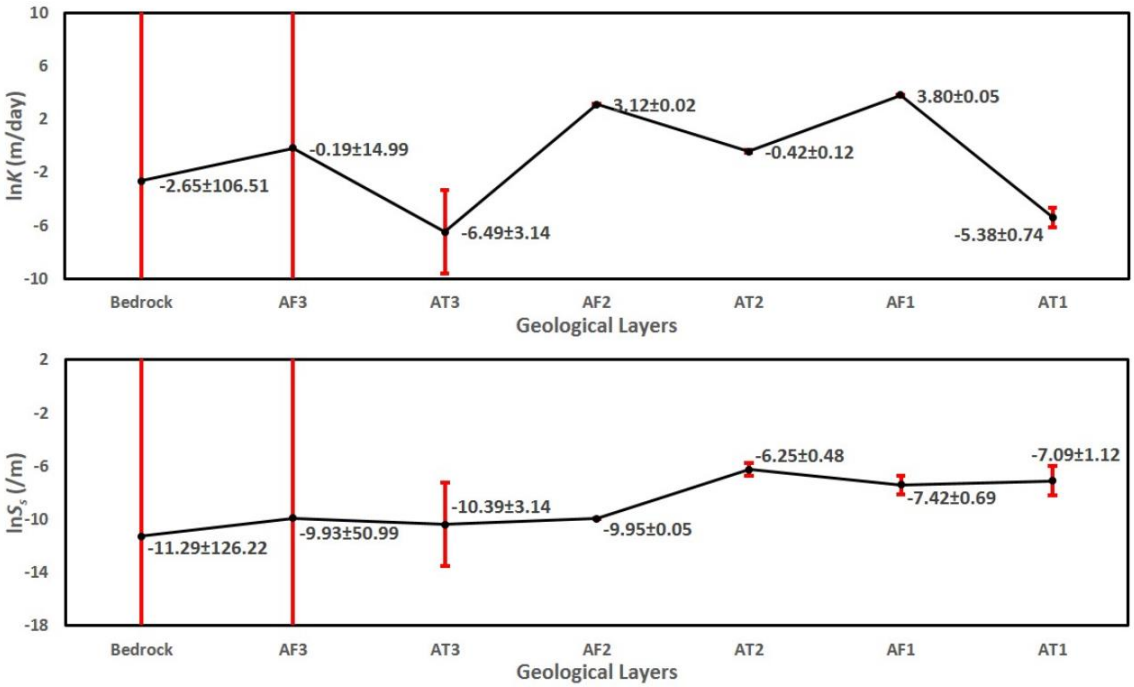

Fig. C16: Estimated ln$K$ and lnS_s values as well as their 95% confidence intervals from the geological model (Case 2) for **Dataset C**.

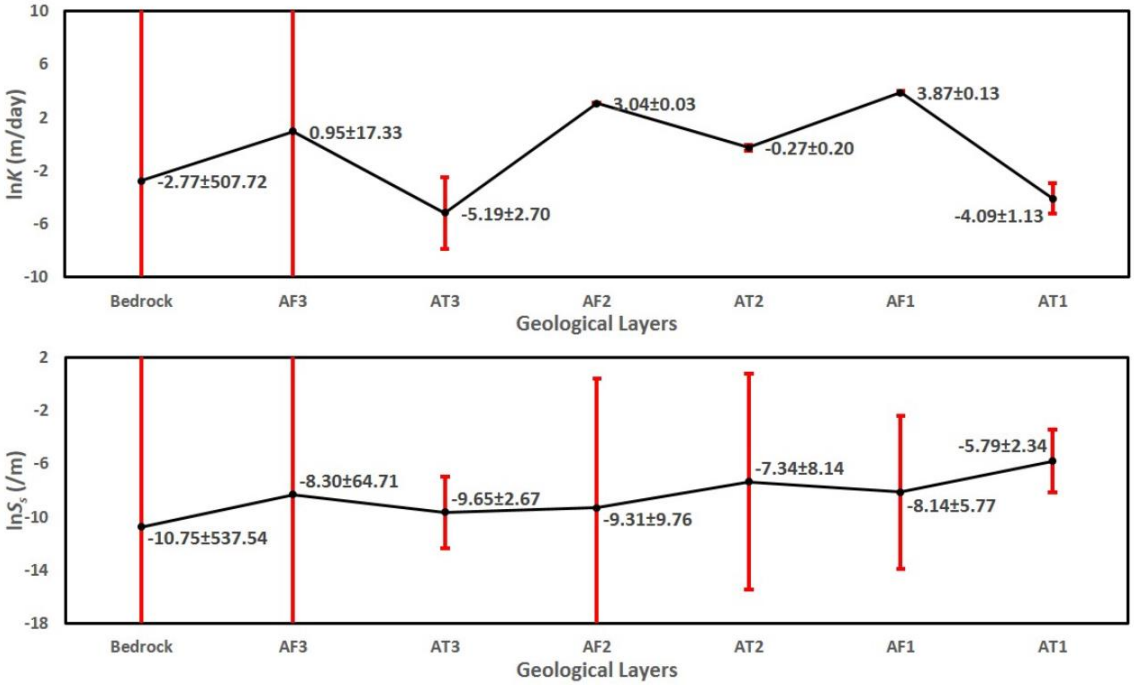

Fig. C17: Estimated ln$K$ and lnS_s values as well as their 95% confidence intervals from the geological model (Case 2) for **Dataset D**.

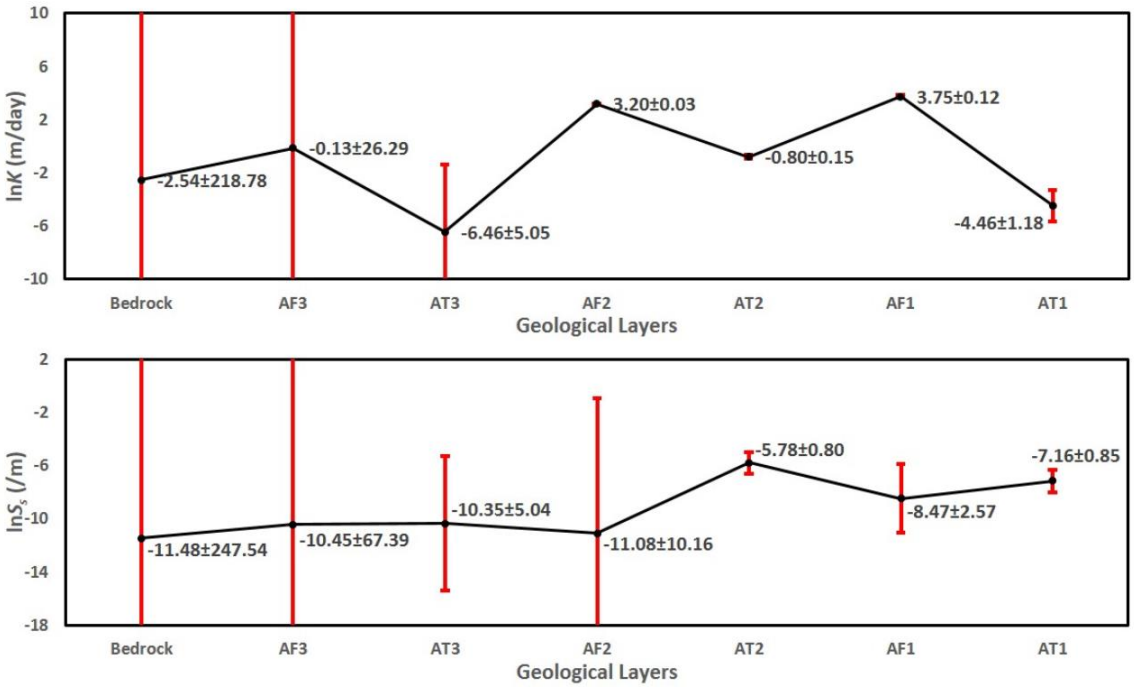

Fig. C18: Estimated ln$K$ and lnS_s values as well as their 95% confidence intervals from the geological model (Case 2) for **Dataset E**.

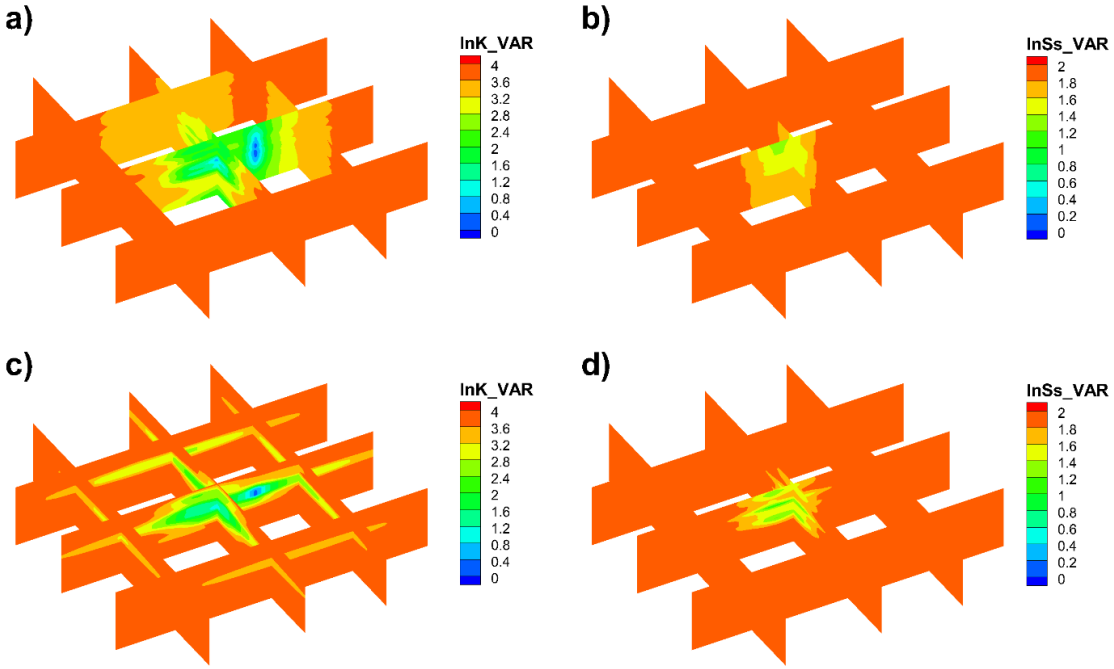

Fig. C19: Variance maps of ln$K$ and lnS_s computed from geostatistical models for **Dataset A**. a) ln$K$ variances for Case 3a, b) lnS_s variances for Case 3b, c) ln$K$ variances for Case 3b, d) lnS_s variances for Case 3b.

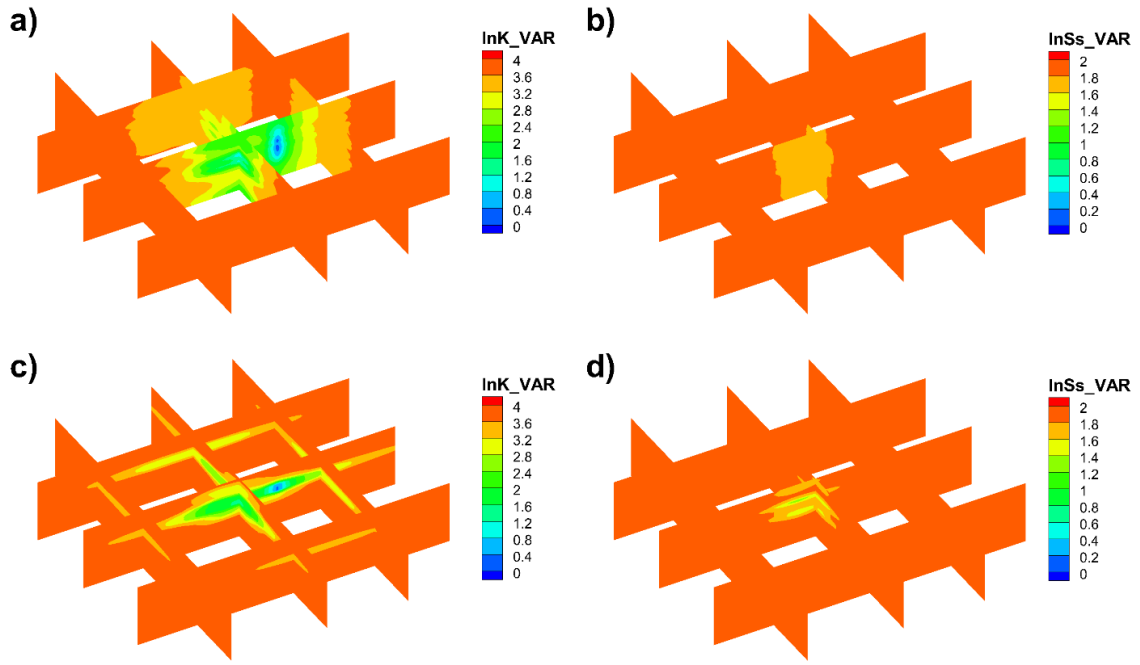

Fig. C20: Variance maps of ln$K$ and lnS_s computed from geostatistical models for **Dataset B**. a) ln$K$ variances for Case 3a, b) lnS_s variances for Case 3b, c) ln$K$ variances for Case 3b, d) lnS_s variances for Case 3b.

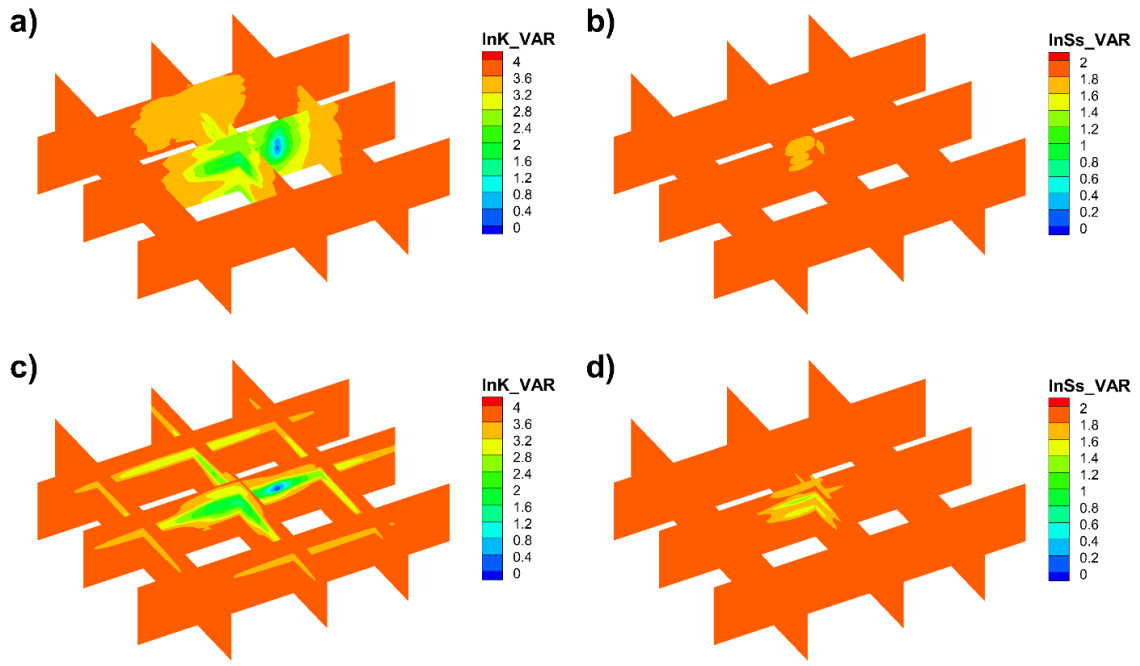

Fig. C21: Variance maps of ln$K$ and lnS_s computed from geostatistical models for **Dataset C**. a) ln$K$ variances for Case 3a, b) lnS_s variances for Case 3b, c) ln$K$ variances for Case 3b, d) lnS_s variances for Case 3b.

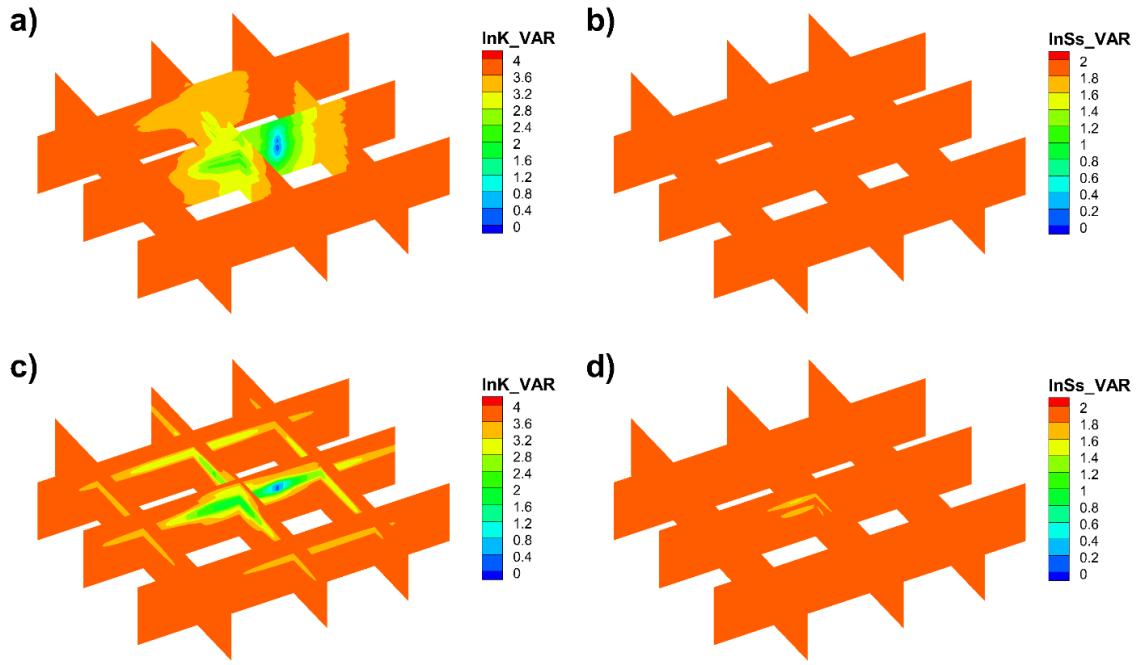

Fig. C22: Variance maps of ln$K$ and lnS_s computed from geostatistical models for **Dataset D**. a) ln$K$ variances for Case 3a, b) lnS_s variances for Case 3b, c) ln$K$ variances for Case 3b, d) lnS_s variances for Case 3b.

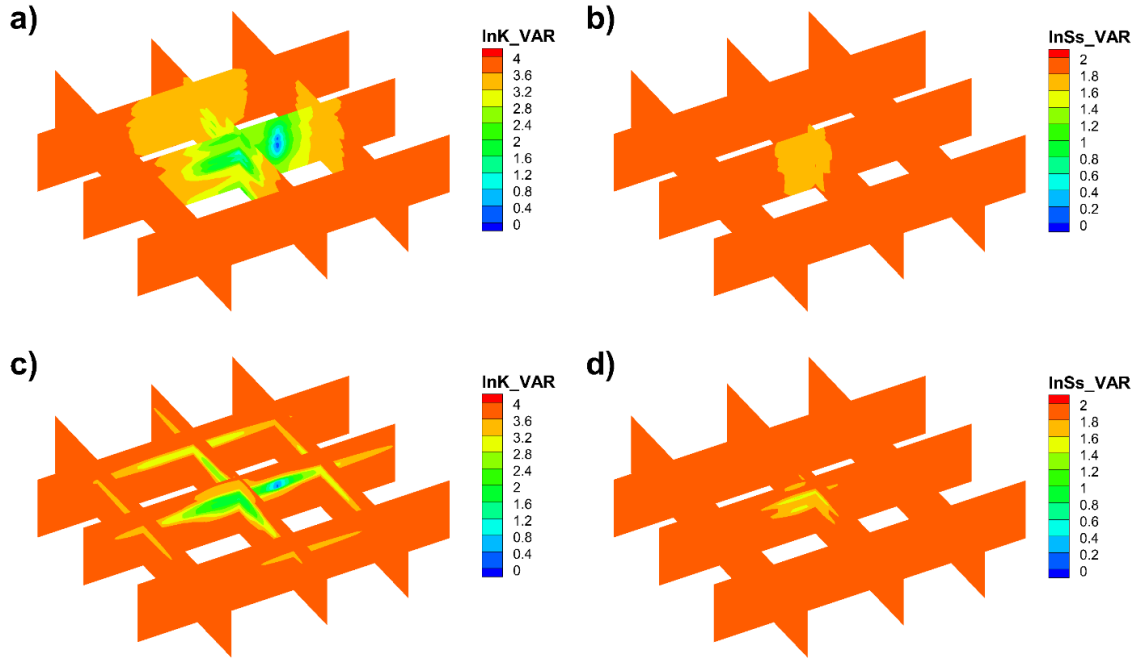

Fig. C23: Variance maps of ln$K$ and lnS_s computed from geostatistical models for **Dataset E**. a) ln$K$ variances for Case 3a, b) lnS_s variances for Case 3b, c) ln$K$ variances for Case 3b, d) lnS_s variances for Case 3b.

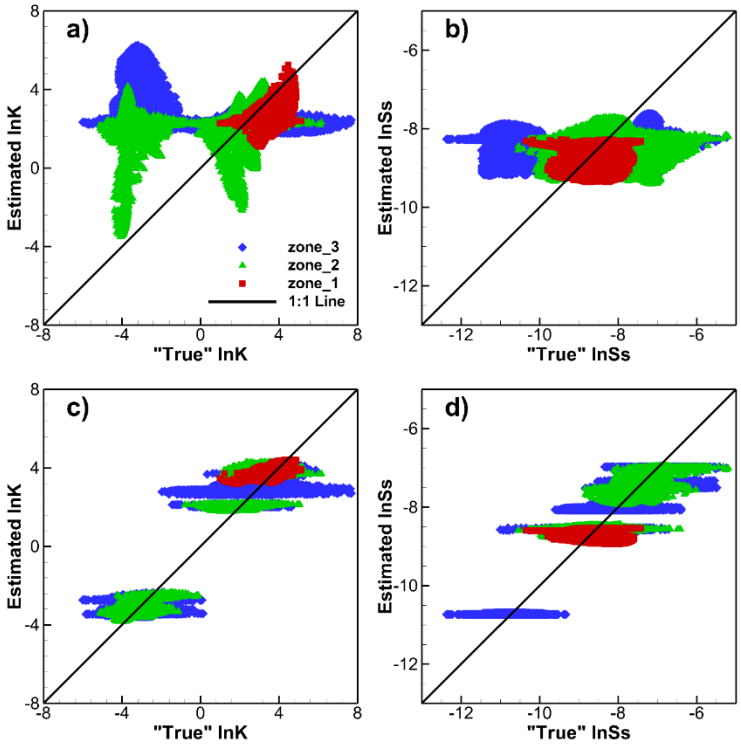

Fig. C24: Scatterplots of estimated versus "true" ln$K$ and lnS_s values at three subdivided zones for **Dataset A**. a) and b) show the comparison of ln$K$ and lnS_s values, respectively, for Case 3a, c) and d) show the comparison of ln$K$ and lnS_s values, respectively, for Case 3b.

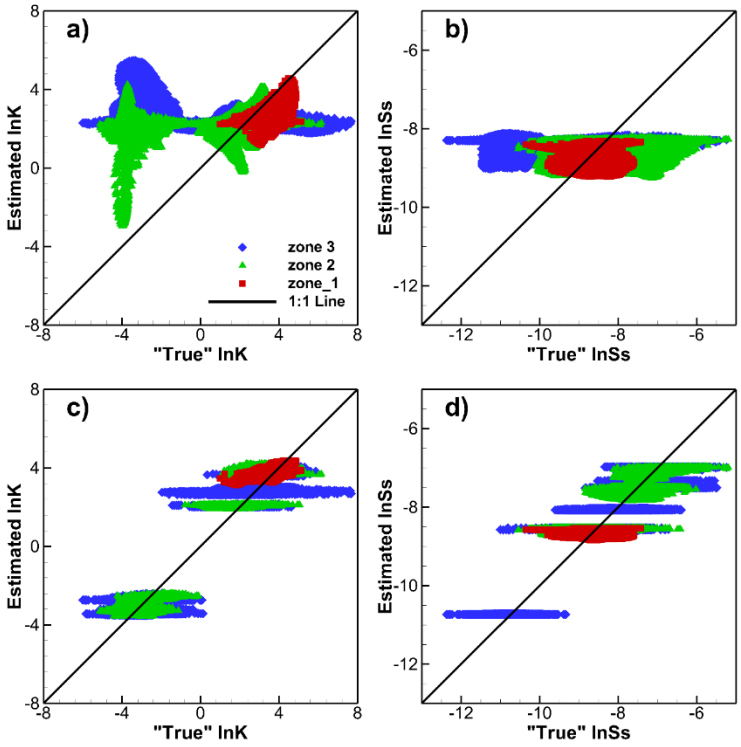

Fig. C25: Scatterplots of estimated versus "true" ln$K$ and lnS_s values at three subdivided zones for **Dataset B**. a) and b) show the comparison of ln$K$ and lnS_s values, respectively, for Case 3a, c) and d) show the comparison of ln$K$ and lnS_s values, respectively, for Case 3b.

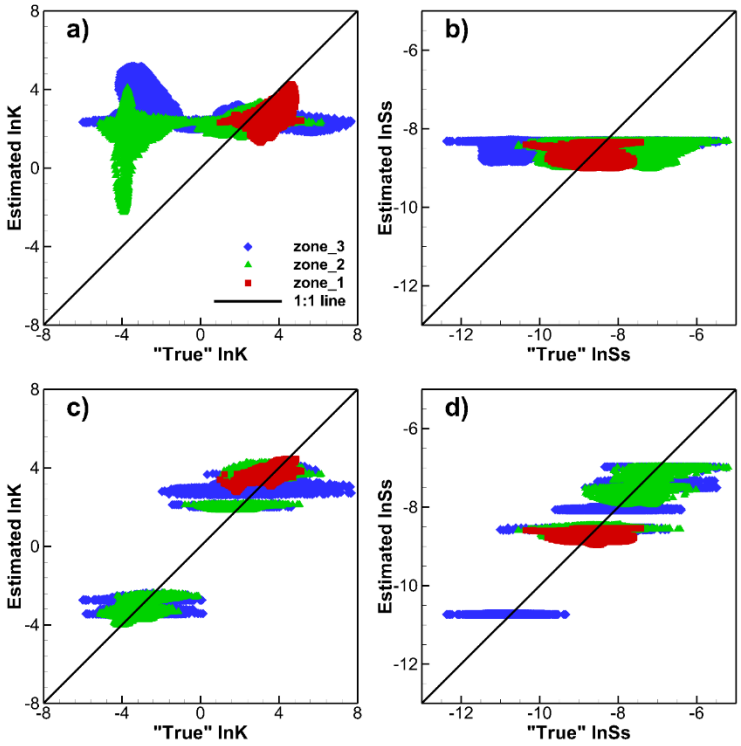

Fig. C26: Scatterplots of estimated versus "true" ln$K$ and lnS_s values at three subdivided zones for **Dataset C.** a) and b) show the comparison of ln$K$ and lnS_s values, respectively, for Case 3a, c) and d) show the comparison of ln$K$ and lnS_s values, respectively, for Case 3b.

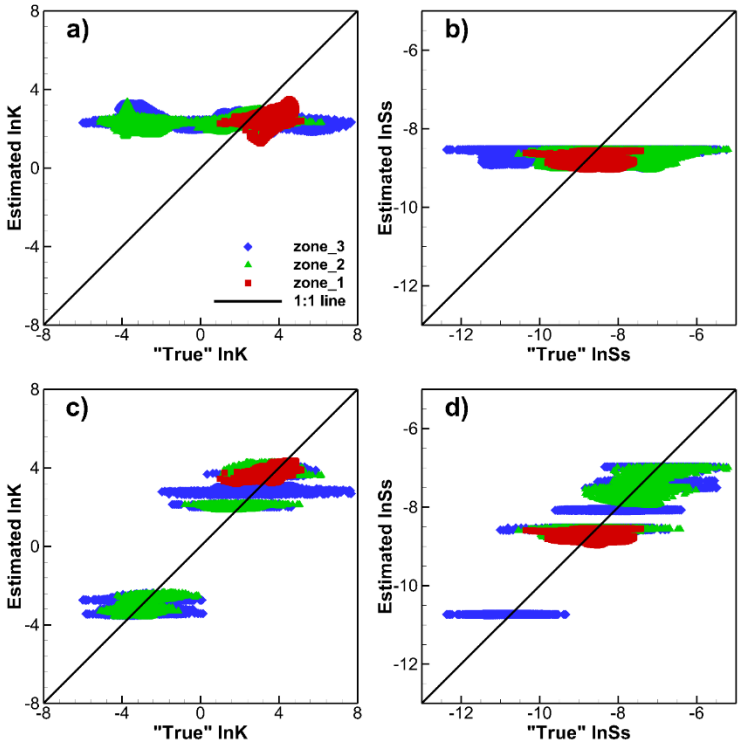

Fig. C27: Scatterplots of estimated versus "true" ln$K$ and lnS_s values at three subdivided zones for **Dataset D.** a) and b) show the comparison of ln$K$ and lnS_s values, respectively, for Case 3a, c) and d) show the comparison of ln$K$ and lnS_s values, respectively, for Case 3b.

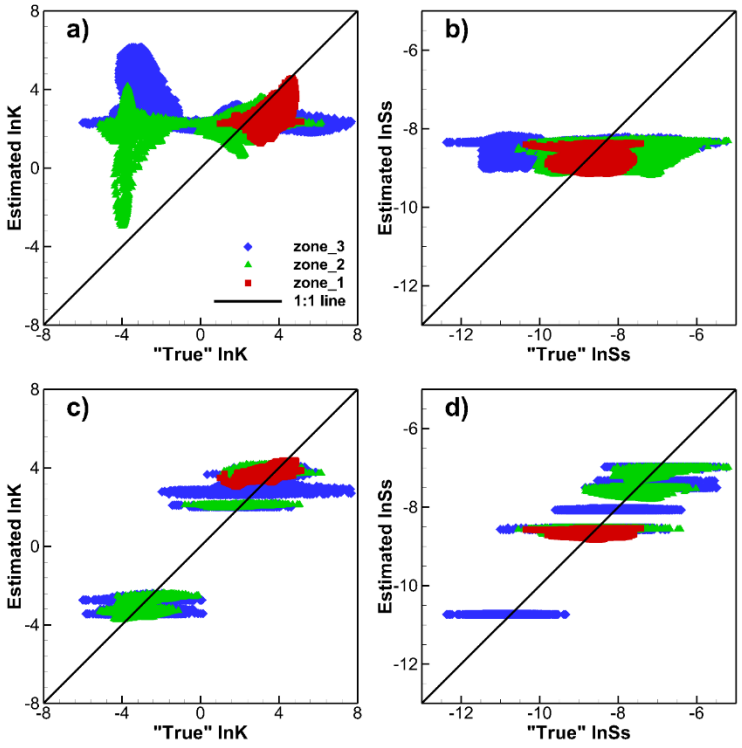

Fig. C28: Scatterplots of estimated versus "true" ln$K$ and lnS_s values at three subdivided zones for **Dataset E**. a) and b) show the comparison of ln$K$ and lnS_s values, respectively, for Case 3a, c) and d) show the comparison of ln$K$ and lnS_s values, respectively, for Case 3b.

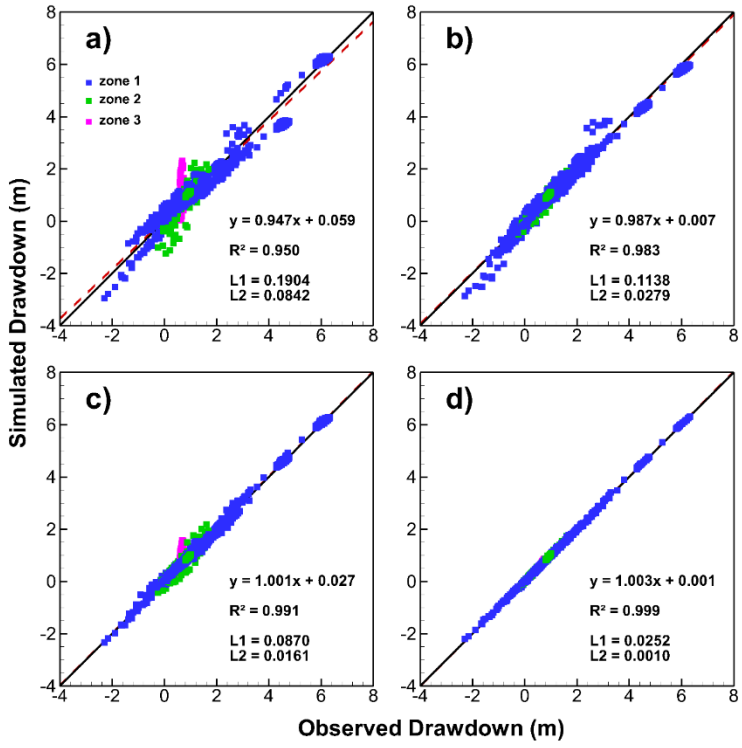

Fig. C29: Calibration scatterplots (**Dataset B**) of simulated versus "observed" drawdowns for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

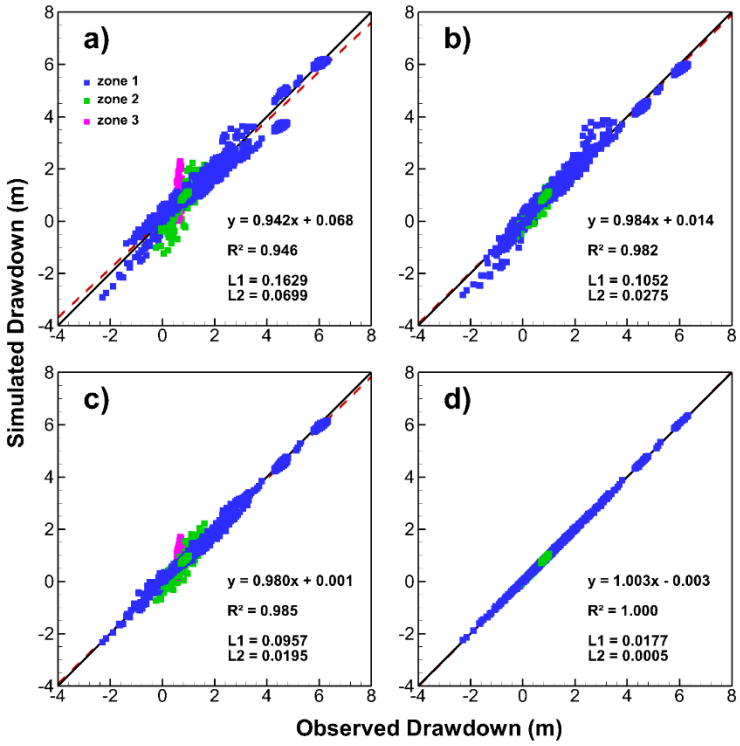

Fig. C30: Calibration scatterplots (**Dataset C**) of simulated versus "observed" drawdowns for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

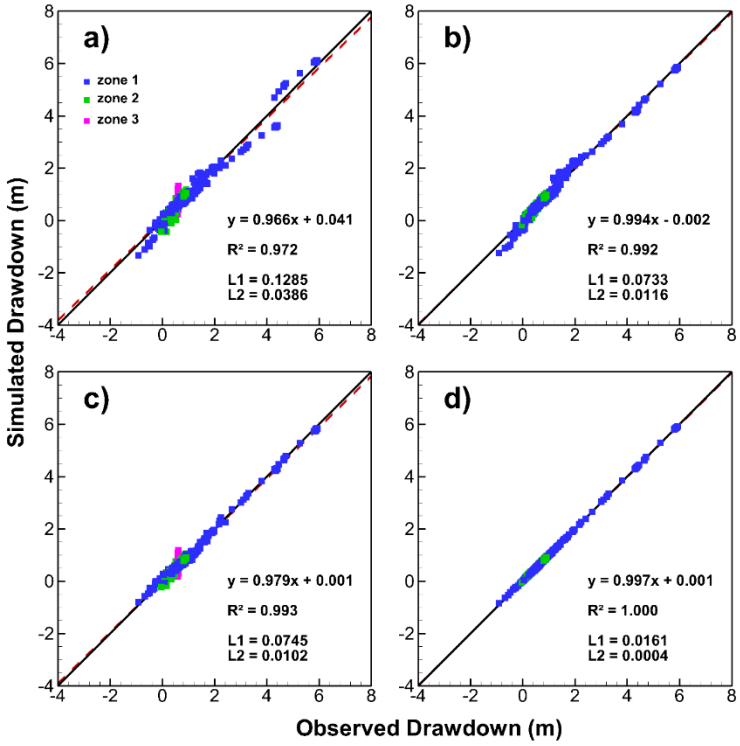

Fig. C31: Calibration scatterplots (**Dataset D**) of simulated versus "observed" drawdowns for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

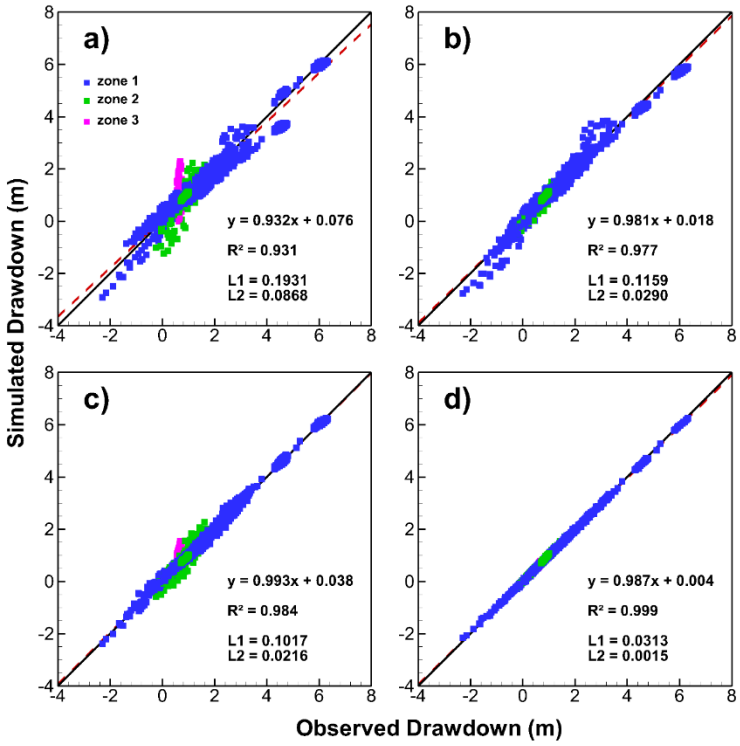

Fig. C32: Calibration scatterplots (**Dataset E**) of simulated versus "observed" drawdowns for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

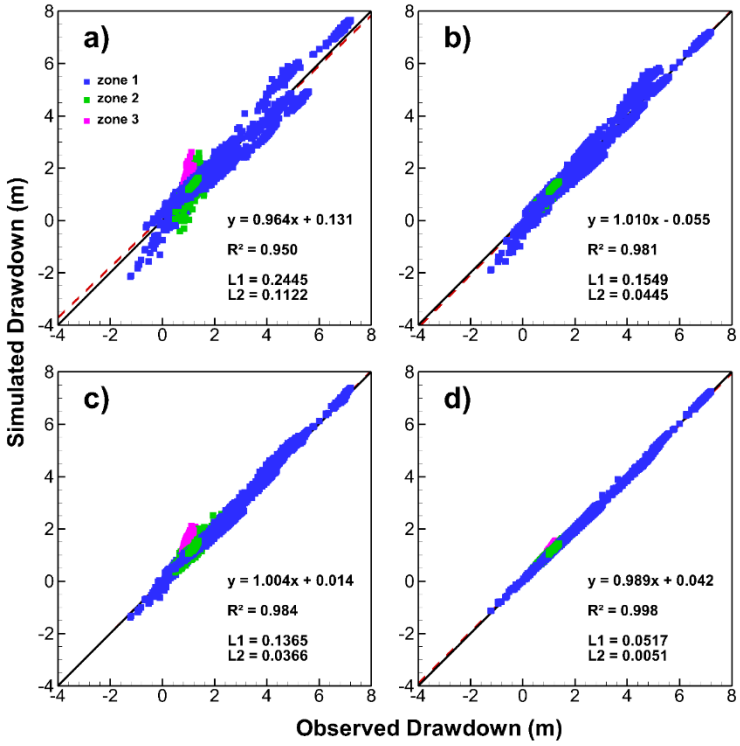

Fig. C33: Validation scatterplots (**Dataset B**) of simulated versus "observed" municipal well data (**Scenario 1**) for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

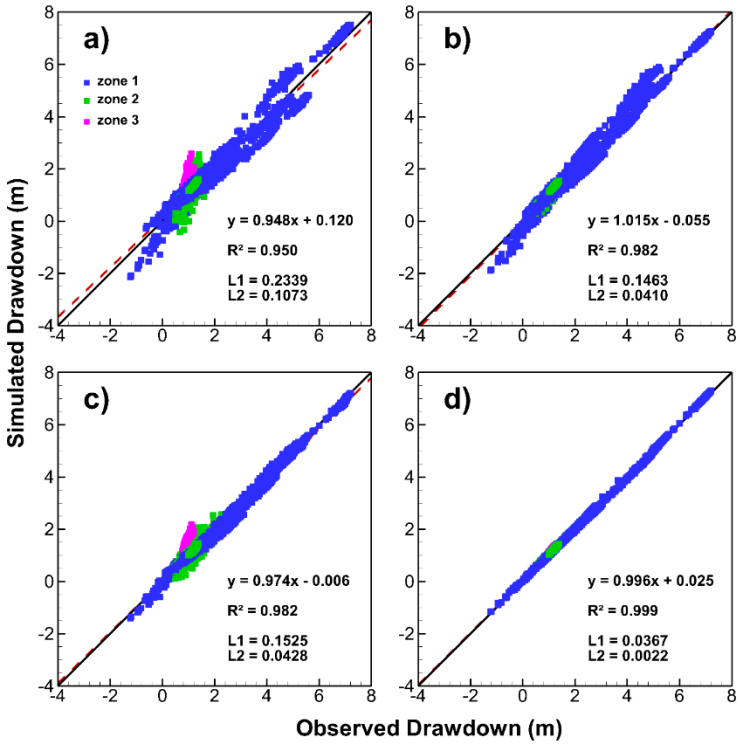

Fig. C34: Validation scatterplots (**Dataset C**) of simulated versus "observed" municipal well data (**Scenario 1**) for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

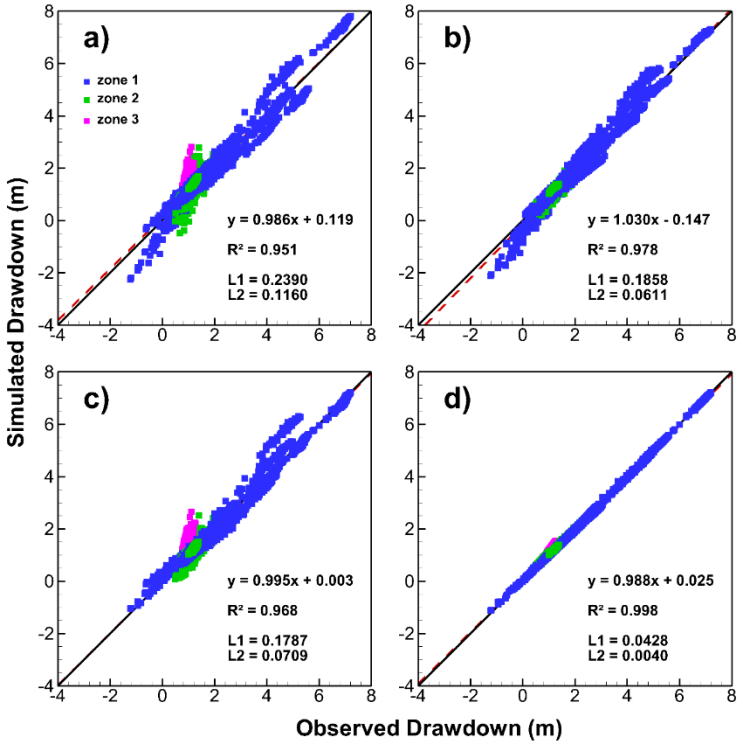

Fig. C35: Validation scatterplots (**Dataset D**) of simulated versus "observed" municipal well data (**Scenario 1**) for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

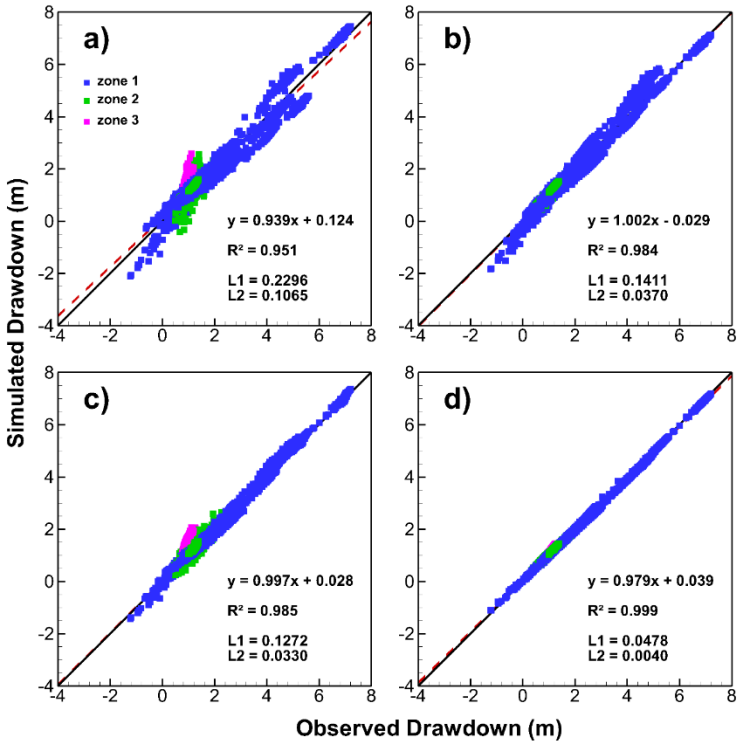

Fig. C36: Validation scatterplots (**Dataset E**) of simulated versus "observed" municipal well data (**Scenario 1**) for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

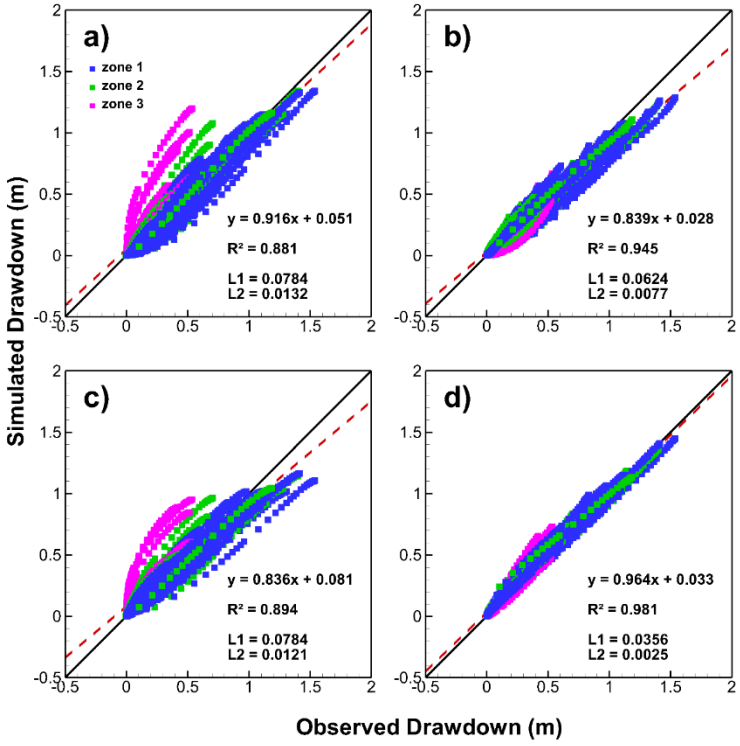

Fig. C37: Validation scatterplots (**Dataset B**) of simulated versus "observed" independent pumping test data (**Scenario 2**) for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

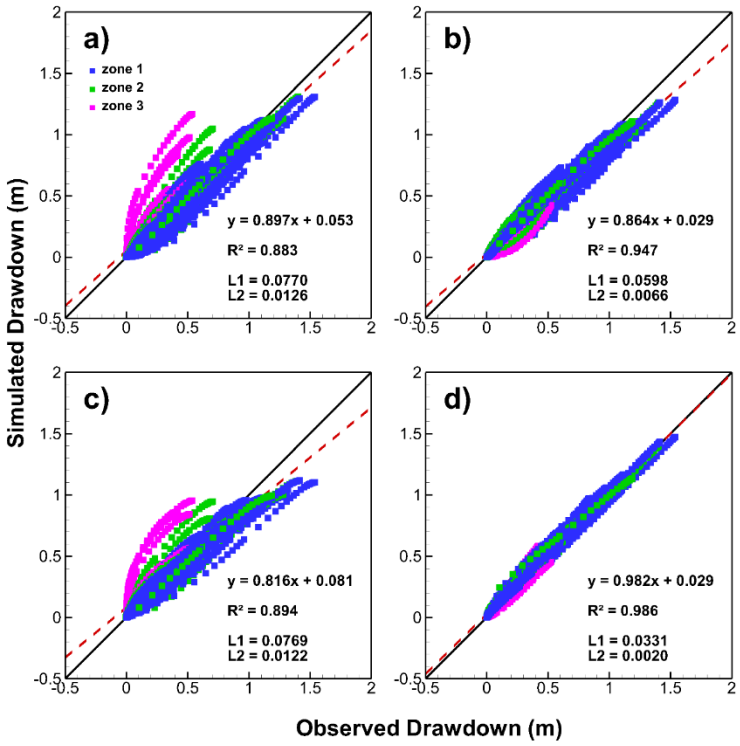

Fig. C38: Validation scatterplots (**Dataset C**) of simulated versus "observed" independent pumping test data (**Scenario 2**) for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

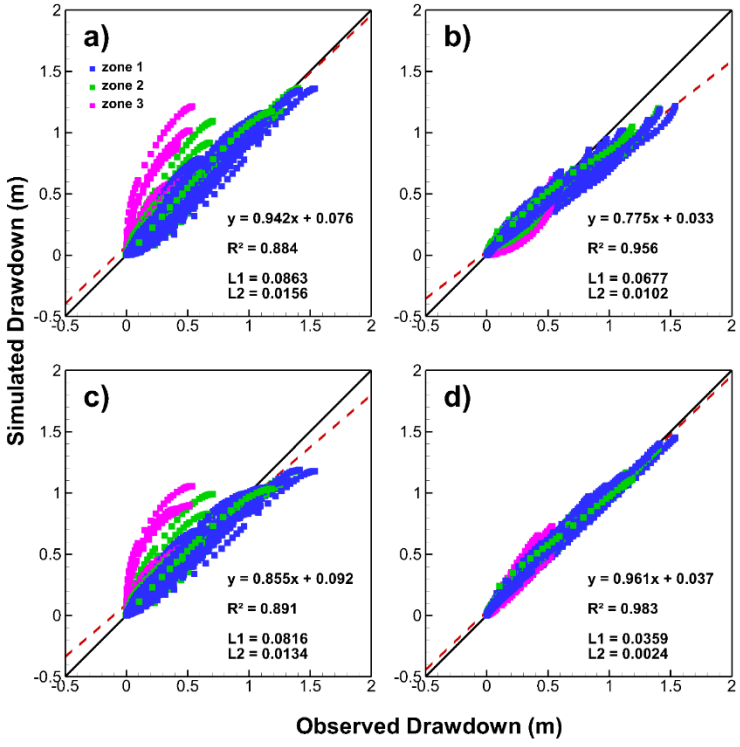

Fig. C39: Validation scatterplots (**Dataset D**) of simulated versus "observed" independent pumping test data (**Scenario 2**) for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

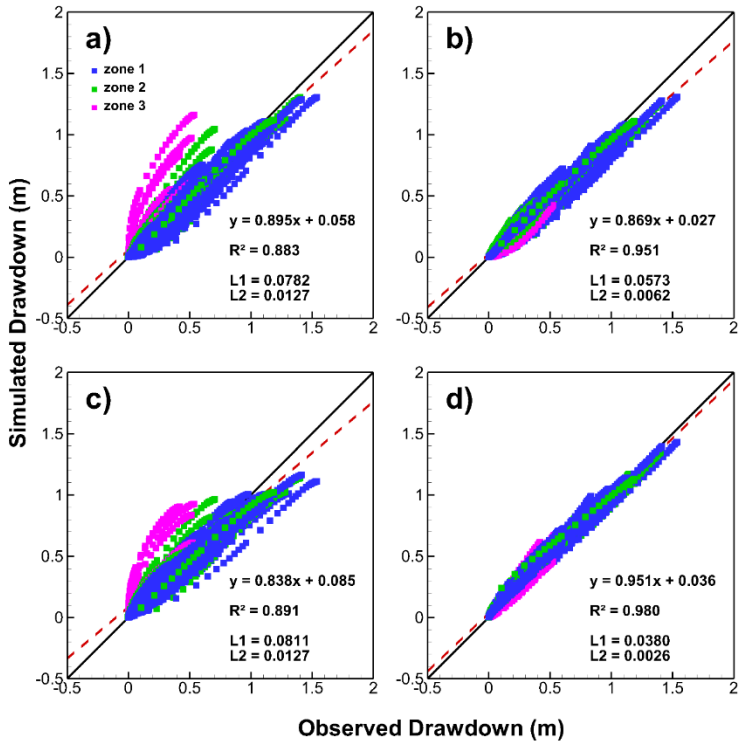

Fig. C40: Validation scatterplots (**Dataset E**) of simulated versus "observed" independent pumping test data (**Scenario 2**) for four model cases. a) Case 1, b) Case 2, c) Case 3a, d) Case 3b.

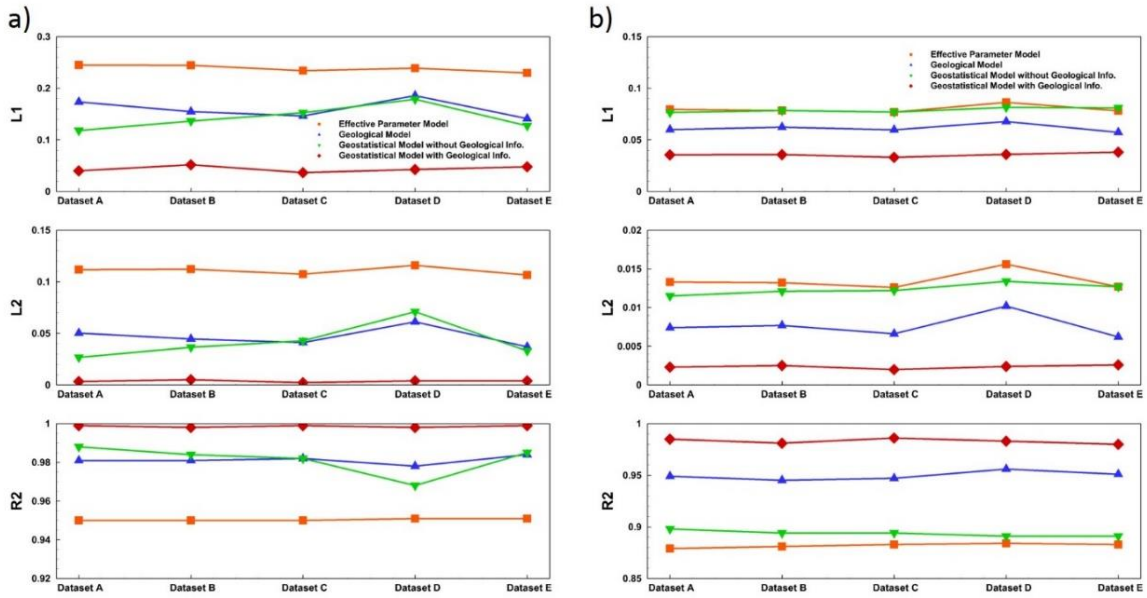

Fig. C41: Statistical Summary (L_1, L_2, and R^2) of validation results for four model cases when different datasets were incorporated for model calibration. a) municipal well data (**Scenario 1**), b) independent pumping test data (**Scenario 2**).

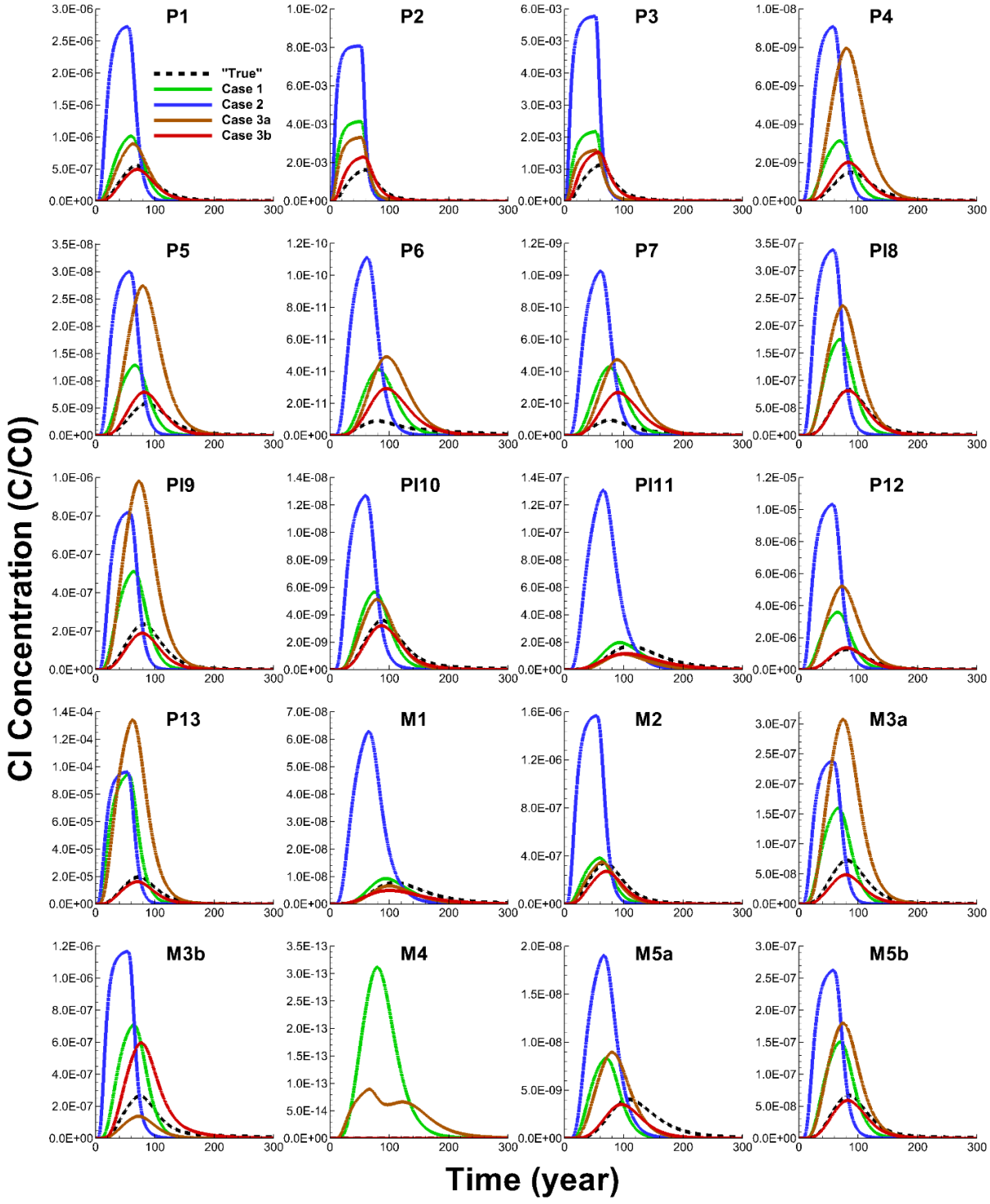

Fig. C42: Simulated and observed breakthrough curves of Cl concentration at all sampling locations for four model cases.

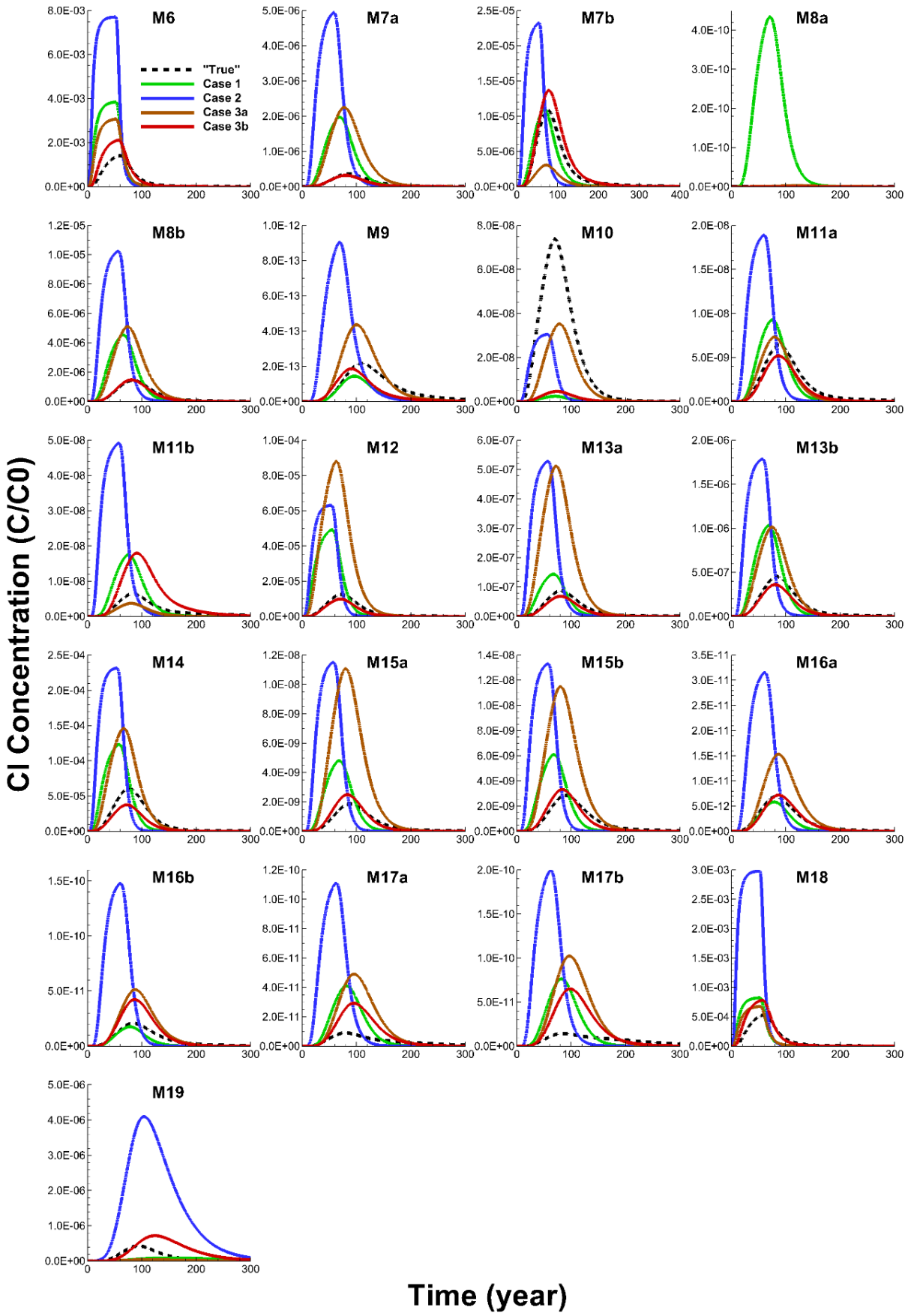

Fig. C42: **continued.**

Table C1: Statistic properties assigned for "true" K and S_s fields generation.

Identified Geological Layers		Assigned Statistical Properties for Random K and S_s Fields						
		ln$K$		lnS_s		Correlation Length (m)		
Layers	Z (masl)	Mean (m/d)	Variance	Mean (/m)	Variance	X	Y	Z
AT1	180-200	-3.3	2	-7.26	0.5	600	600	10
AF1	150-180	2	1	-8.57	0.5	600	600	10
AT2	140-150	-2.4	1	-6.81	0.5	600	600	10
AF2	110-140	3.5	1	-8.57	0.5	600	600	10
AT3	80-110	-3.2	1	-7.48	0.5	600	600	20
AF3	50-80	2.3	4.5	-8.18	0.5	600	600	10
Bedrock	0-50	-2.8	2	-10.77	0.5	600	600	20

Table C2: Summarized properties of selected datasets for model calibration.

Datasets	**Duration (days)**	**Period (day)**	**Observation Points**
A	30	1 – 30	840
B	60	1 – 60	1680
C	120	1 – 120	3360
D	30	31 – 60	840
E	120	10 – 33, 53 – 70, 80 – 85, 92 – 97, 102 - 120	2044

Table C3: Statistical summary (L_1, L_2, and R^2) of validation results for four model cases when different datasets were incorporated for model calibration. Dark green indicates the best case, followed by light green and light yellow, and dark orange shows the worst case.

Dataset A: 30 days						
	Scenario 1			**Scenario 2**		
	L1	L2	R^2	L1	L2	R^2
Case 1: Effective Parameter Model	0.2450	0.1118	0.950	0.0795	0.0133	0.879
Case 2: Geological Model	0.1737	0.0503	0.981	0.0600	0.0074	0.949
Case 3a: Geostatistical Model without Geological Info.	0.1182	0.0266	0.988	0.0764	0.0115	0.898
Case 3b: Geostatistical Model with Geological Info.	0.0398	0.0033	0.999	0.0355	0.0023	0.985
Dataset B: 60 days						
	Scenario 1			**Scenario 2**		
	L1	L2	R^2	L1	L2	R^2
Case 1: Effective Parameter Model	0.2445	0.1122	0.950	0.0784	0.0132	0.881
Case 2: Geological Model	0.1549	0.0445	0.981	0.0624	0.0077	0.945
Case 3a: Geostatistical Model without Geological Info.	0.1365	0.0366	0.984	0.0784	0.0121	0.894
Case 3b: Geostatistical Model with Geological Info.	0.0517	0.0051	0.998	0.0356	0.0025	0.981
Dataset C: 120 days						
	Scenario 1			**Scenario 2**		
	L1	L2	R^2	L1	L2	R^2
Case 1: Effective Parameter Model	0.2339	0.1073	0.950	0.0770	0.0126	0.883
Case 2: Geological Model	0.1463	0.0410	0.982	0.0598	0.0066	0.947
Case 3a: Geostatistical Model without Geological Info.	0.1525	0.0428	0.982	0.0769	0.0122	0.894
Case 3b: Geostatistical Model with Geological Info.	0.0367	0.0022	0.999	0.0331	0.0020	0.986
Dataset D: second 30 days						
	Scenario 1			**Scenario 2**		
	L1	L2	R^2	L1	L2	R^2
Case 1: Effective Parameter Model	0.2390	0.1160	0.951	0.0863	0.0156	0.884
Case 2: Geological Model	0.1858	0.0611	0.978	0.0677	0.0102	0.956
Case 3a: Geostatistical Model without Geological Info.	0.1787	0.0709	0.968	0.0816	0.0134	0.891
Case 3b: Geostatistical Model with Geological Info.	0.0428	0.0040	0.998	0.0359	0.0024	0.983
Dataset E: selected days						
	Scenario 1			**Scenario 2**		
	L1	L2	R^2	L1	L2	R^2
Case 1: Effective Parameter Model	0.2296	0.1065	0.951	0.0782	0.0127	0.883
Case 2: Geological Model	0.1411	0.0370	0.984	0.0573	0.0062	0.951
Case 3a: Geostatistical Model without Geological Info.	0.1272	0.0330	0.985	0.0811	0.0127	0.891
Case 3b: Geostatistical Model with Geological Info.	0.0478	0.0040	0.999	0.0380	0.0026	0.980

Appendix D: Supplementary Information for Study III

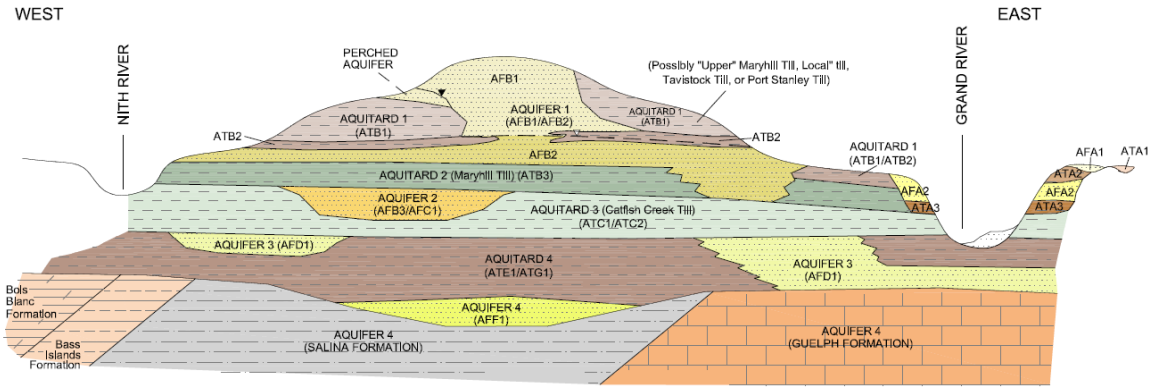

Fig. D1: Conceptual hydrogeologic model of the Waterloo Moraine, modified after the work of **Bajc and Shirota (2007)**.

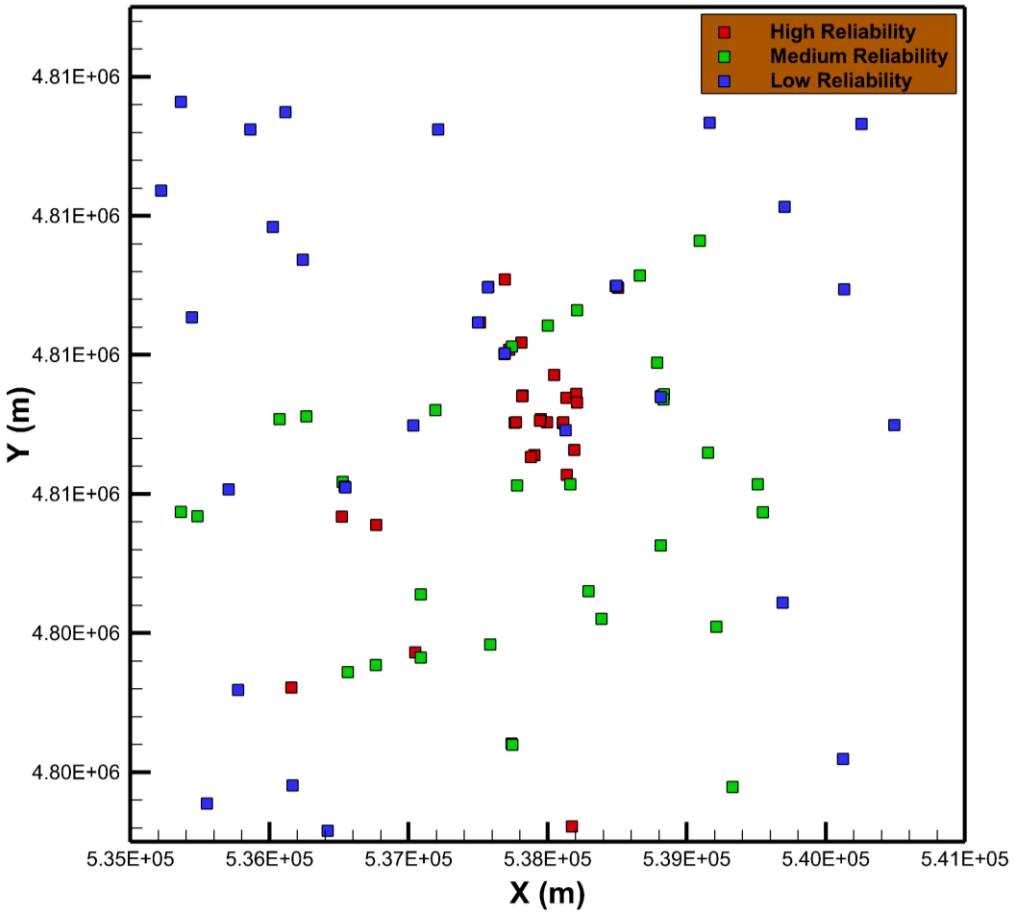

Fig. D2: Distribution of selected borehole logs for geological model construction as well as their reliabilities.

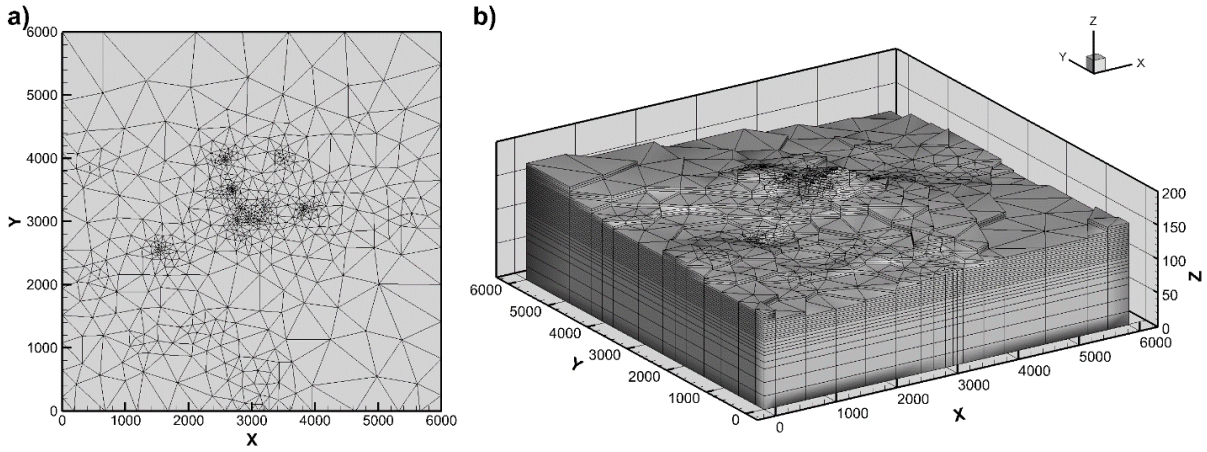

Fig. D3: Generated mesh for forward and inverse groundwater flow modeling within the simulation domain. (a) plan view and (b) perspective view.

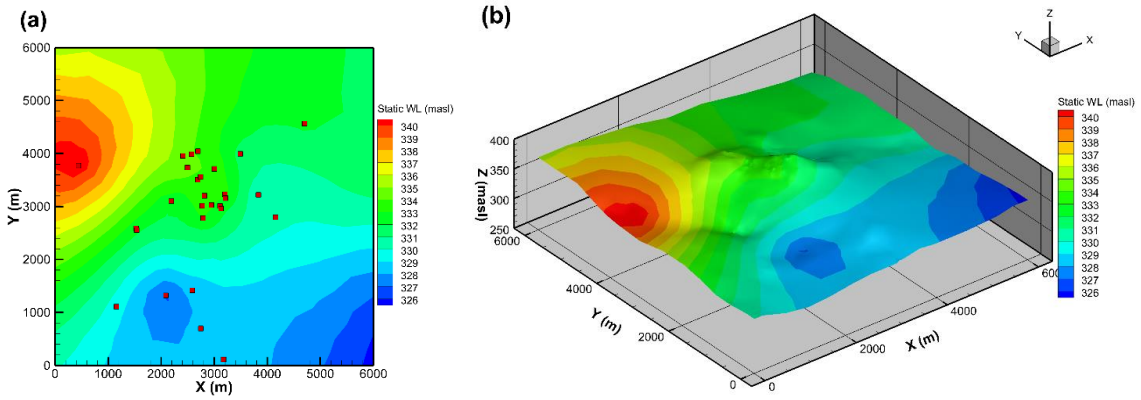

Fig. D4: Interpolated initial and lateral boundary conditions for inverse modeling using kriging. (a) 2D head map along with the distribution of static water-level measurements, and (b) Static head distribution at the surface dem of the simulation domain.

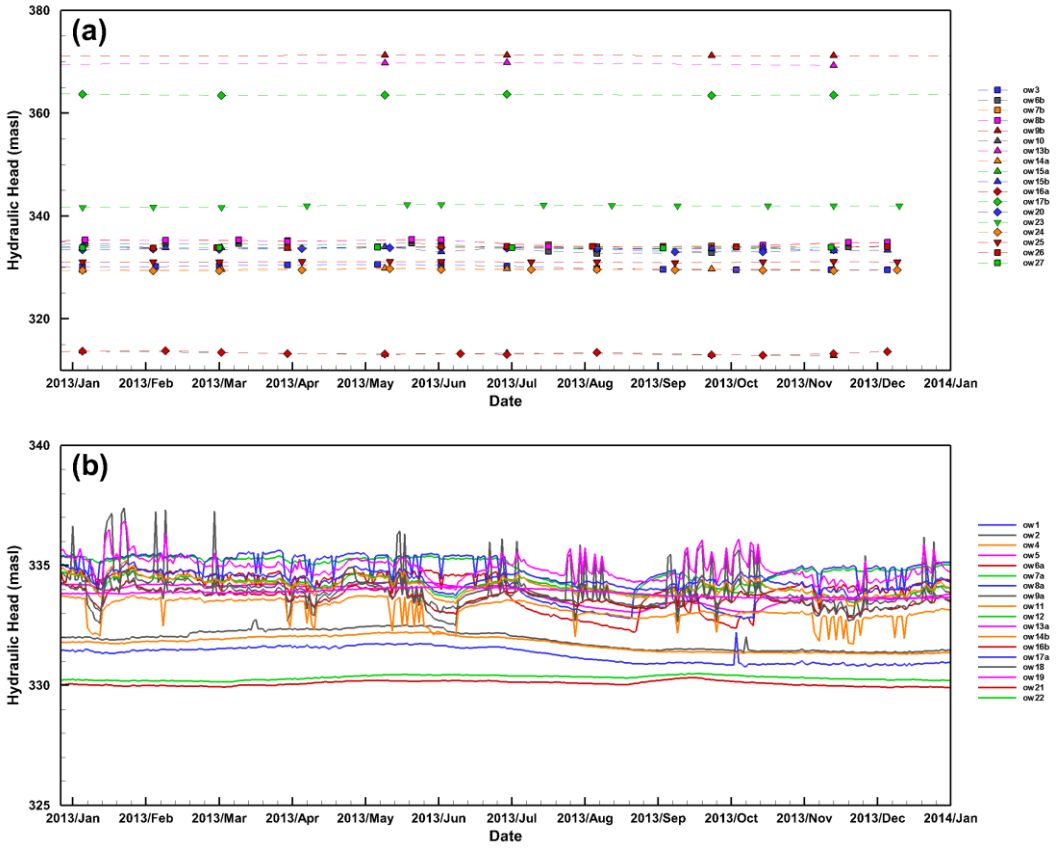

Fig. D5: Existing hydrographs in selected 36 observation wells during the year of 2013 from WRAS+ database. (a) shows the observation wells with manual measurement of head data, while (b) shows the wells with pressure transducer.

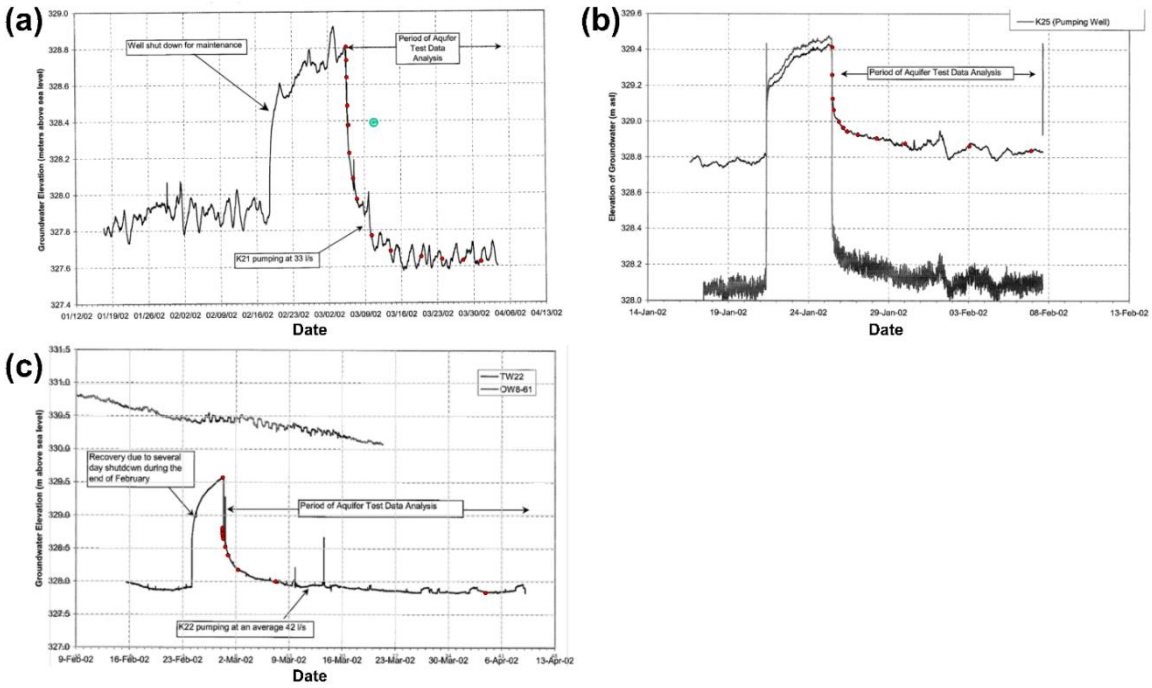

Fig. D6: Drawdown curves for the previously conducted dedicated pumping tests within the Mannheim wellfield. (a) P1, (b) P2, and (c) ow20. Extracted data points utilized for model validation are labeled as red dots in each plot.

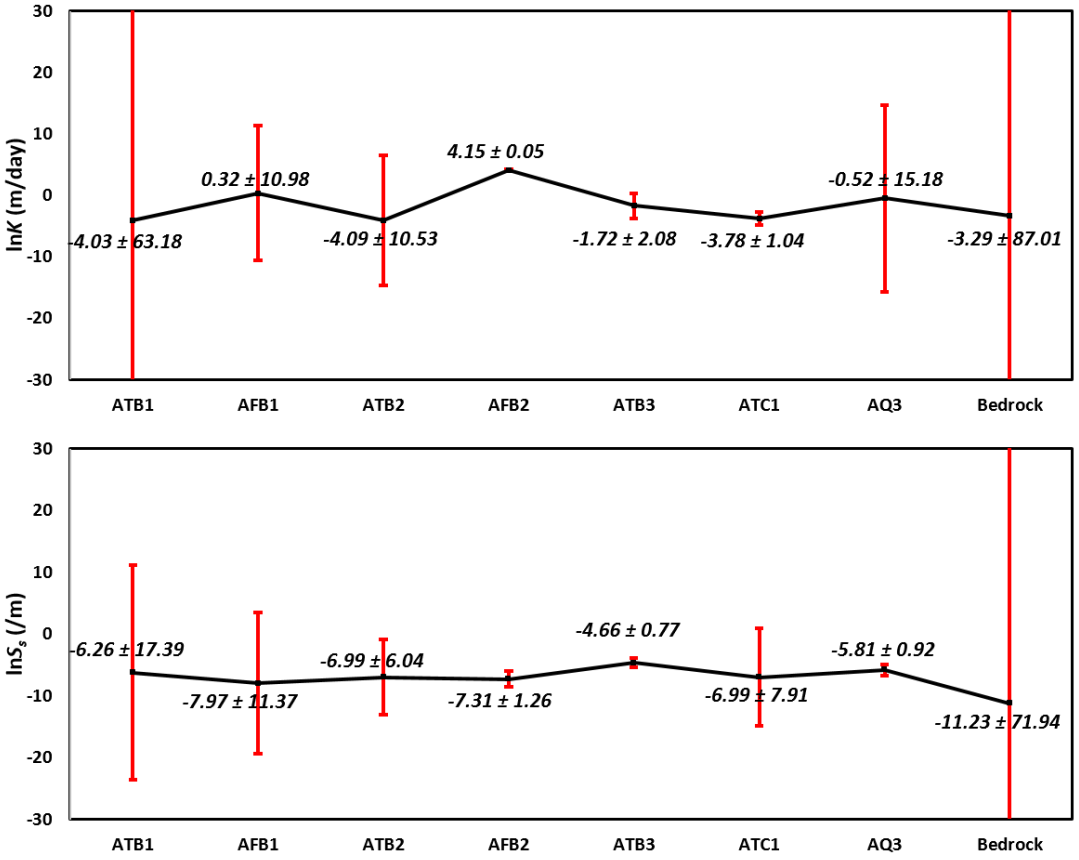

Fig. D7: Estimated ln$K$ and lnS_s values for each geological units from the geology-based zonation model as well as their 95% confidence intervals.

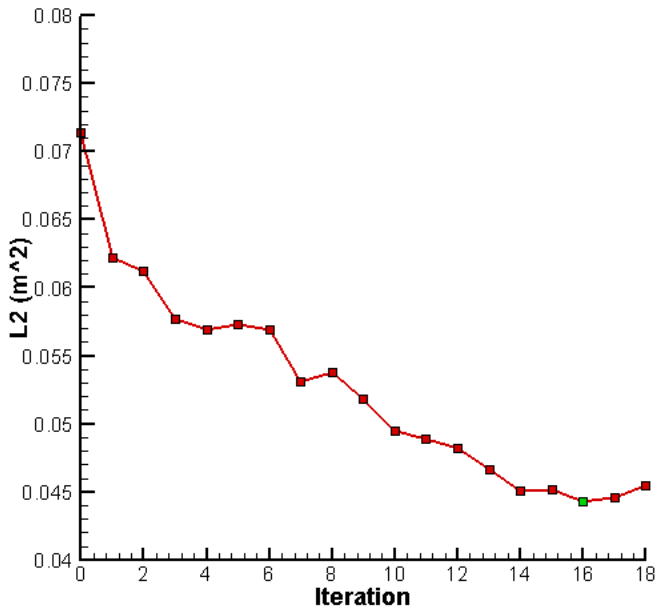

Fig. D8: Changes of the L_2 norm (m^2) between the simulated and modified water-level variations for all completed iterations during geostatistical inversions. Green dot indicates the selected iteration for model calibration results.

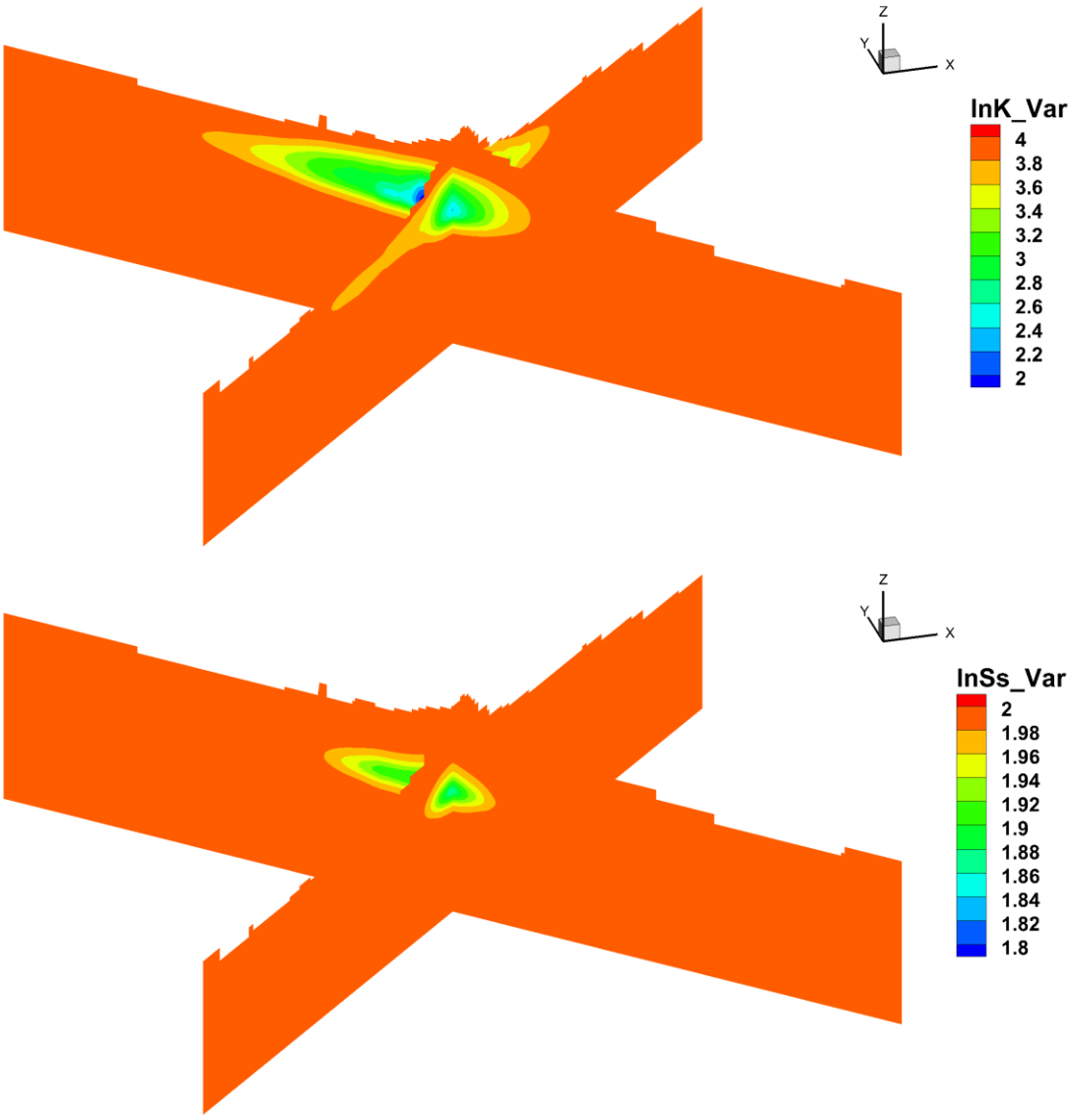

Fig. D9: Variance maps of ln$K$ and lnS_s computed from the geostatistical model.

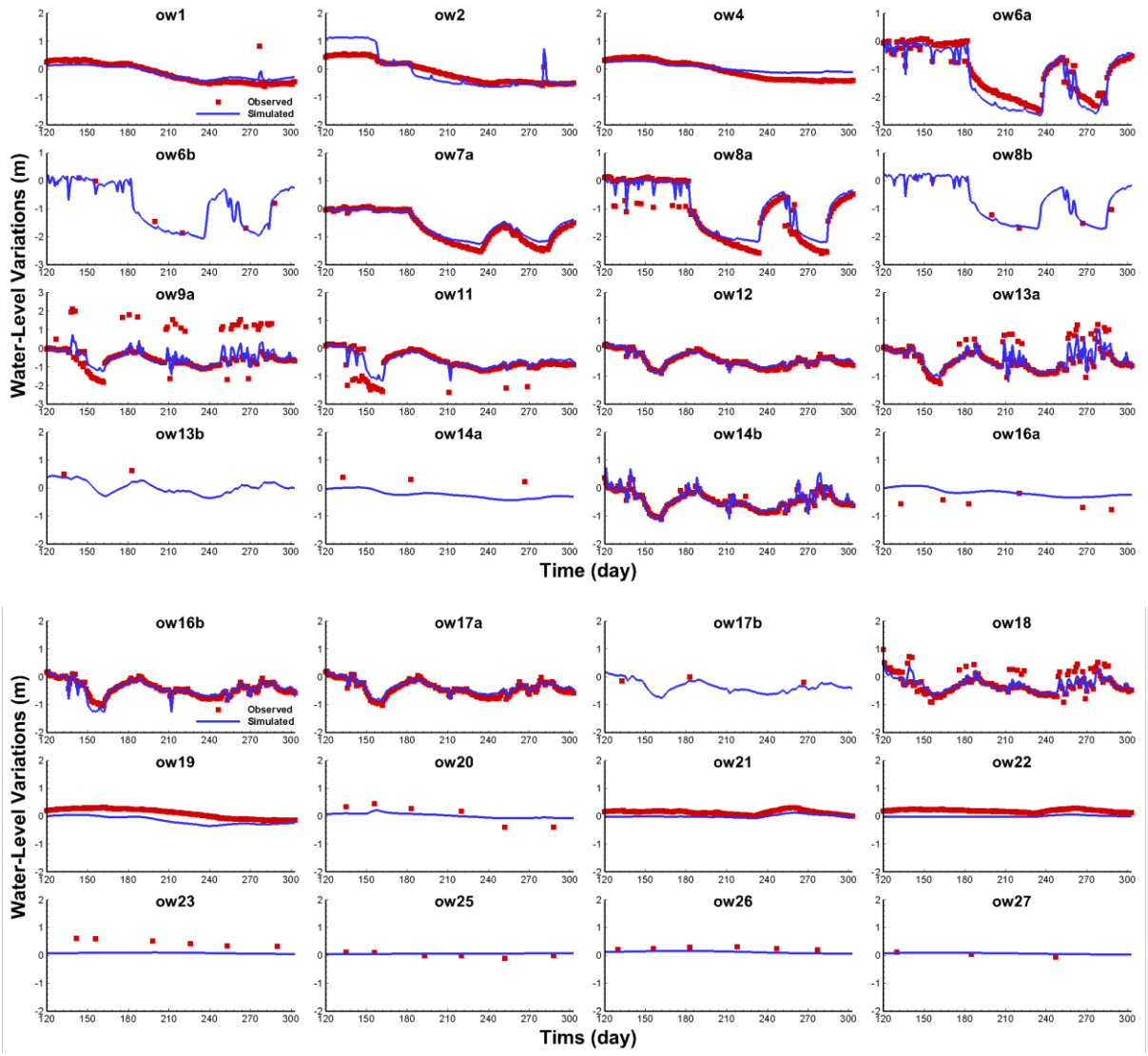

Fig. D10: Model calibration results of simulated versus observed water-level variations for the geology-based zonation model.

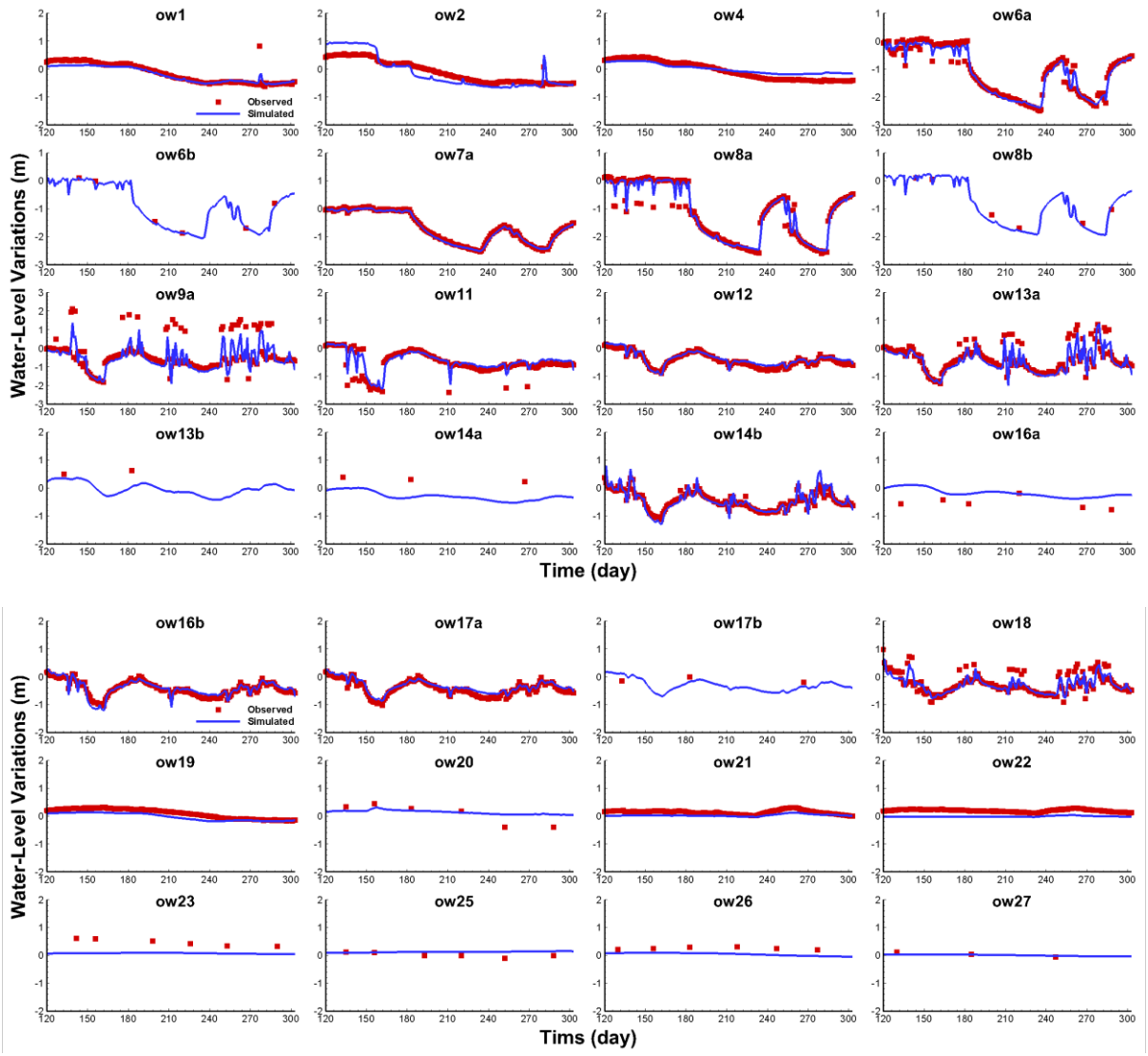

Fig. D11: Model calibration results of simulated versus observed water-level variations for the geostatistical model.

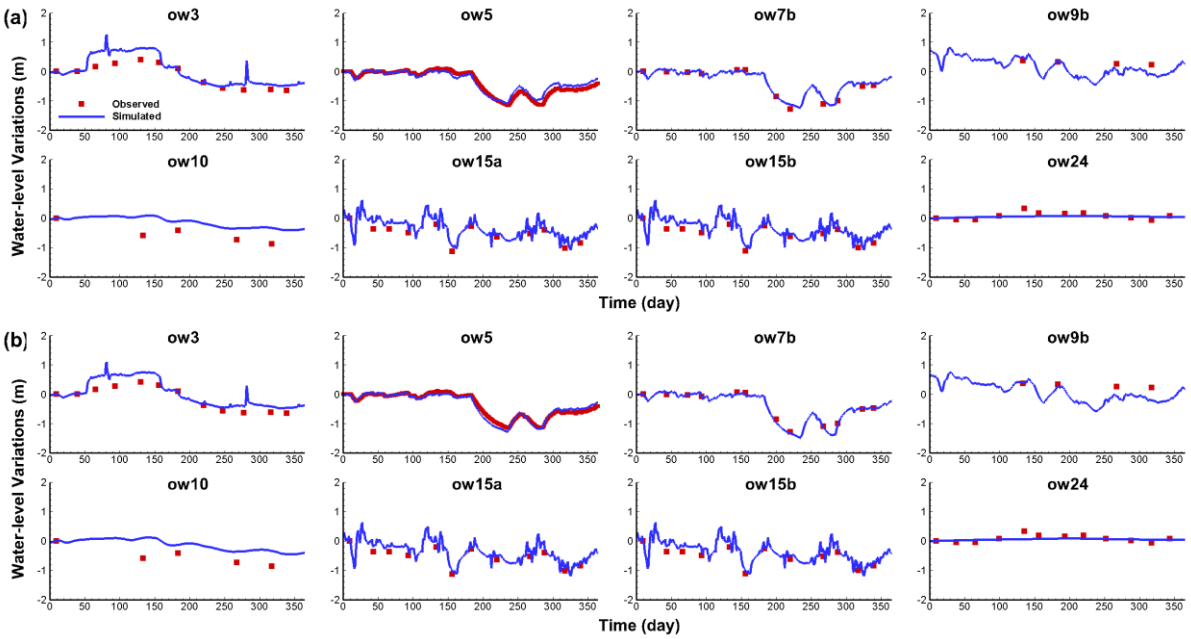

Fig. D12: Model validation results of simulated versus observed water-level variations for (a) the geology-based zonation model, and (b) the geostatistical model.

Table D1: Identified geological units for the newly constructed geological model, as well as the geological models from previous studies, along with the predominant materials in each geological units and the corresponding initial K and S_s estimates for inverse modeling.

Identified Geological Units			OGS Interpreted Units	Predominant Materials	Initial Estimates	
New Model	Tong et al., (2021)	Baic and Shirota (2007)			K (m/day)	S_s (/m)
ATB1	ATB1	ATB1	Upper Maryhill/Port Stanley Till	Silty to clayey till	1.12E-02	1.92E-03
AFB1	AFB1	AFB1	Upper Waterloo Moraine stratified sediments and equivalents	Mainly fine sand, some gravel	5.53	1.66E-04
ATB2	ATB2	ATB2	Middle Maryhill Till and equivalents	Silty to clayey till, silt, clay	1.12E-02	1.92E-03
AFB2	AFB2	AFB2	Middle Waterloo Moraine stratified sediments and equivalents	Mainly fine sand, some gravel	5.53	1.66E-04
ATB3	ATB3	ATB3	Lower Maryhill Till	Silty to clayey till	1.12E-02	1.92E-03
ATC1	ATC1	ATC1	Upper/Main Catfish Creek Till	Stony, silty to sandy till	4.32E-02	1.10E-03
		AFC1	Catfish stratified deposits	Sand and gravel		
		ATC2	Lower Catfish Creek Till	Stony, silty to sandy till		
AQ3	AFD1	AFD1	Pre-Catfish Creek sand and gravel	Sand and gravel	8.64E-01	3.32E-04
	ATE1	ATE1	Canning Drift (Pre-Catfish Creek Till)	Silty to clayey till, silt, clay		
	AFF1	AFF1	Pre-Canning coarse-textured glaciofluvial/glaciolacustrine sediments	Sand and gravel		
	ATG1	ATG1	Pre-Canning coarse-textured till	Stony, silty to sandy till		
Bedrock	Bedrock	Bedrock	Guelph, Salina, Bass islands, Bois Blanc Formation	Limestone, Dolostone, Shale	4.32E-02	1.51E-05

Table D2: Summary of dedicated pumping tests utilized for model validation.

Tests	**Pumping Well**	**Observation Well**	**Distance (m)**	**Pumping Rate (m^3/min)**	**Duration (day)**	**Extracted Points**
#1	P1	ow-p1	19	1.98	~ 26	14
#2	P2	ow3	21	2.76	~ 13	12
#3	ow20	ow-ow20	8	2.52	~ 35	15

Table D3: Transmissivity (T) estimates obtained from previous aquifer tests conducted in the Mannheim wellfield using analytical solutions, as well as computed hydraulic conductivity (K) based on the aquifer thickness obtained from the newly constructed geological model in this study.

Well Site	Well	Aquifer thickness (m)	Transmissivity (m^2/day)	Hydraulic Conductivity (m/day)
East	P1	19.2	1435	74.74
	P2	26.2	5550	211.83
Peaking	P4	46.6	2277	48.86
	P5	41.2	1318	31.99
	P6	39.7	1976	49.77
	P7	42.1	1581	37.55
ASR	P8	44.8	5292	118.13
	P9	37.2	1370	36.83
	P10	32.3	2118	65.57
	P11	47.6	2830	59.45
	P12	38.4	3023	78.72
	P13	45.1	2885	63.97
West	ow20	26.5	1770	66.79
	P16	31.4	13600	433.12

Table D4: Statistical summary of the validation results using existing municipal well records.

Well	Geology-based Zonation Model					Geostatistical Model				
	L_1	L_2	Linear Model			L_1	L_2	Linear Model		
			Slope	*Intercept*	R^2			*Slope*	*Intercept*	R^2
ow3[a]	0.196	0.071	1.260	0.246	0.921	0.189	0.058	1.153	0.228	0.930
ow5[b]	0.098	0.015	0.816	-0.035	0.929	0.072	0.008	1.006	-0.007	0.954
ow7b[b]	0.077	0.011	0.844	-0.022	0.977	0.041	0.003	1.029	-0.012	0.990
ow15a[b]	0.173	0.037	0.903	0.133	0.948	0.154	0.028	1.005	0.171	0.977
ow15b[b]	0.167	0.035	0.903	0.127	0.945	0.148	0.026	1.006	0.164	0.975
ow9b[c]	0.184	0.043	4.369	-1.090	0.986	0.213	0.075	4.470	-1.226	0.995
ow10[c]	0.380	0.196	0.694	0.276	0.457	0.375	0.196	0.676	0.258	0.376
ow24[d]	0.077	0.011	0.103	0.050	0.429	0.076	0.010	0.112	0.048	0.478

[a] Observation well located close to the production well.

[b] Observation wells located within the production area and screened in AFB2.

[c] Observation wells located within the production area, but screened in the upper and lower units.

[d] Observation well located away from the production area, but screened in AFB2.

Appendix E: Supplementary Information for Study IV

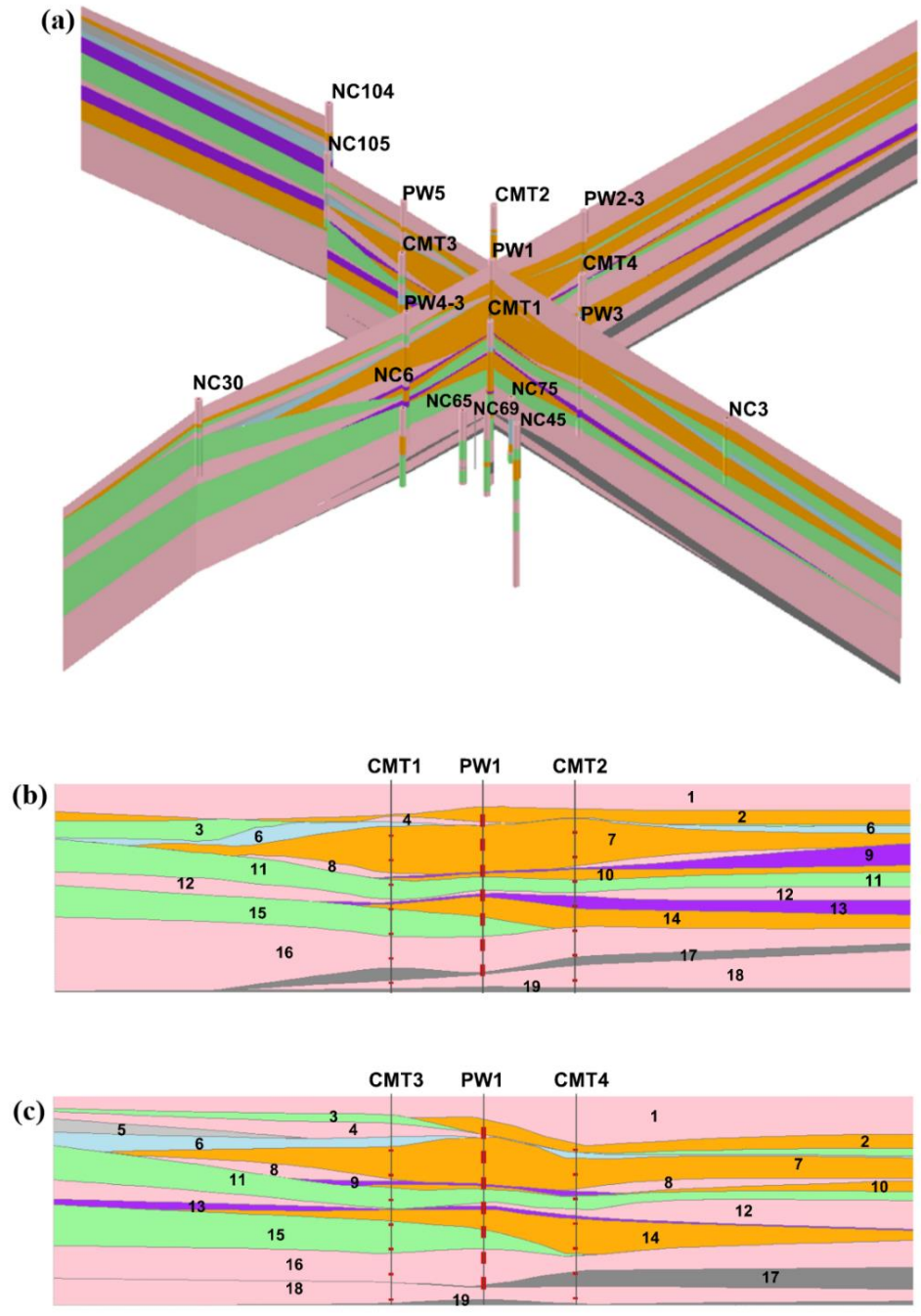

Fig. E1: 19-layer geological model constructed for the NCRS showing locations of boreholes used to collect stratigraphy data. Numbers in (b) and (c) indicate layers of different materials: Clay (1, 4, 8, 12, 16, 18); Silt and Clay (17, 19); Silt (2, 7, 10, 14); Sandy Silt (6, 9, 13); Sand and Silt (5); Sand (3, 11); Sand and Gravel (15). Modified after the work of Zhao and Illman (2017).

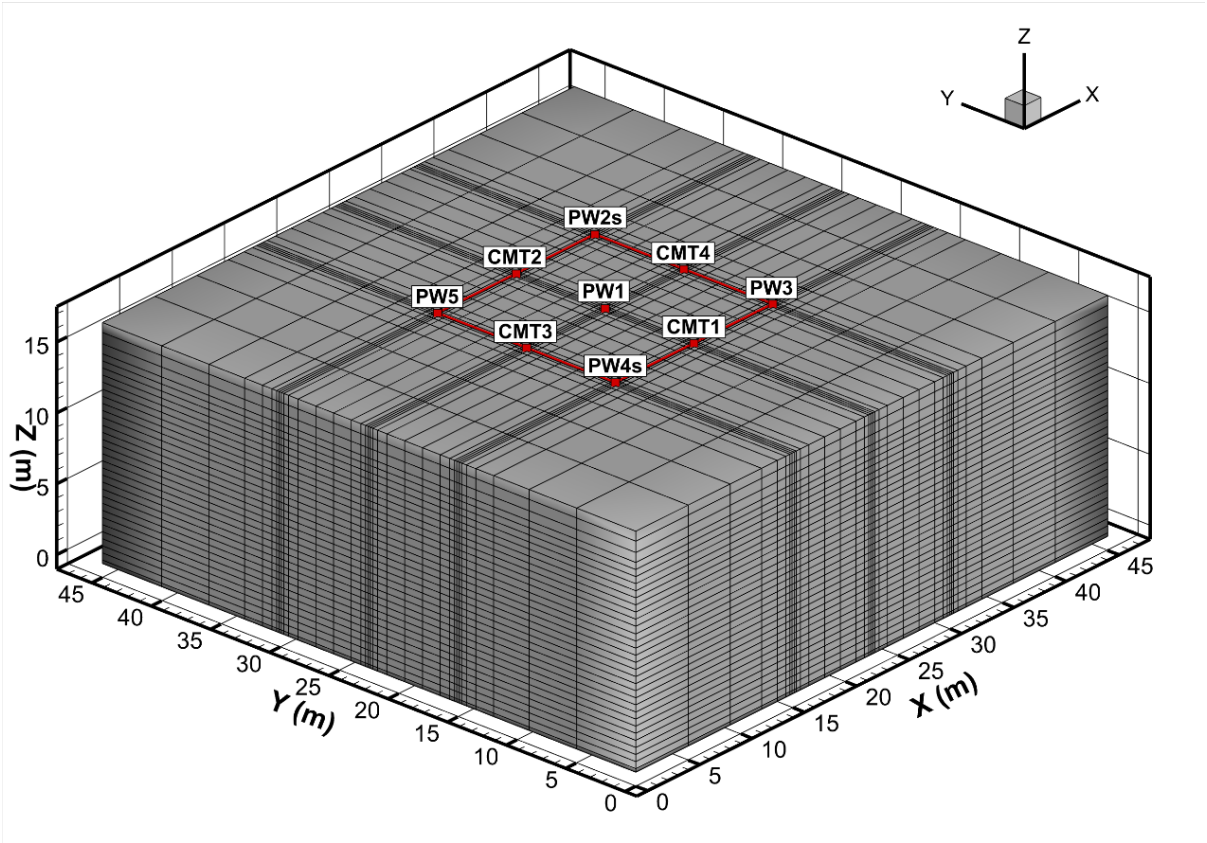

Fig. E2: Discretization of simulation domain for groundwater flow simulation.

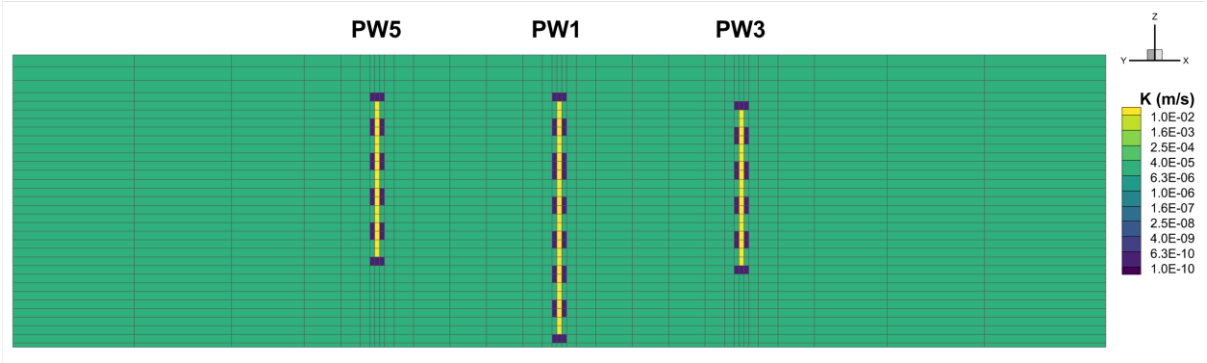

Fig. E3: Cross-section along A-B (Figure 1a) showing numerical multilevel-screened wells at the NCRS.

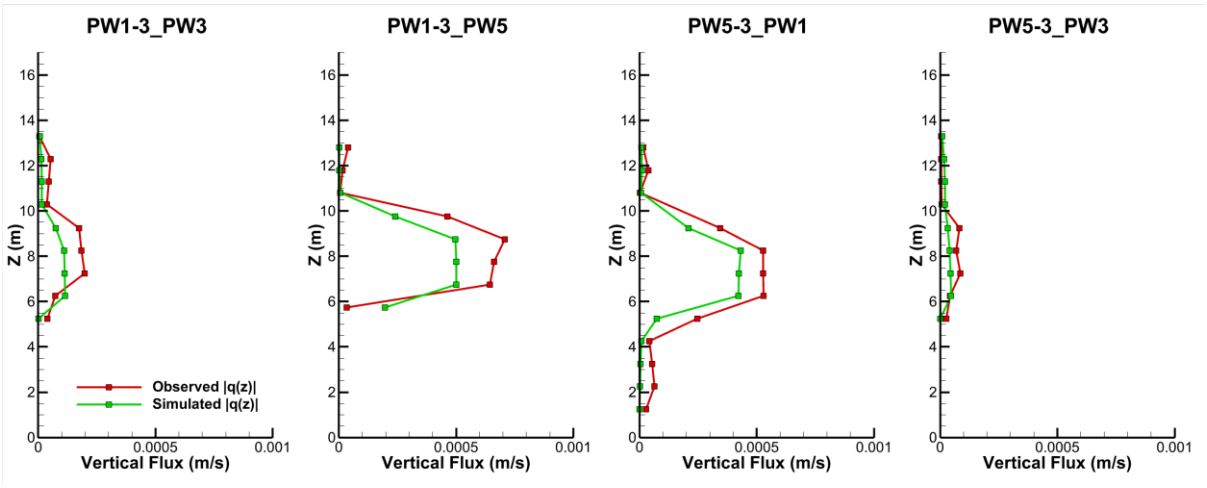

Fig. E4: Comparison of the simulated versus observed cross-hole flow logs using previous characterization result (Case 3d in the work of Zhao and Illman (2017)) with numerical multilevel-screened wells.

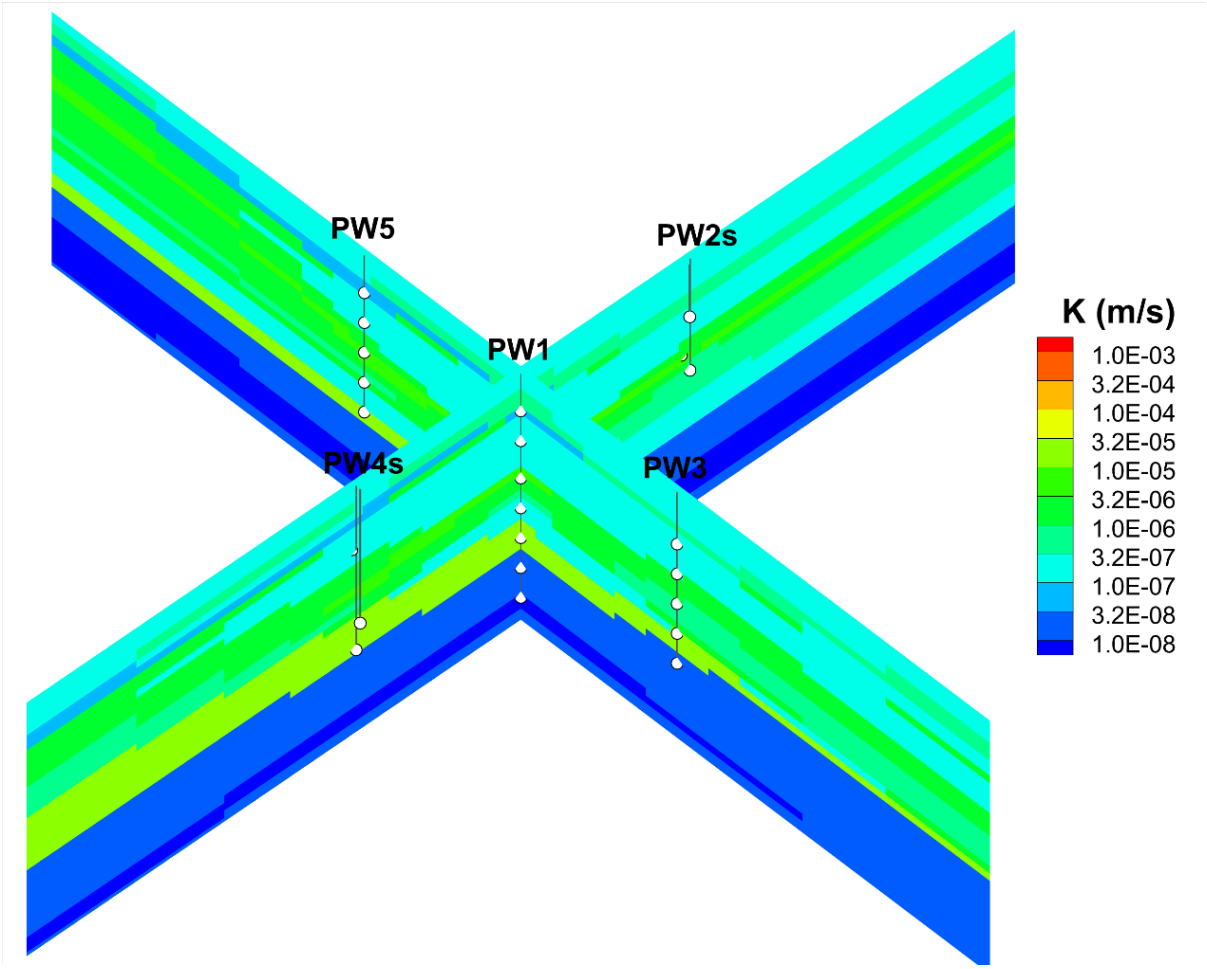

Fig. E5: Zonation *K* distribution based on the constructed 19-layer geological model and laboratory permeameter test results, incorporated as prior geological information for inverse modeling along cross-sections A-B and C-D on Fig. 6.1a. Monitoring ports are indicated as white circles.

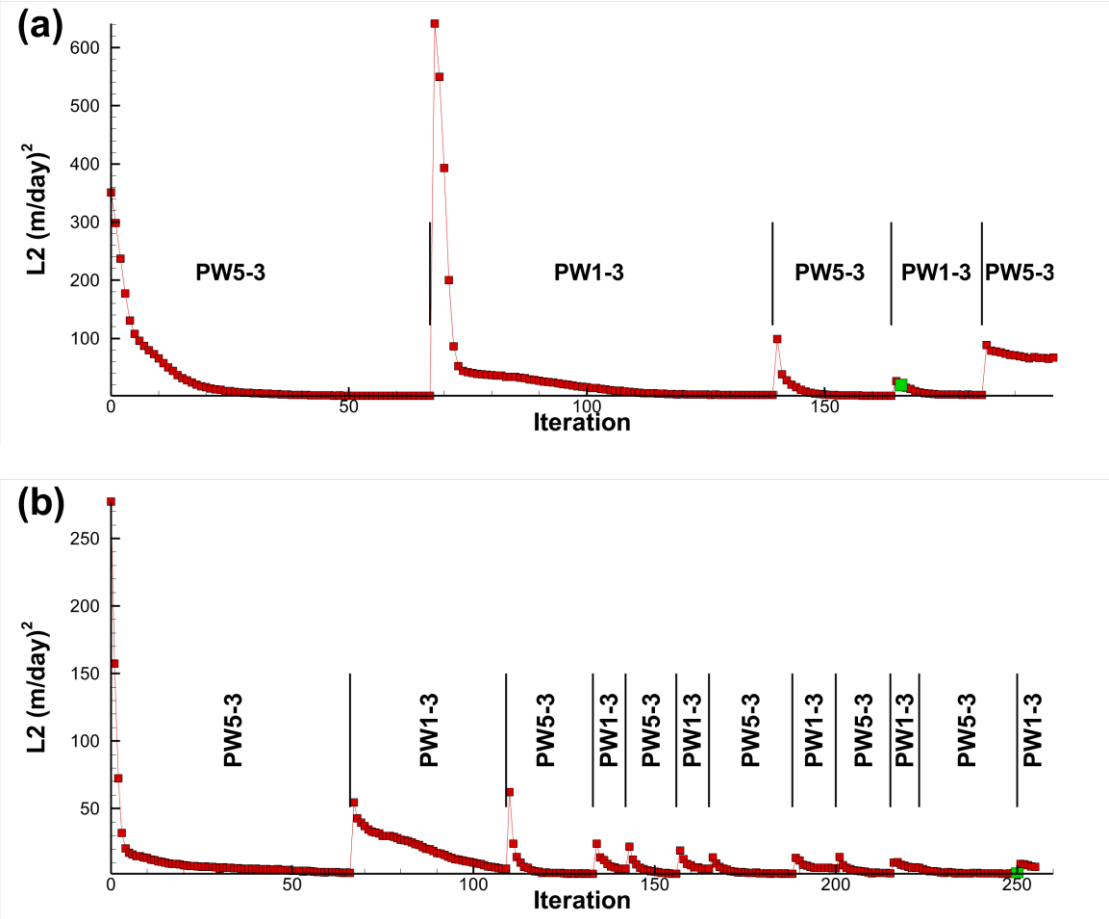

Fig. E6: Variations of L_2 norm that indicate the convergence of geostatistical inversions of cross-hole flowmeter data for site characterization using the method of loop-iteration. (a) cross-hole flowmeter data only, (b) with prior geological information.

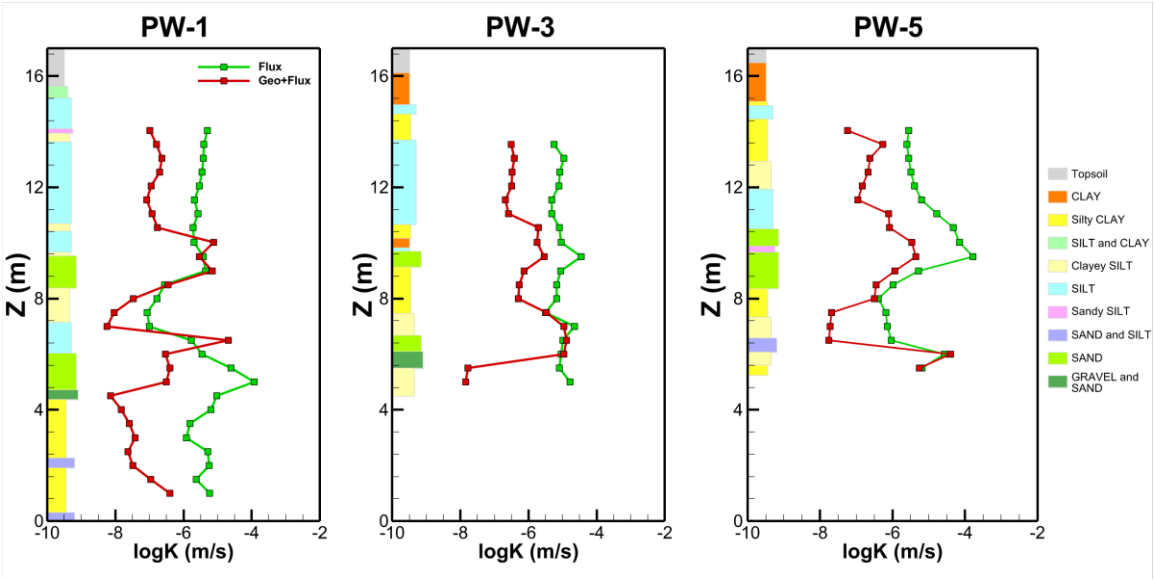

Fig. E7: Vertical distributions of estimated K values through geostatistical inverse analysis of cross-hole flowmeter data at multi-screen well locations. In each plot, green curve shows the scenario without prior geological information and red one indicates the scenario with prior geological information.

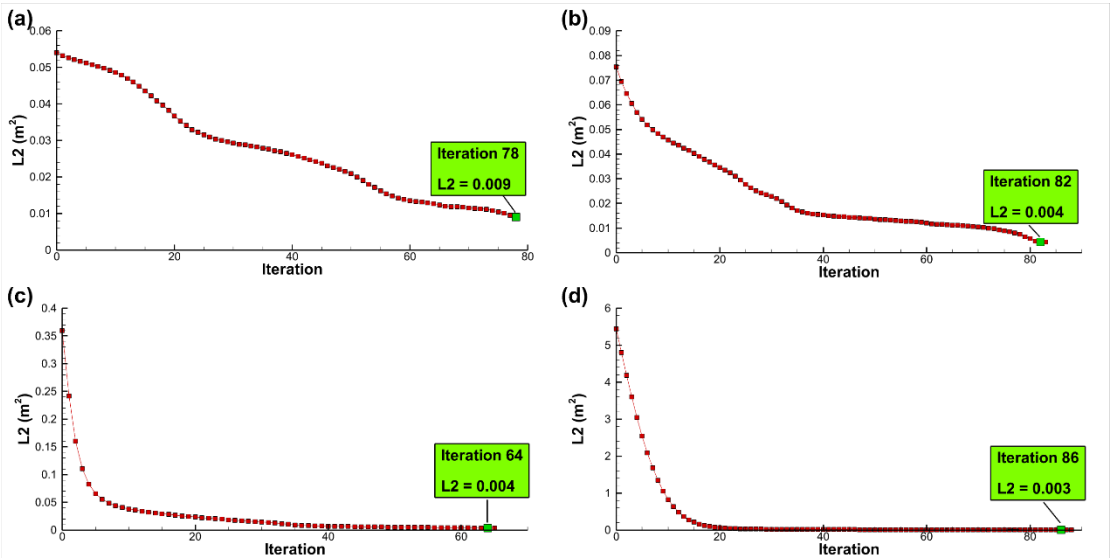

Fig. E8: L_2 norms showing the convergence of SSHT analyses of head data with/without the integration of different additional datasets. (a) head data only, (b) cross-hole flowmeter and head data, (c) head data with prior geological information, and (d) cross-hole flowmeter and head data with prior geological information.

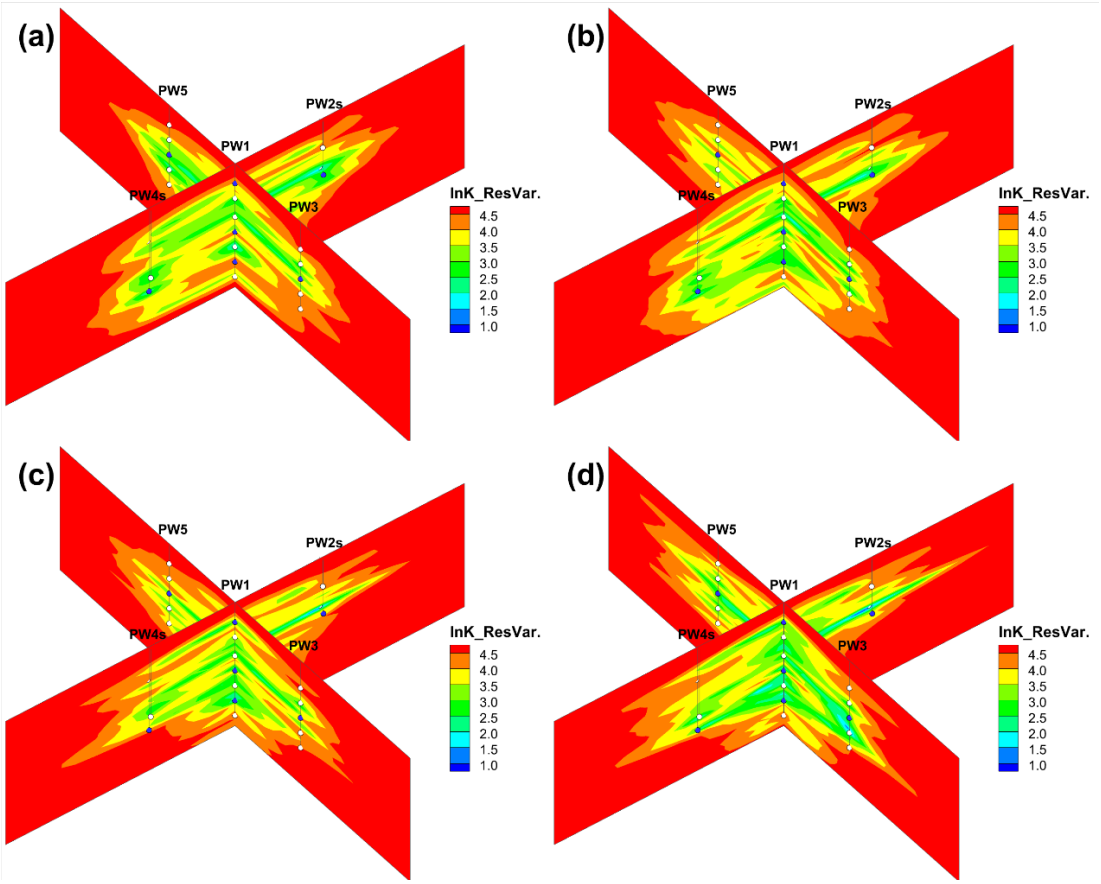

Fig. E9: Computed residual variance maps of ln*K* from Cases 1 to 4, shown as (a) though (d), respectively along cross-sections A-B and C-D on Figure 1a. Pumping/injection and monitoring ports are indicated as blue and white circles, respectively.

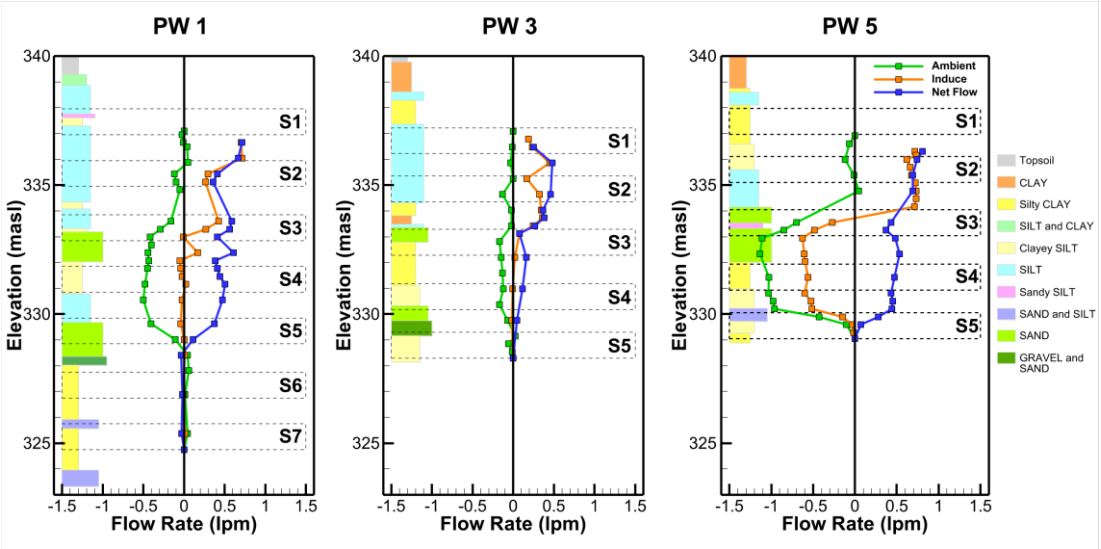

Fig. E10: Single-hole flowmeter measurements at multi-screen wells. Negative and positive values indicate groundwater flow downward and upward, respectively. Green curves show the measured flow logs under the ambient condition, orange curves show the measured flow logs under the induced condition, and blue curves show the computed net flow indicating the vertical flow induced by pumping only.

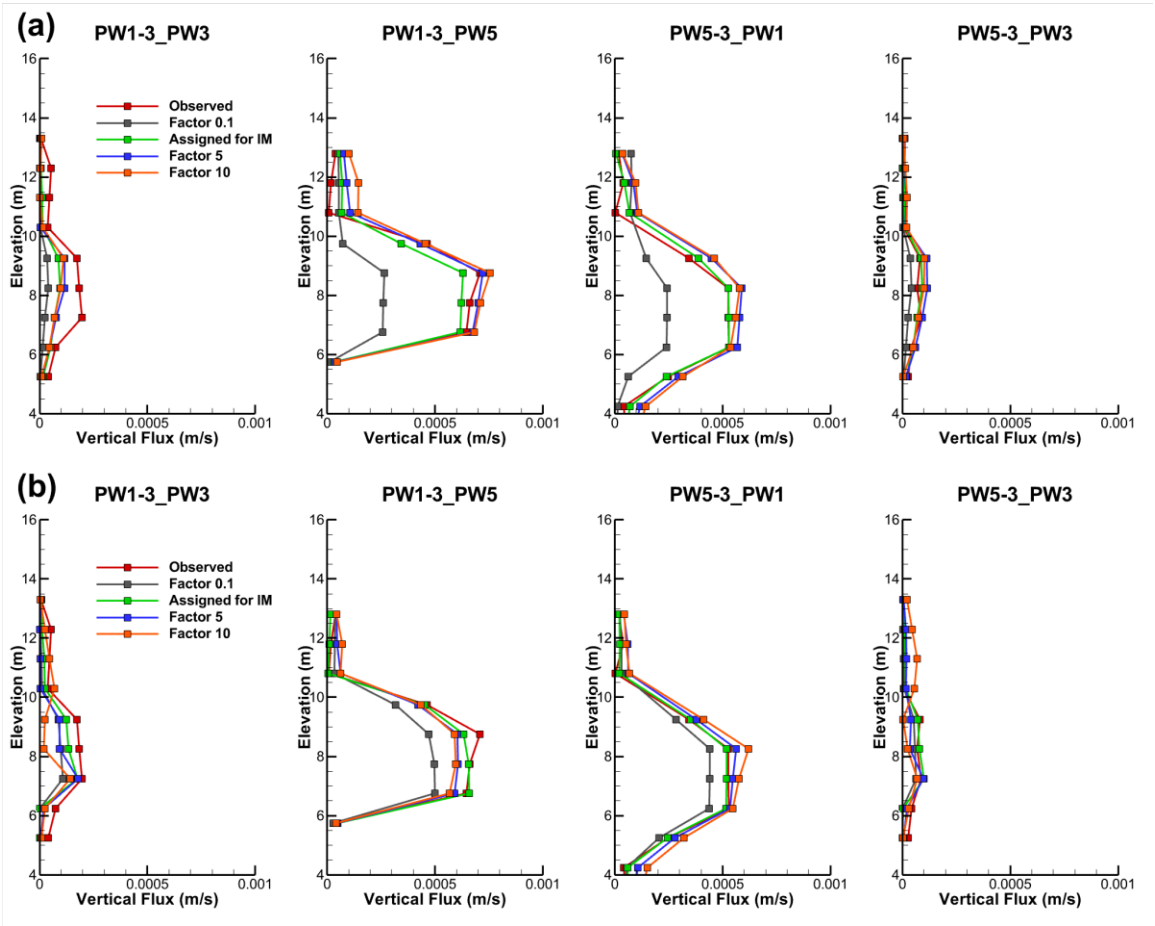

Fig. E11: Comparison of the simulated versus observed vertical flux magnitudes at multilevel-screened wells induced by steady-state pumping at PW1-3/PW5-3 using K tomograms estimated through geostatistical inversions of flux data but with different sets of K values assigned to numerical boreholes. (a) for Scenario 1 and (b) for Scenario 2. Factor 0.1: vertical conduit $K = 1 \times 10^{-3}$ m/s and cased interval $K = 1 \times 10^{-7}$ m/s; Factor 5: vertical conduit $K = 5 \times 10^{-2}$ m/s and cased interval $K = 2 \times 10^{-9}$ m/s; Factor 10: vertical conduit $K = 1 \times 10^{-1}$ m/s and cased interval $K = 1 \times 10^{-9}$ m/s.