Three-dimensional hydraulic tomography analysis of long-term municipal wellfield operations: Validation with synthetic flow and solute transport data

NING LUO[1], Walter Arthur Illman[1], Yuanyuan Zha[2], Young-Jin Park[1], and Steven J. Berg[3]

[1]University of Waterloo
[2]Wuhan University
[3]Aquanty Inc.

November 21, 2022

Abstract

This study proposes the utilization of municipal well records as an alternative dataset for large-scale heterogeneity characterization of hydraulic conductivity () and specific storage () using hydraulic tomography (HT). To investigate the performance of HT and the feasibility of utilizing municipal well records, a three-dimensional aquifer/aquitard system is constructed and synthetic groundwater flow and solute transport experiments are conducted to generate data for inverse modeling and validation of results. In particular, we simultaneously calibrate four groundwater models with varying parameterization complexity using five datasets consisting of different time durations and periods. Calibration and validation results are qualitatively and quantitatively assessed to evaluate the performance of investigated models. The estimated and tomograms from different model cases are also validated through the simulation of independently conducted pumping tests and conservative solute transport. Our study reveals that: 1) the HT analysis of municipal well records is feasible and yields reliable heterogeneous and distributions where drawdown records are available; 2) accurate geological information is of critical importance when data density is low and should be incorporated for geostatistical inversions; 3) the estimated and tomograms from the geostatistical model with geological information are capable in providing robust predictions of both groundwater flow and solute transport. Overall, this synthetic study provides a general framework for large-scale heterogeneity characterization using HT through the interpretation of municipal well records, and provides guidance for applying this concept to field problems.

Three-dimensional hydraulic tomography analysis of long-term municipal wellfield operations: Validation with synthetic flow and solute transport data

Ning Luo[1], Walter A. Illman[1], Yuanyuan Zha[2], Young-Jin Park[1], and Steven J. Berg[1,3]

[1] Department of Earth & Environmental Sciences, University of Waterloo, Waterloo, Ontario, Canada

[2] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, China

[3] Aquanty Inc., Waterloo, Ontario, Canada

Corresponding author: Ning Luo (n2luo@uwaterloo.ca)

Key Points:

- Hydraulic tomography analysis of long-term municipal wellfield records is feasible although data selection requires careful consideration.
- Continuous records with large water-level variations should be included for more accurate estimation of hydraulic parameters.
- Results from hydraulic tomography are validated through synthetic flow and solute transport experiments showing robust performance.

Abstract

This study proposes the utilization of municipal well records as an alternative dataset for large-scale heterogeneity characterization of hydraulic conductivity (K) and specific storage (S_s) using hydraulic tomography (HT). To investigate the performance of HT and the feasibility of utilizing municipal well records, a three-dimensional aquifer/aquitard system is constructed and synthetic groundwater flow and solute transport experiments are conducted to generate data for inverse modeling and validation of results. In particular, we simultaneously calibrate four groundwater models with varying parameterization complexity using five datasets consisting of different time durations and periods. Calibration and validation results are qualitatively and quantitatively assessed to evaluate the performance of investigated models. The estimated K and S_s tomograms from different model cases are also validated through the simulation of independently conducted pumping tests and conservative solute transport. Our study reveals that: 1) the HT analysis of municipal well records is feasible and yields reliable heterogeneous K and S_s distributions where drawdown records are available; 2) accurate geological information is of critical importance when data density is low and should be incorporated for geostatistical inversions; 3) the estimated K and S_s tomograms from the geostatistical model with geological information are capable in providing robust predictions of both groundwater flow and solute transport. Overall, this synthetic study provides a general framework for large-scale heterogeneity characterization using HT through the interpretation of municipal well records, and provides guidance for applying this concept to field problems.

1. Introduction

Planning for the optimized use and management of groundwater resources requires the accurate characterization of subsurface heterogeneity in hydraulic conductivity (K) and specific storage (S_s), which are two importantt hydraulic properties for the construction of groundwater flow models, and in particular, for solute transport simulations (Ni et al., 2009; Illman et al., 2012). In the past few decades, numerous efforts have been dedicated to map the spatial distribution of K and S_s. Typically, geostatistical interpretation of small-scale K estimates obtained from core samples, slug tests, flowmeter surveys, and single-hole pumping/injection tests is applied to map its spatial distribution (e.g., Salamon et al., 2007; Sudicky et al., 2010; Alexander et al., 2011), while S_s is commonly treated to be homogeneous as its variability is considered to be much less than K in natural geological formations (Gelhar, 1993). Using this approach, a sufficient number of small-scale estimates is required to fully capture the heterogeneity patterns of hydraulic properties (Rehfeldt et al., 1992). On the other hand, Kuhlman et al. (2008) demonstrated that the interpolated K and S_s fields strongly relied on the estimated small-scale values, which may be biased in representing realistic conditions due to the restricted assumptions implied in analytical solutions (e.g., Theis (1935)) for these estimates.

An alternative approach to the geostatistical interpretation of small-scale values, hydraulic tomography (HT) was proposed and developed (e.g., Gottlieb and Dietrich, 1995; Yeh and Liu, 2000) for subsurface heterogeneity characterization. Fundamentally, the HT approach involves the inverse modeling of groundwater response data collected at various locations during a series of spatially varying pumping/injection tests. Yeh et al. (2008) concluded that the data collected in such tomographic surveys provide many constraints for model calibration, yielding more accurate estimations of K and S_s fields with less uncertainty in comparison to traditional characterization methods.

The robust performance of HT in revealing K and S_s heterogeneity has been demonstrated through numerous numerical (e.g., Yeh and Liu, 2000; Bohling et al., 2002; Zhu and Yeh, 2005), laboratory (e.g., Liu et al., 2002; Illman et al., 2007, 2010, 2015; Berg and Illman, 2011a; Zhao et al., 2015; Luo et al., 2017) and dedicated field experiments (e.g., Straface et al., 2007; Bohling et al., 2007; Illman et al., 2009; Berg and Illman, 2011b; Zha et al., 2016; Zhao and Illman, 2017). Nevertheless, most of these 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 using the approach of HT (e.g., Illman et al., 2009; Zha et al., 2016, 2019). In small-scale studies, each single-well pumping/injection test is able to stress the entire aquifer and generate a head response throughout the domain which can be monitored with a well-designed monitoring network. However, designing and conducting HT surveys for large-scale heterogeneity characterization is typically expensive, time-consuming, and sometimes impractical. Illman et al. (2009) applied the approach of transient HT to characterize a kilometer-scale fractured granite site at Mizunami, Japan, using data from two large-scale cross-hole pumping tests. The estimated K and S_s tomograms qualitatively agreed well with observed drawdown records, available fault information, and coseismic groundwater responses during several large earthquakes (Niwa et al., 2012). However, they pointed out that the estimated fault and fracture zones might still involve great uncertainty due to the limited hydraulic data for inverse modeling. Through the inclusion of datasets from two additional pumping tests conducted at the same site, Zha et al. (2016) yielded K and S_s tomograms with improved delineation of fault zones in terms of their locations and patterns. However, they indicated that with limited number of wells, the collected hydraulic data from four large-scale pumping tests were still insufficient to map the detailed distribution of fractures and faults.

Instead of the traditional data collection strategy used for HT surveys, alternative datasets have been proposed and utilized for large-scale heterogeneity characterization using HT (e.g., Kuhlman et al., 2008; Yeh et al., 2009; Wang et al., 2017; Zha et al., 2019). For instance, Kuhlman et al. (2008) applied the HT approach to characterize subsurface heterogeneity (T and S) at the basin scale using head response data collected from multiple simultaneous pumping wells. Through cycling the operation of different sets of pumping wells, the regional aquifer is repeatedly stressed to yield groundwater responses that cover most of the simulation domain. They concluded that the head data collected from potentially disparate aquifer tests could be jointly interpreted to estimate basin-wide aquifer properties using HT. It was suggested that this characterization approach could be applied to municipal or pump-and-treat wellfields with existing monitoring networks. However, due to the operational requirements of municipal well fields, it is unlikely to be able to cease pumping/injection to conduct dedicated pumping tests for aquifer characterization, thus making the application of typical HT methodologies infeasible.

Yeh et al. (2008) provided an opinion of using natural stimuli (e.g., 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 corresponding natural stimuli can be interpreted for hydraulic properties estimation.

Following this thought, Yeh et al. (2009) proposed river stage tomography as a new approach for basin-scale subsurface heterogeneity characterization. Specifically, the migration of river stage perturbation along the river was treated as a natural stimuli that induces groundwater fluctuations over the entire basin. The temporal and spatial variations of river stage as well as the corresponding groundwater response data were then incorporated for inverse modeling to estimate the spatial distribution of hydraulic properties (*T* and *S*) of the basin. The efficiency of river stage tomography in revealing basin-scale hydraulic heterogeneity was later evaluated through a field experiment conducted in Zhoushui River alluvial fan, Taiwan (Wang et al., 2017). 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 monitoring or municipal water-supply wellfields, but they are rarely adopted for mapping the heterogeneity of hydraulic properties. 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. In particular, hydraulic head changes during four distinct events (e.g., system shutdown and resumption, changes in pumping/injection operations, and significant variations of flow rates) were extracted from the existing head records and utilized for inverse modeling. Here, it should be noted that the characteristics of well hydrographs might be quite different from one 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. 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. On the other hand, 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 condition for groundwater modeling.

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. In particular, a synthetic three-dimensional 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 error (e.g., model identification and head measurement) yields a general framework for subsurface heterogeneity characterization using the existing long-term pumping/injection and water-level records.

The main objectives of this synthetic study were to: 1) explore the feasibility of utilizing municipal well data for subsurface heterogeneity characterization using HT, 2) evaluate the performance of three different modeling approaches (homogeneous, geological, and geostatistical models) for HT analyses with well data 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 the simulation of nonreactive tracer migration through the municipal wellfield.

2. 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 Figures 1a and 1b. The size of the model was set to be 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 was not considered. In each geological layer, random K and S_s fields were generated by assuming the Gaussian distributions of ln$K$ and lnS_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 lnS_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 Figures 4d and 5d, respectively, while the statistical details of hydraulic parameters are summarized in Table S1 of the Supplementary Information section. To better evaluate the results, the entire simulation domain was subdivided into three zones (ZONE 1, ZONE 2, and ZONE 3, as shown in Figures 1a and 1b) based on the density of well screens.

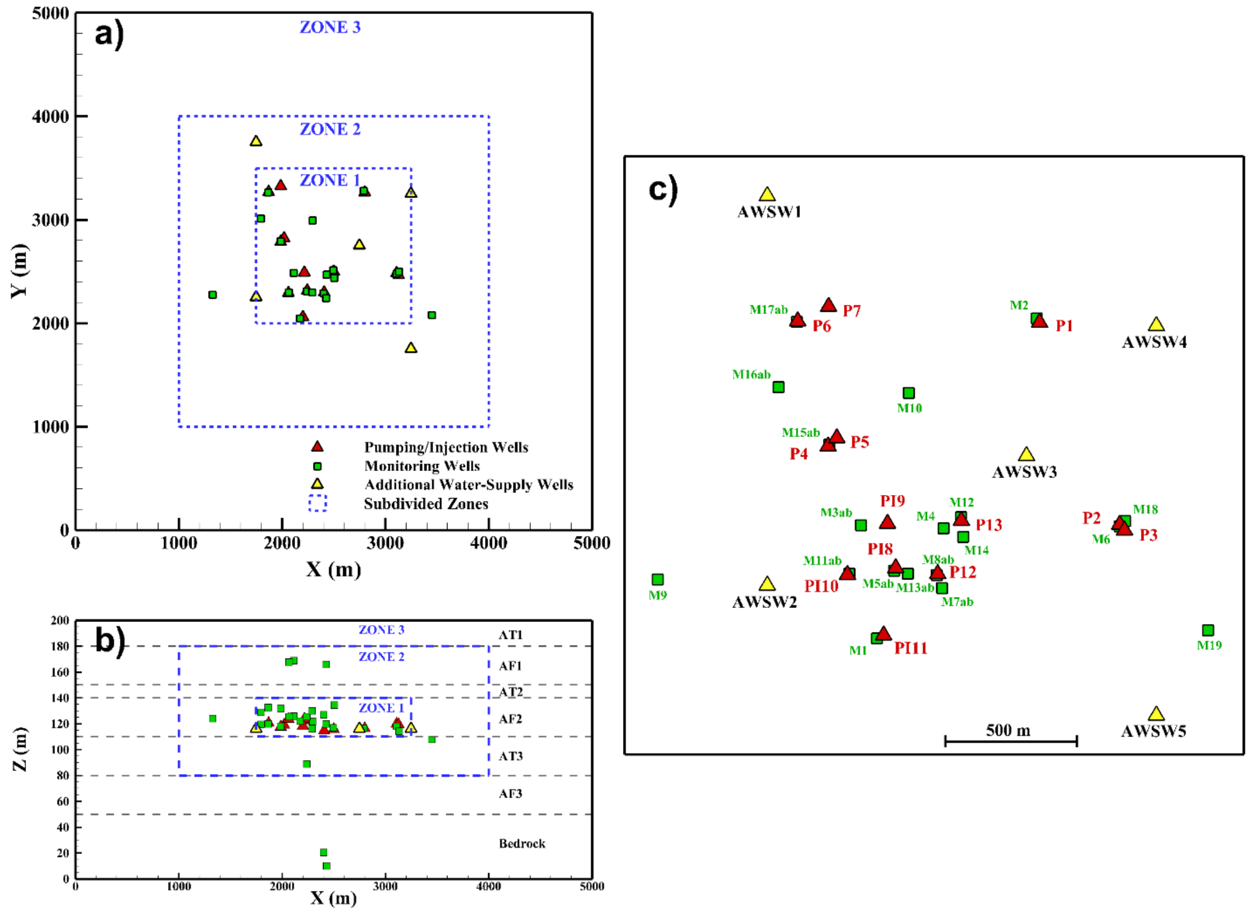

Figure 1: The synthetic layer-cake geological model domain along with the distribution of pumping and monitoring wells. a) and b) illustrate the plan-view and cross-section of the simulation domain, respectively; c) shows spatial distribution of assigned wells with IDs.

To mimic the hydraulic condition in a municipal water-supply wellfield, the same well configuration as the Mannheim East wellfield was applied for the synthetic study. Within the wellfield, a subdivide 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 Figure 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 Figure S1 of the Supplementary Information section. The mesh was refined around wells, but became 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 equation can be expressed as:

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

subject to initial and boundary conditions:

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

where, in Eq. (1), $\boldsymbol{\nabla}$ is the gradient operator, $K(\mathbf{x})$ is hydraulic conductivity (L T^{-1}), h is hydraulic head (L), $Q(\mathbf{x}_p)$ is the rate of pumping per unit volume (T^{-1}) at location $\mathbf{x}_p$, and $S_s(\mathbf{x})$ is specific storage (L^{-1}). In Eq. (2), h_0 represents the initial hydraulic head, h_1 is a constant head (L) at boundary Γ_1, q is the specific discharge (L T^{-1}) at the Neumann boundary Γ_N, and $\mathbf{n}$ is a unit vector normal to Γ_2.

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 (Figure 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 Figure S2 of the Supplementary Information section) for the analyses presented in this study. 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 Figure S3 of the Supplementary Information section). 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 at the study site.

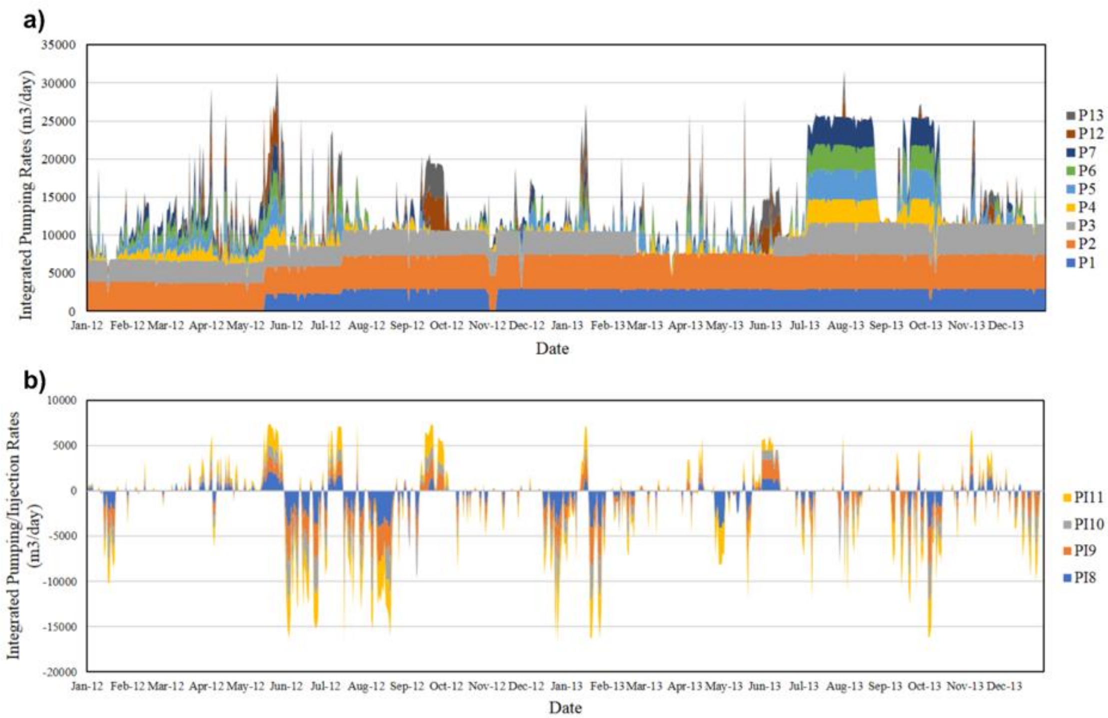

Figure 2: Extraction (positive) and injection (negative) pumping rate records at all 13 municipal wells during the years of 2012 and 2013 from the WRAS+ database.

3. 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) are 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 are selected for model calibration in the synthetic study, as shown in Figure 3, with their properties summarized in Table S2. In particular, Dataset A includes daily observation data in all 28 monitoring wells during the first 30 days, while Datasets B and C extend the simulation durations to 60 and 120 days, respectively. Daily observation data during the second 30 days are extracted as Dataset D, which shares the same simulation duration as Dataset A, but with a relatively small magnitude of water-level variations. Dataset E has the same simulation duration as Dataset C (120 days); however, instead of incorporating all observation points, only the periods with large water-level variations are selected and utilized for model calibration.

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 is adopted here to minimize the effect of uncertain initial conditions. In particular, pumping/injection rate records prior to the observation data are utilized for model spin-up and incorporated for model calibration. The model spin-up time is 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 are then compared quantitatively to the observed ones (Figure S1), with the comparison results illustrated in Figure S4 of the Supplementary Information section. Results 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 are extracted and incorporated for model calibration in this synthetic study.

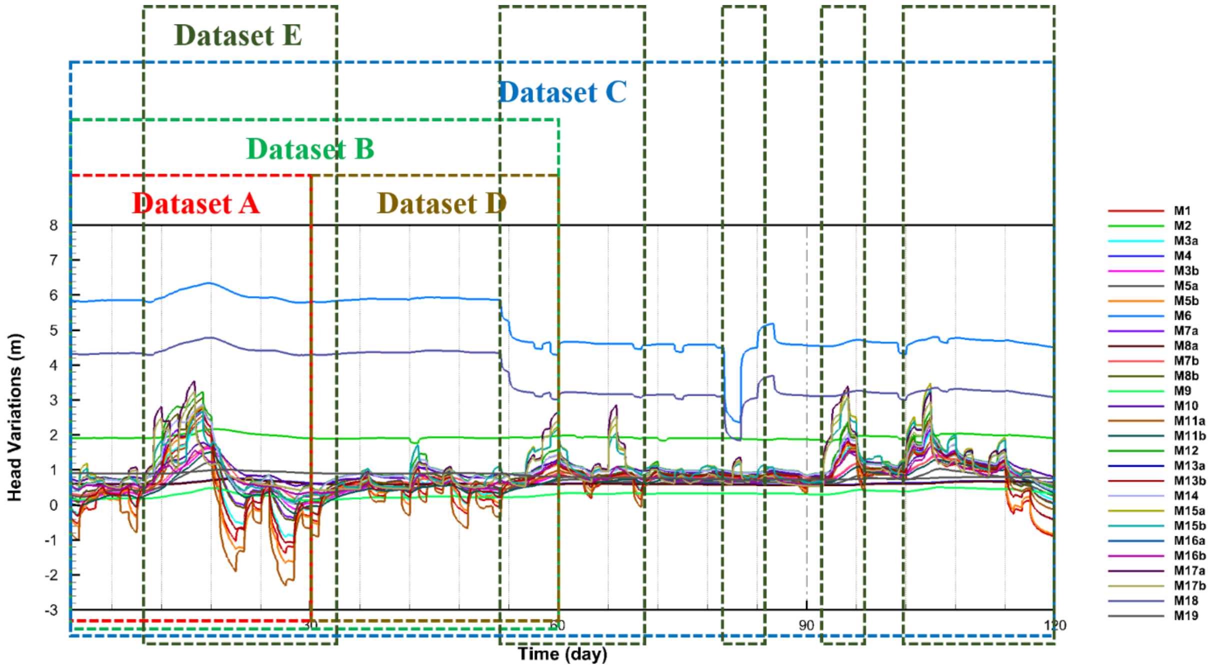

Figure 3: Selected datasets for model calibration. Dataset A (0 – 30 days), Dataset B (0 – 60 days), Dataset C (0 – 120 days), Dataset D (30 – 60 days), Dataset E (0 – 120 days with selected large drawdown variation periods).

4. Groundwater Flow Modeling Approaches

4.1 Case 1: Effective Parameter Model

The synthetic multi-aquifer/aquitard system is 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, 2005), and is referred to as the 'effective parameter' model. The effective parameter model provides zero-resolution on subsurface heterogeneity; however, it may still able to describe the overall behavior of groundwater flow in the system. Furthermore, the estimated effective K and S_s values can 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 is 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.

4.2 Case 2: Geological Model

The response from the synthetic multi-aquifer/aquitard system is then used to calibrate the geological model, which is normally adopted for groundwater flow modeling at large scales (e.g., regional or basin scales). In this approach, each geological layer is characterized as a homogeneous, isotropic medium, and a uniform set of K and S_s is 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 (Zhao et al., 2016; Luo et al., 2017). 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 is 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 are 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 note in previous studies (Berg and Illman, 2011b).. As a result, the heterogeneous K and S_s fields based on the stratigraphic information are applied as initial guesses. In total, 14 parameters are estimated for subsurface heterogeneity characterization using the geological model.

4.3 Case 3: Geostatistical Models

As a third case, the head response to the synthetic municipal well data are interpreted with highly parameterized geostatistical models for subsurface heterogeneity characterization. All geostatistical inversions are 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. In this study, geostatistical inversion using SimSLE assumes a transient groundwater flow field, and the natural logarithm of K and S_s are both treated as multi-Gaussian, second-order stationary, stochastic processes. Other than estimating the spatial distributions of hydraulic parameters (K and S_s tomograms), SimSLE also provides the variance maps of ln$K$ and lnS_s to describe the uncertainty of estimated values, with large variance meaning high uncertainty in the estimated parameters and vice versa.

Based on the differences in initial K and S_s fields, two geostatistical inversion cases are investigated. For Case 3a, homogeneous initial K and S_s fields are 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) are utilized as initial guesses and assigned to the entire domain. For Case 3b, geological information is incorporated for model calibration. Geostatistical inversions in this case start from heterogeneous initial K and S_s fields which are the same as those utilized for geological model calibration. For both cases, the variances of ln$K$ and lnS_s ($\sigma^2_{\ln K}$, $\sigma^2_{\ln Ss}$) are initially set to be 4.0 and 2.0, respectively, while the correlation scales are 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 statistical properties of heterogeneous K and S_s fields are commonly unknown for field studies, the input properties for geostatistical inversions are set to be different from those utilized for "true" K and S_s fields generation (as shown in Table S1 of the Supplementary Information section).

5. Results and Discussion

In this study, five datasets (Datasets A-E) are interpreted with four different models (Cases 1, 2, 3a and 3b). Results from all investigated models are summarized and examined. In particular, K and S_s values estimated from different models are first compared to the "true" fields to illustrate the accuracy of these estimates. Then, calibration and validation results are assessed qualitatively and quantitatively by plotting scatterplots of simulated versus observed head variations and evaluating model errors, respectively. The evaluation of model errors is 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:

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

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

$$R^2 = \left[\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(\hat{x}_i - \mu_{\hat{x}})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2 \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - \mu_{\hat{x}})^2}}\right]^2 \qquad (5)$$

where n is the total number of head data, x_i and $\hat{x}_i$ represent ith simulated and observed head data, respectively, μ_x and $\mu_{\hat{x}}$ represent averaged simulated and observed head data, respectively. These values were calculated to quantitatively analyze the discrepancy and correspondence between the simulated and observed head data.

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 the Supplementary Information section. Then, statistical summary (L_1, L_2, and R^2) of the validation results obtained from all investigated models is presented to show the effect of data selection on inverse modeling.

5.1 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 ± 0.14 m/day and S_s = 2.56 × 10^{-4} ± 2.5 × 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 to 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.

Figure 4 illustrates the obtained K tomograms from the geological (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. Figure5 illustrates the same, but for the estimated S_s tomograms.

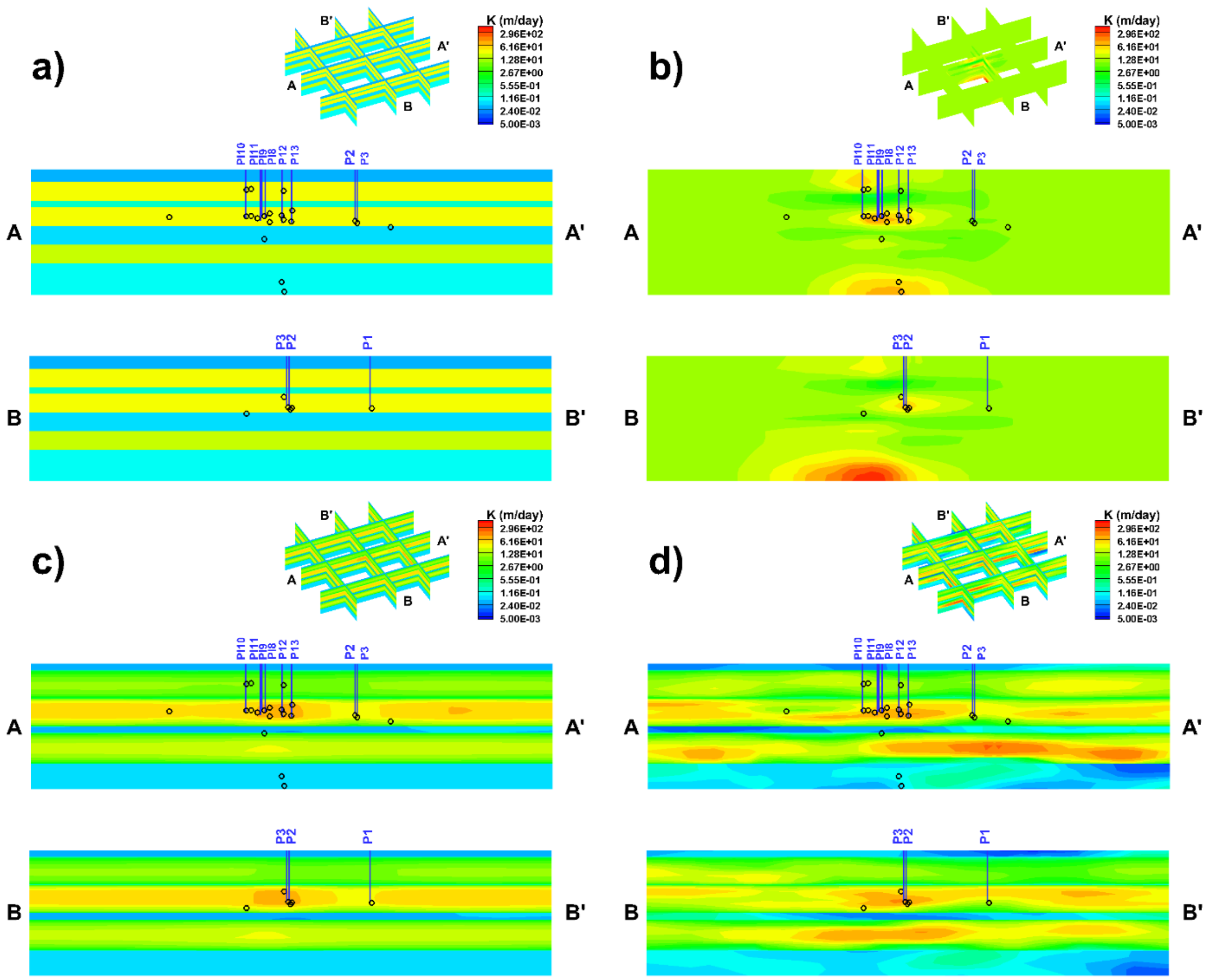

Figure 4: Estimated K tomograms from three model cases through the interpretation of Dataset A as well as the "true" K field. a) Case 2: geological model, b) Case 3a: geostatistical model without geological information, c) Case 3b: geostatistical model with geological information, and d) "true" K field. In each contour map, small black circles represent the location of monitoring well screens.

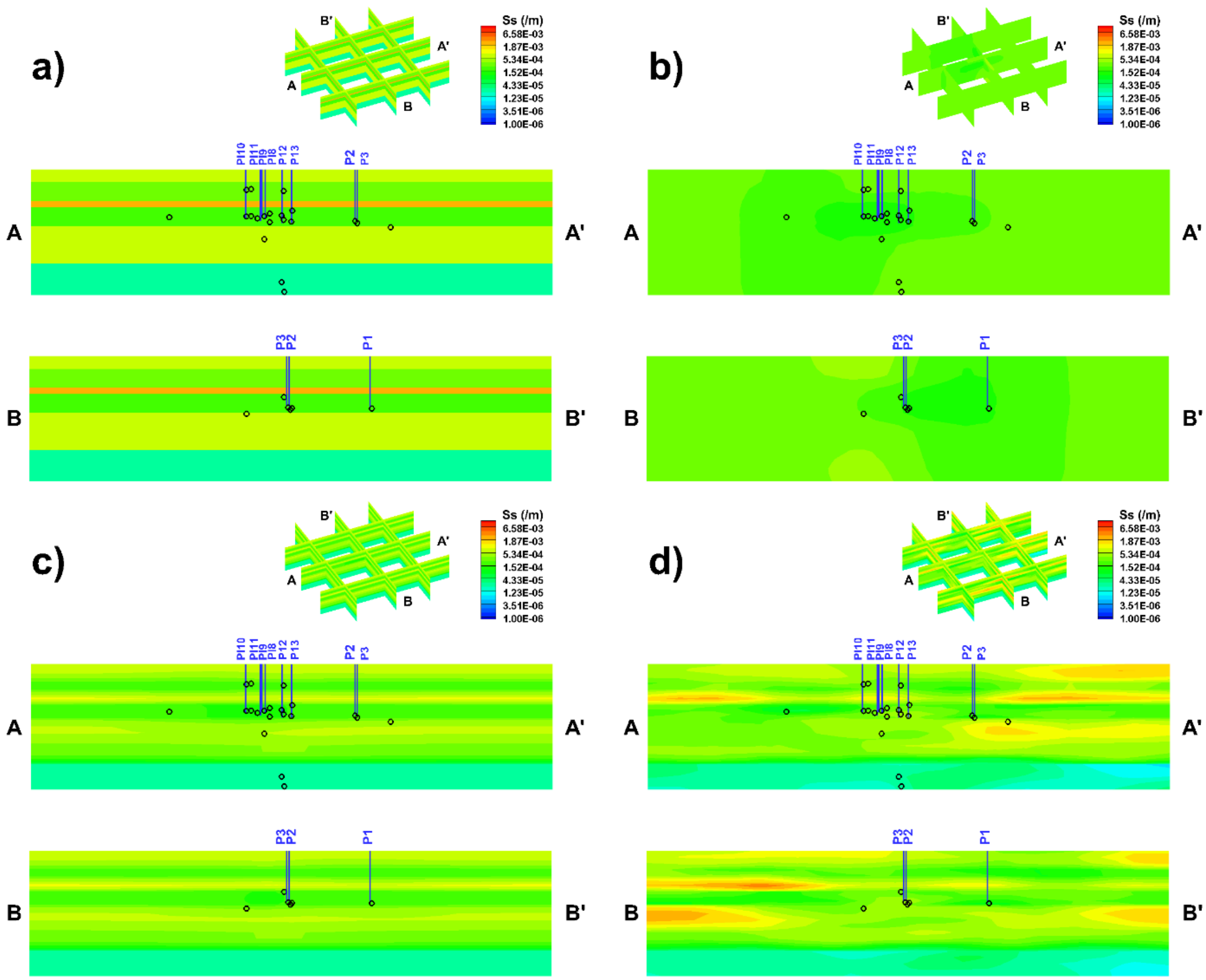

Figure 5: Estimated S_s tomograms from three model cases through the interpretation of Dataset A as well as the “true” S_s field. a) Case 2: geological model, b) Case 3a: geostatistical model without geological information, c) Case 3b: geostatistical model with geological information, and d) “true” S_s field. In each contour map, small black circles represent the location of monitoring well screens.

As shown in Figures 4a and 5a, the estimated K and S_s values from the geological model (Case 2) are found to roughly describe the average hydraulic properties of each geological layer in comparison to the “true” field (Figures 4d and 5d for K and S_s, respectively). Here, it should be noted that geological models in this synthetic study are 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 significant uncertainty in the geological model constructed with sparse boreholes and utilized as prior information. To further evaluate these estimates, natural logarithm of these K and S_s estimates as well as their 95% confidence intervals are plotted as Figure S13 of the Supplementary Information section. 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 provided by Luo et al. (2017) 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 the 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 are 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, such a scenario of including a dense dataset was not included in this study.

The obtained K and S_s tomograms from the geostatistical models are illustrated as Figures 4b – 4c and Figures 5b – 5c, respectively. As shown in Figure 4b, 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 is found with 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 Figure 5b. 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, as shown in Figures 4c and 5c, respectively. In particular, greater detail in K heterogeneity is revealed within the water-supply aquifer (AF2), resulting the spatial distribution of K in this layer comparable to that in 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 further be 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 still revealed after the incorporation of geological information, as shown in Figure 5c.

The estimated K and S_s values from geostatistical models (Case 3a and 3b) are 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 (as shown in Figure S18 of the Supplementary Information section), with larger variances indicate 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 attribute to the factor of the temporal resolution of observation data for model calibration, as discussed above. The estimated K and S_s values are then compared to the "true" values by plotting the scatterplots of corresponding estimated versus "true" ln$K$ and lnS_s values (as shown in Figure S23 of the Supplementary Information section). Comparison results reveal 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 methods. 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 are then assessed qualitatively and quantitatively by plotting the scatterplots of calibration results, as shown in Figure 6. 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 Figure 6 reveals that the calibration results in terms of head data matching improve when a larger number of estimated parameters are accounted for inverse modeling (from Case 1 to Cases 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 observation data. After incorporating the geological information, the geostatistical model (Case 3b) yields the best fit of simulated and observed head variations (Figure 6d).

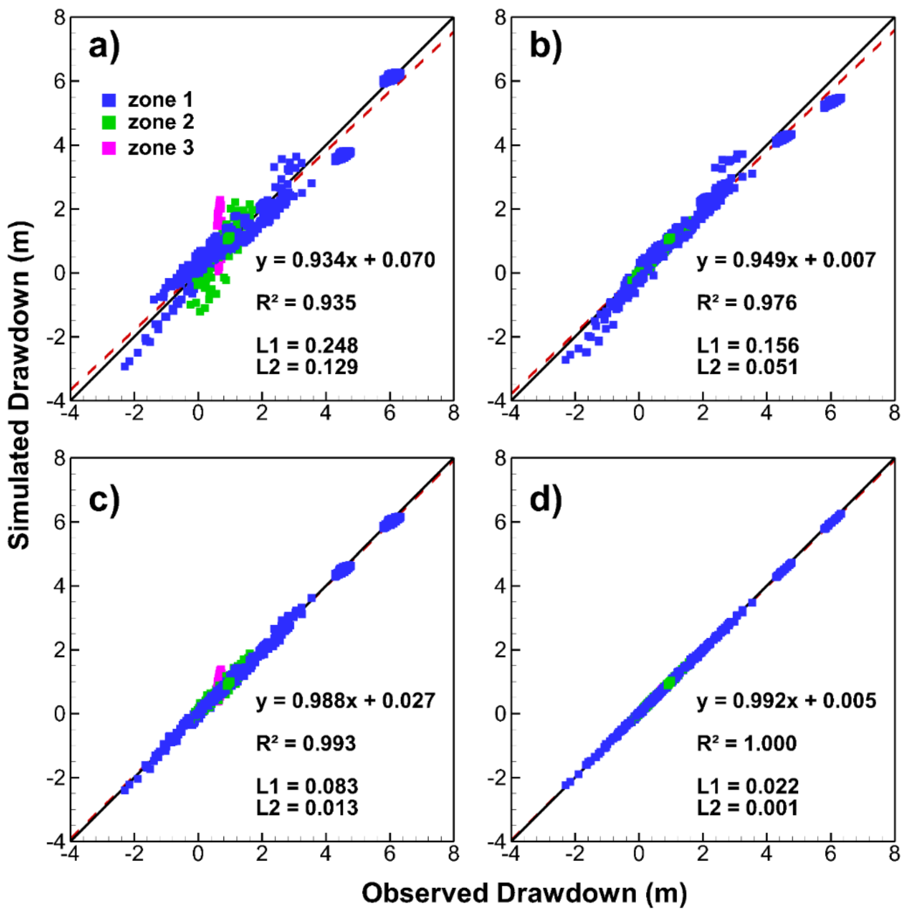

Figure 6: Calibration scatterplots (Dataset A) of simulated versus "observed" drawdowns for four model cases. a) Case 1: effective parameter model, b) Case 2: geological model, c) Case 3a:

geostatistical model without geological information, d) Case 3b: geostatistical model with geological information.

5.2 Model Validation

Model validation in this study is performed in two scenarios. For Scenario 1, the municipal well data during the second half year of 2013 are utilized for model validation. Specifically, the obtained K and S_s tomograms are 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) that not used in the calibration effort are utilized for model validation. The validation scatterplots of different model cases associated with Dataset A are illustrated in Figures 7 and 8, for Scenarios 1 and 2, respectively.

Examination of Figure 7 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 (Figure 7d) performs the best in continuously predicting drawdown variations for the entire domain, followed by Case 3a (Figure 7c), 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) reveal greater loss of heterogeneity details (as shown in Figures 4b and 5b 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.

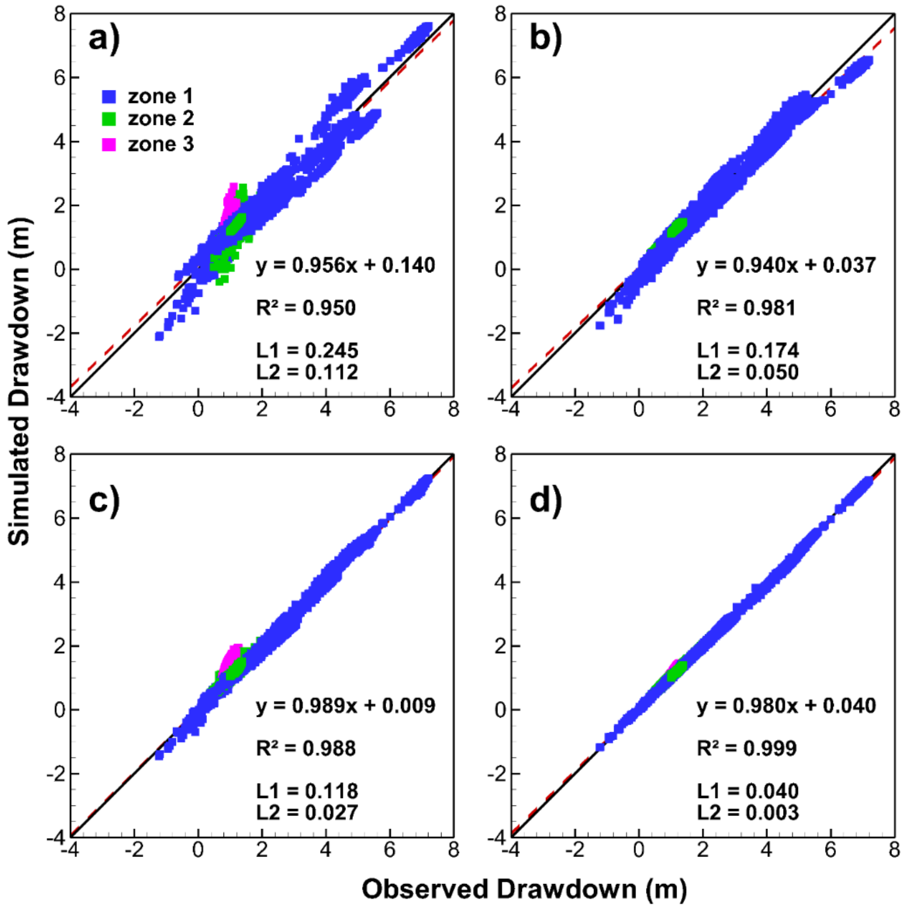

Figure 7: Validation scatterplots (Dataset A) of simulated versus "observed" municipal well data (Scenario 1) for four model cases. a) Case 1: effective parameter model, b) Case 2: geological model, c) Case 3a: geostatistical model without geological information, d) Case 3b: geostatistical model with geological information.

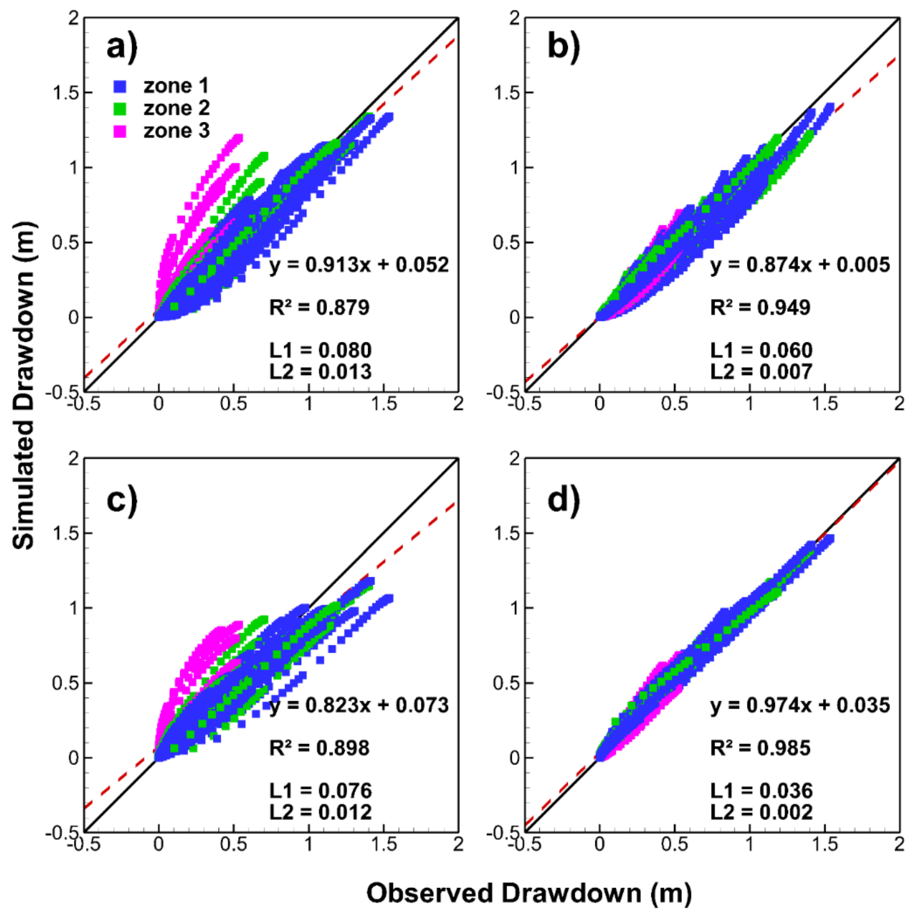

Figure 8: Validation scatterplots (Dataset A) of simulated versus "observed" independent pumping test data (Scenario 2) for four model cases. a) Case 1: effective parameter model, b) Case 2: geological model, c) Case 3a: geostatistical model without geological information, d) Case 3b: geostatistical model with geological information.

To ensure a more credible validation of the different models, independent pumping test data that was not used in the calibration effort are utilized for model validation (Scenario 2), as suggested by Illman et al. (2010). As shown in Figure 8, Case 1 (Figure 12a) still performs the worst in predicting drawdowns from the independent pumping tests. However, it is surprising to find that Case 2 (Figure 8b) provides much better prediction results in comparison to Case 3a (Figure 8c), 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 by Luo et al. (2017). After incorporating geological information, Case 3b (Figure 8d) 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.

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 for 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 Luo et al. (2017), close attention should be paid in constructing accurate geological model when using the actual municipal well data for site characterization.

5.3 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 are computed and plotted in Figure 9 with all values summarized in Table S3 of the Supplementary Information section. In general, when different datasets are included for model calibration, the effective parameter model (Case 1) always performs the worst in predicting groundwater flow, while the geostatistical model with geological information (Case 3b) always performs the best. On the other hand, the performance of the geological model (Case 2) and the geostatistical model without geological information (Case 3a) vary from one dataset to another.

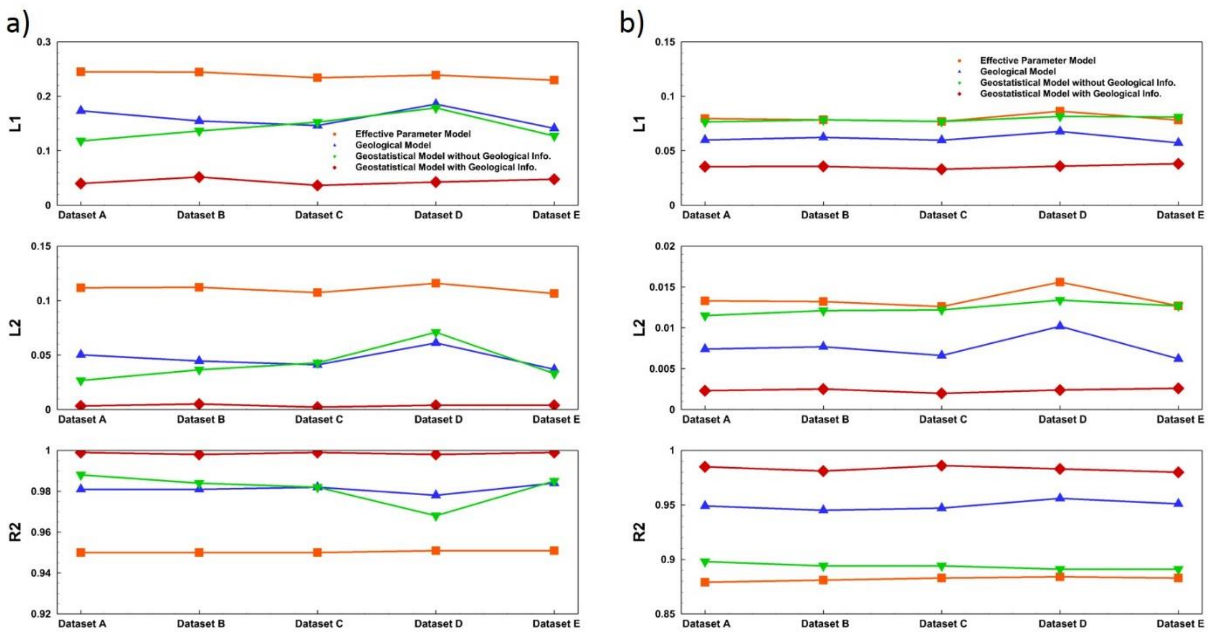

Figure 9: Statistical Summary (L_1, L_2, and R^2) of validation results for four model cases when different datasets were incorporated for model calibration. a) municipal well data (Scenario 1), b) independent pumping test data (Scenario 2).

When more observation points with longer simulation durations are included for model calibration (from Dataset A to Dataset C), the estimated K and S_s tomograms from Case 2 show distinct improvement in continuously predicting municipal well data (Scenario 1, as shown in Figure 9a) 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 Figure 9b); however, slightly better prediction results are 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 are 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 of Dataset A, but with much smaller magnitude of head variations. Results reveal that the validation results associated with Dataset D are 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 yields 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.

The results presented above reveal that the effects of data selection on inverse modeling are different for different modeling approaches. Through the comparison of the validation results from all investigated models, the geostatistical model with geological information (Case 3b) is suggested to interpret continuous head data with large variations for large-scale heterogeneity characterization. However, new approaches need to be developed for big data synthesis and intelligent data selection for inverse modeling.

6. 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. To investigate this issue, additional model runs are performed by simulating solute transport using the estimated K and S_s tomograms. Results are 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 are utilized.

6.1 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 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. 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 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 can be expressed as:

$$-\boldsymbol{\nabla}\cdot(\mathbf{q}C-\theta_s\mathbf{D}\boldsymbol{\nabla}C)\pm Q_c=\theta_s\frac{\partial C}{\partial t} \qquad (6)$$

$$D_{ij}=[\alpha_L-\alpha_T]\frac{v_i v_j}{|\mathbf{v}|}+\alpha_T|\mathbf{v}|\delta_{ij}+D_0\delta_{ij}, \qquad where\ i,j=x,y,z \qquad (7)$$

subject to initial and boundary conditions:

$$C(\mathbf{x},t)|_{t=0}=C_0,$$

$$C(\mathbf{x},t)|_{\Gamma_D}=C_D,\ [\theta_s\mathbf{D}\boldsymbol{\nabla}C]\cdot\mathbf{n}|_{\Gamma_N}=0,\text{ and }[-\mathbf{q}C+\theta_s\mathbf{D}\boldsymbol{\nabla}C]\cdot\mathbf{n}|_{\Gamma_C}=\mathbf{q}C_0 \qquad (8)$$

where in Eq. (1), $\mathbf{q} = -K(\mathbf{x})\boldsymbol{\nabla}h(\mathbf{x})$ is the flux (L T^{-1}) and θ_s is the effective porosity (dimensionless). C is the solute concentration (M L^{-3}), and Q_c is the rate at which solutes are injected (source) or extracted (sink) (M L^{-3} T^{-1}). **D** is the macrodispersion coefficient (L^2 T^{-1}) evaluated from velocity and dispersivities, as shown in Eq. (2). α_L and α_T are longitudinal and transverse dispersivity (L^2 L^{-1}), respectively. v_i and v_j are velocities (L T^{-1}) at different directions, and $|\mathbf{v}|$ is the magnitude of the velocity. D_0 is the effective molecular diffusion coefficient (L^2 T^{-1}), and δ_{ij} is the identity tensor. In Eq. (3), C_0 is the initial concentration in the entire system, C_D is the specified concentration at the Dirichlet boundary (Γ_D), no dispersive flux is applied at the Neumann boundary (Γ_N), and $\mathbf{q}C_0$ is the mass flux (M L^{-2} T^{-1}) at the Cauchy boundary (Γ_C).

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

6.2 Simulation Results

Figure 10 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 Figure 10 reveals that Case 3b provides the best prediction results (Figure 10d), yielding Cl plumes quite similar to the observed ones (Figure 10e) at all time stages. Without incorporating the stratigraphic information for inverse modeling, Case 1 and Case 3a fail to capture the migration of Cl (shown as Figures 10a and 10c, respectively), 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 (Figure 10b) 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.

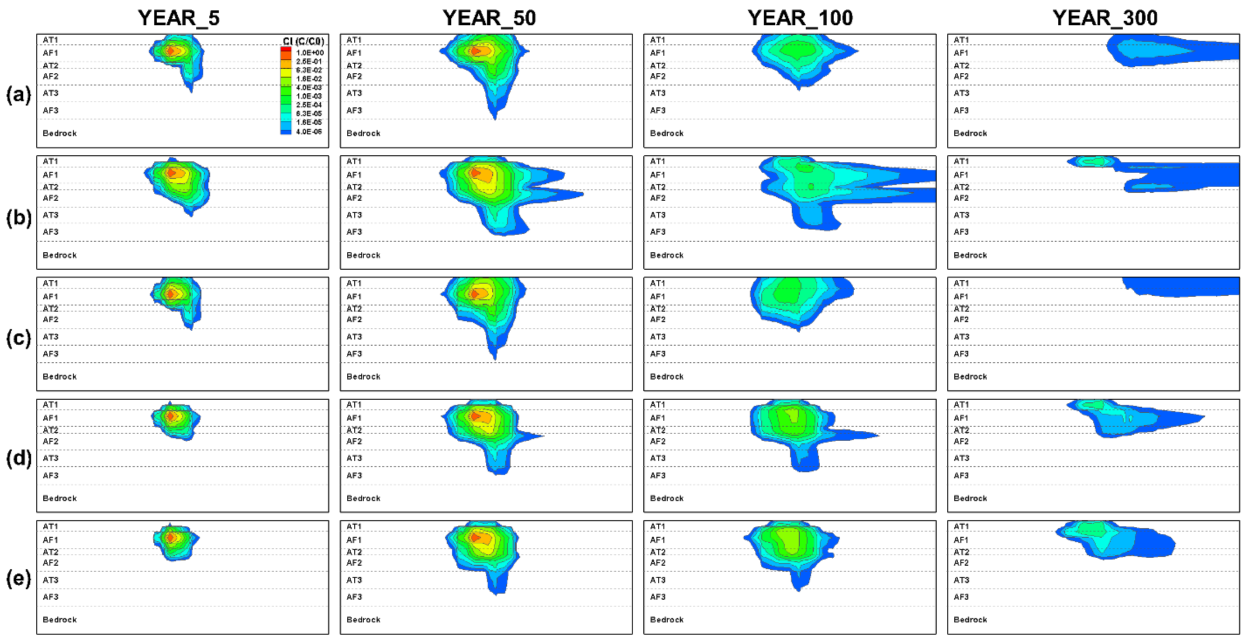

Figure 10: Simulated Cl plumes at four different time stages for four model cases and using the "true" K and S_s fields. a) Case 1: effective parameter model, b) Case 2: geological model, c) Case 3a: geostatistical model without geological information, d) Case 3b: geostatistical model with geological information, and e) "true" K and S_s fields.

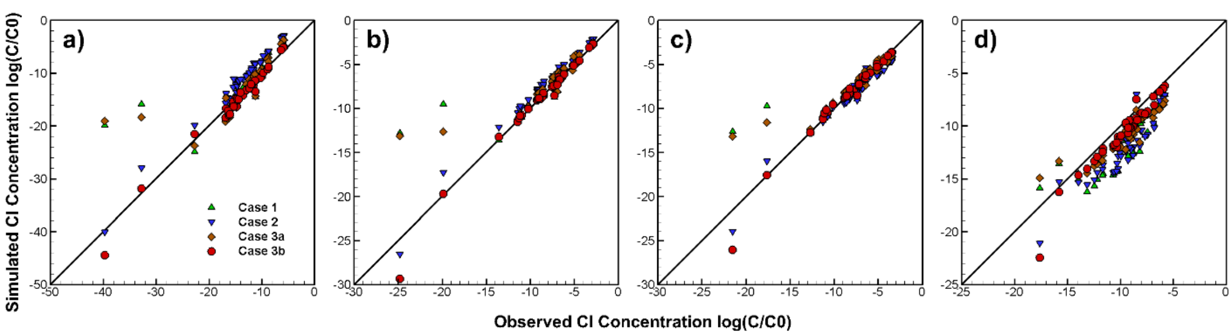

Figure 11: Scatterplots of simulated and observed Cl concentrations at all wells for four model cases at four time stages: a) Year 5 (early time); b) Year 50 (source removal); c) Year 100 (peak concentration arrival); d) Year 300 (late time).

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 Figure 11. Examination of Figure 11 reveals that prediction results for all model cases are improved from the early time stage to peak concentration arrival (Figures 11a through 11c) 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 enhanced again, resulting biased prediction results with enlarged scatters for all model cases at the late time stage (Figure 11d). 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.

6.3 Breakthrough Curves

Figure 12 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 in Figure S40 of the Supplementary Information section. In each plot, the “true” breakthrough curve is illustrated as the black dash line, while the simulated ones from different model cases are illustrated as solid lines with different colors.

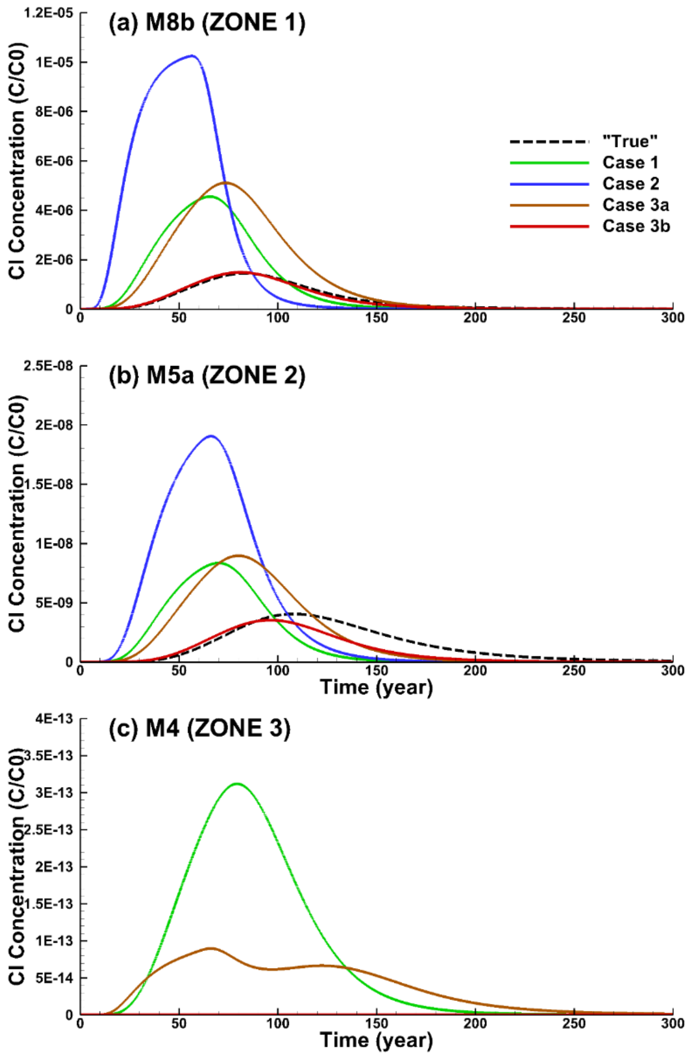

Figure 12: Simulated and observed breakthrough curves of Cl concentration at selected sampling locations (one for each subdivided zone) for four model cases.

As shown in Figure 12a, 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 points M8b be consistent with the “true” one. In contrast, the geological model (Case 2) yields quite poor prediction result, 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 Figure 12b. 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 Figure 12c. 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 performed to quantitatively compare the simulated breakthrough curves to the "true" ones.

6.4 Temporal Moment Analysis

Instead of characterizing the breakthrough curves at all wells, two sampling points (M4 and M8a) at the bottom layer (Bedrock) are excluded for temporal moment analysis since the Cl plume is simulated 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:

$$M_n = \int_0^{\infty} t^n C(x, y, z, t) dt \qquad (9)$$

where t is the time, and C is the Cl concentration. The zeroth (M_0), first (M_1), and second (M_2) 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):

$$\mu = \frac{M_1}{M_0} \qquad (10)$$

The variance σ^2 of breakthrough curves is calculated through

$$\sigma^2 = \frac{M_2}{M_0} - (\frac{M_1}{M_0})^2. \qquad (11)$$

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 are then compared, with the comparison scatterplots illustrated in Figure 13.

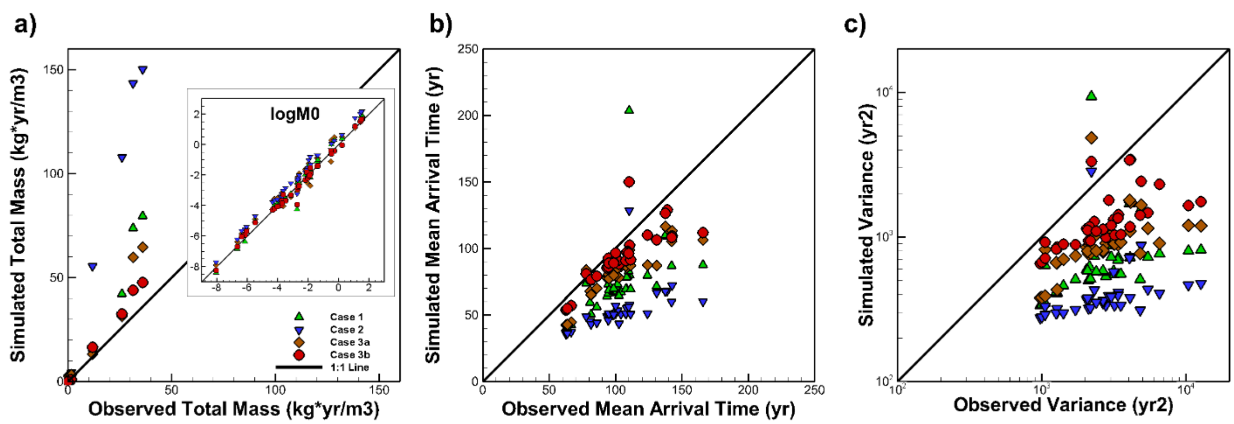

Figure 13: Temporal moment analysis of simulated versus observed breakthrough curves for four model cases. a) total mass (M_0), b) mean arrival time (μ), and c) variance (σ^2).

Figure 13a 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 Figure 13a. 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 Figure 13b. 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.

Figure 13c 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.

7. Conclusions

In this study, a synthetic 3D multi-aquifer/aquitard system that mimics the Mannheim East wellfield 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. Our 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 are accounted for during model calibration. In this study, pumping/injection records 180 days prior to the observation data were used; however, the appropriate duration is dependent on site specific conditions. 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 is able to provide 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.

The numerical experiments presented in this study provide a general framework for large-scale heterogeneity characterization using HT through the interpretation of long-term pumping/injection and water-level records. Groundwater variations in this study are considered to be induced by only pumping/injection operations, while ignoring the influence from other complexities (e.g., precipitation, evapotranspiration, surface water/groundwater interaction, long-term decline of groundwater levels due to dewatering operations, and etc.) which might be very important for long simulation periods. When applying this concept and technique to field problems with actual municipal well data, a more sophisticated groundwater model with well characterized sources of groundwater variations should be developed.

Acknowledgements

This research was supported by a grant from the Region of Waterloo to the University of Waterloo. Additional support for Ning Luo was provided by the Discovery grant from the Natural Sciences & Engineering Research Council of Canada (NSERC) awarded to Walter Illman. The authors thank Tammy Middleton from the Region of Waterloo for her assistance in accessing data used in this study. The data utilized in this paper can be found in the Supplementary Information section graphically and the corresponding values are available from the link 10.5281/zenodo.3723880.

References

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.

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

Bohling, G.C., Zhan, X., Butler, J.J., Zheng, L. (2002), Steady shape analysis of tomographic pumping tests for characterization of aquifer heterogeneities. *Water Resour. Res.*, 38(12). DOI:10.1029/2001wr001176

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

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

Doherty, J. (Ed.) (2005), PEST Model-Independent Parameter Estimation User Manual, fifth ed. Watermark Numer. Comput., Brisbane, Australia.

Gelhar, L.W. (1993), Stochastic subsurface hydrology. Englewood Cliffs, N.J.: Prentice-Hall.

Gottlieb, J. & Dietrich, P. (1995), Identification of the permeability distribution in soil by hydraulic tomography. *Inverse Probl*, 11: 353-360.

Illman, W.A., Liu, X., Craig, A. (2007), Steady-state hydraulic tomography in a laboratory aquifer with deterministic heterogeneity: Multi-method and multiscale validation of hydraulic conductivity tomograms. *J. Hydrol.*, 341(3–4): 222-234. DOI:http://dx.doi.org/10.1016/j.jhydrol.2007.05.011

Illman, W.A., Liu, X., Takeuchi, S., Yeh, T.-C.J., Ando, K., Saegusa, H. (2009), Hydraulic tomography in fractured granite: Mizunami Underground Research site, Japan. *Water Resour. Res.*, 45(1). DOI:10.1029/2007WR006715

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

Illman, W.A., Berg, S.J., Yeh, T.C.J. (2012), Comparison of approaches for predicting solute transport: Sandbox experiments. *Ground Water*, 50(3): 421-431.

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

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

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

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

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:https://doi.org/10.1016/j.jhydrol.2017.09.045

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

Niwa, M., Takeuchi, R., Onoe, H., Tsuyuguchi, K., Asamori, K., Umeda, K., Sugihara, K. (2012), Groundwater pressure changes in Central Japan induced by the 2011 off the Pacific coast of Tohoku Earthquake. *Geochemistry, Geophysics, Geosystems*, 13(5). DOI:10.1029/2012gc004052

Regional Municipality of Waterloo (2014), Region of Waterloo: hydrogeology & source water - WRAS database design manual. Regional Municipality of Waterloo, Waterloo, ON.

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

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.

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

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).

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.

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:https://doi.org/10.1016/j.jhydrol.2017.02.032

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

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.

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.

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:https://doi.org/10.1016/j.jhydrol.2019.03.089

Zhao, Z., Illman, W.A., Yeh, T.-C.J., Berg, S.J., Mao, D. (2015), Validation of hydraulic tomography in an unconfined aquifer: A controlled sandbox study. *Water Resour. Res.*, 51(6): 4137-4155. DOI:10.1002/2015WR016910

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

Zhao, Z. & Illman, W.A. (2017), On the Importance of Geological Data for Three-dimensional Steady-State Hydraulic Tomography Analysis at a Highly Heterogeneous Aquifer-Aquitard System. *J. Hydrol.*

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