PAPER

Hydraulic tomography analysis of municipal-well operation data with geology-based groundwater models

Xin Tong[1] · Walter A. Illman[1] · Steven J. Berg[1,2] · Ning Luo[1]

Abstract
The sustainable management of groundwater resources is essential to municipalities worldwide due to increasing water demand. Planning for the optimized use of groundwater resources requires reliable estimation of hydraulic parameters such as hydraulic conductivity (K) and specific storage (S_s). However, estimation of hydraulic parameters can be difficult with dedicated pumping tests while municipal wells are in operation. In this study, the K and S_s of a highly heterogeneous, multi-aquifer/aquitard system are estimated through the inverse modeling of water-level data from observation wells collected during municipal well operations. In particular, four different geological models are calibrated by coupling HydroGeoSphere (HGS) with the parameter estimation code PEST. The joint analysis of water-level records resulting from fluctuating pumping and injection operations amounts to a hydraulic tomography (HT) analysis. The four geological models are well calibrated and yield reliable estimates that are consistent with previously studies. Overall, this research reveals that: (1) the HT analysis of municipal well records is feasible and yields reliable K and S_s estimates for individual geological units where drawdown records are available; (2) these estimates are obtained at the scale of intended use, unlike small-scale estimates typically obtained through other characterization methods; (3) the HT analysis can be conducted using existing data, which leads to substantial cost savings; and (4) data collected during municipal well operations can be used in the development of new groundwater models or in the calibration of existing groundwater models, thus they are valuable and should be archived.

Keywords Hydraulic tomography · Conceptual models · Groundwater management · Model calibration and validation · Subsurface heterogeneity

✉ Xin Tong
   x23tong@gmail.com

   Walter A. Illman
   willman@uwaterloo.ca

   Steven J. Berg
   sberg@aquanty.com

   Ning Luo
   nluo1222@gmail.com

[1] Department of Earth & Environmental Sciences, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada

[2] Aquanty Inc, 564 Weber St N., Unit 2, Waterloo, Ontario N2L 5C6, Canada

Introduction

The Region of Waterloo (RoW) in Ontario (Fig. 1a,b) is one of the largest municipalities in Canada, which relies mostly (>75%) on groundwater supplies for its drinking water. The dramatic growth of the region and increasing water demand have prompted the development of municipal wellfields as well as the need to sustainably manage groundwater resources.

There are more than 40 wellfields, consisting of more than 120 water supply wells within the region, which supply in excess of 269,000 m³/day of groundwater to urban citizens (Lake Erie Region Source Protection Committee 2012). Groundwater is extracted from a complex multi-aquifer-aquitard system within the Waterloo Moraine (Fig. 1c), which was formed by interlobate glacial activity, with seven well fields pumping groundwater from the upper aquifer units, while ten additional well fields pump groundwater from the deeper units (Bajc and Shirota 2007).
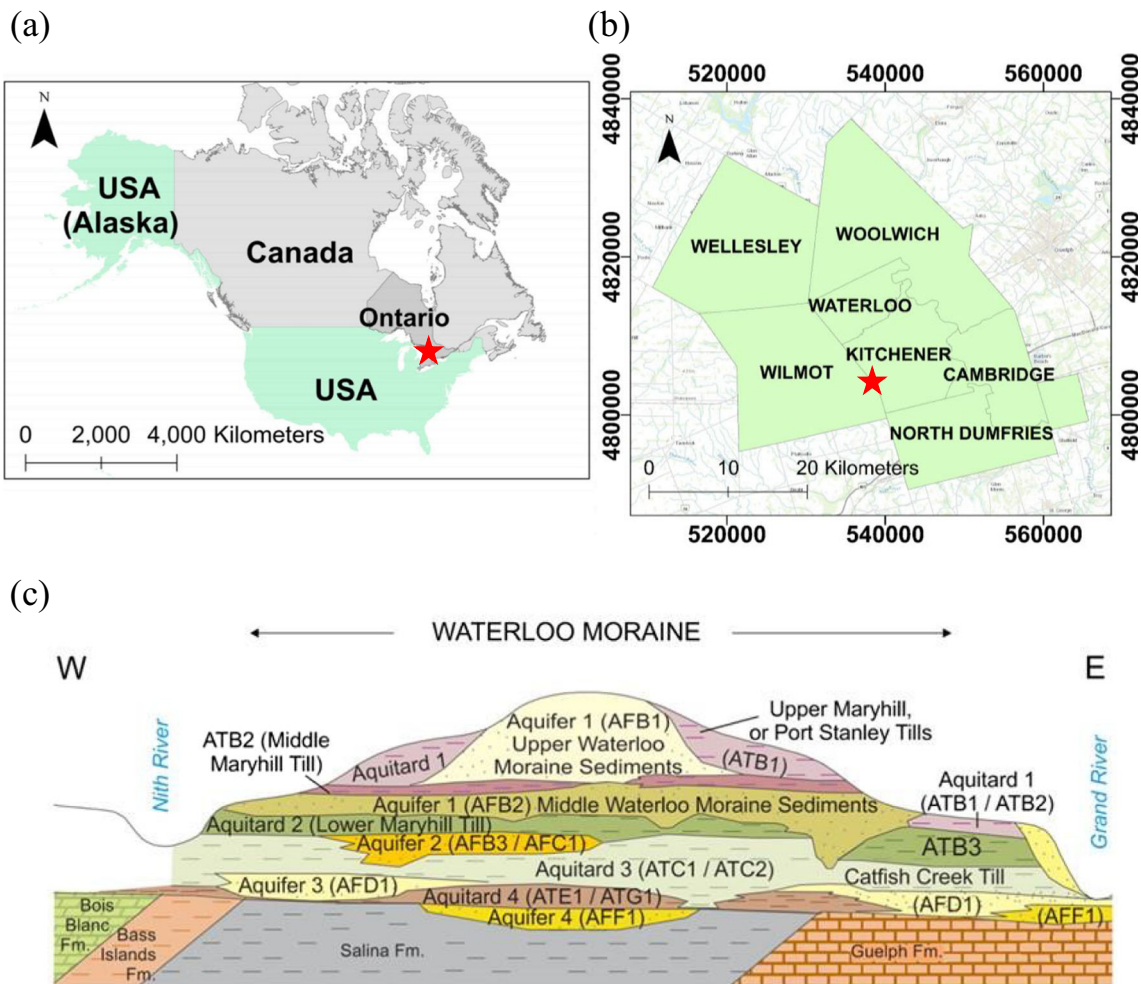
(a)

(b)

(c)

Fig. 1 **a** The location of the study area in relation to Canada and the US. **b** The location of the Mannheim East Well Field within the Region of Waterloo. **c** Conceptual hydrogeologic model of the Waterloo Moraine (after Blackport et al. 2014)

The complexity and susceptibility of the Waterloo Moraine to overexploitation of groundwater resources and its potential contamination requires a sound understanding of local hydrogeology, including the reliable estimation of hydraulic parameters such as hydraulic conductivity (K) and specific storage (S_s).

A number of hydraulic parameter estimation approaches have been developed and studied during the past several decades which include: (1) the analysis of small-scale data including grain size distributions (Hazen 1911; Kozeny 1927; Shepherd 1989), collection of core samples for laboratory permeameter analyses (Sudicky 1986; Sudicky et al. 2010; Alexander et al. 2011), slug tests (Hvorslev 1951; Bouwer and Rice 1976), flowmeter tests (Rehfeldt et al. 1992); and (2) performing pumping tests and fitting data to analytical solutions to determine the large-scale hydraulic properties of the aquifer (Theis 1935; Cooper and Jacob 1946). However, the large area of the municipal well field raises the question of whether small-scale hydraulic parameter estimates are suitable and reliable in predicting groundwater levels and flow. Another concern is that it is difficult to conduct dedicated

pumping tests within a municipal well field where pumping/ injection schedules cannot be readily modified or terminated. When dedicated pumping tests can be conducted, existing analytical solutions that treat the subsurface to be homogeneous are typically utilized, which can yield biased and questionable parameter estimates (Wu et al. 2005; Berg and Illman 2011a, b, 2013, 2015).

As part of their management and operation, most municipal well fields contain a network of observation wells to monitor the response of the aquifer system to pumping, and to inform optimal extraction rates. In addition to monitoring the response of the aquifer and informing pumping/injection rates, these monitoring data can potentially be used to better characterize regional groundwater flow and estimate hydraulic parameters (Yeh and Lee 2007; Harp and Vesselinov 2011).

In a previous study by Luo and Illman (2016), long-term pumping/injection events and water-level variation records were used to estimate hydraulic parameters including transmissivity (T) and storativity (S) for the shallow aquifer (AFB2) within the Waterloo Moraine. A set of T and S values

were estimated between each production and monitoring borehole by fingerprinting the water-level variations to pumping/injection rate changes. The fingerprinting process was accomplished through the Theis (1935) model implemented in the WELLS code (Mishra and Vesselinov 2011) coupled with a nonlinear parameter estimation code PEST (Doherty 2005). The Luo and Illman (2016) study showed that long-term municipal water-level records were amenable for hydraulic parameter estimation, as the geometric means of the individual T and S estimates were similar to previous pumping test results at the study site. However, the wide range of estimated T (9–55,335 m/day) and S (0.002–0.736) indicated the high degree of heterogeneity of the investigated aquifer. Furthermore, poor validation results using data that were not used for calibration purposes suggested that T and S estimates from individual pumping and monitoring boreholes may not be suitable for the drawdown prediction of other monitoring wells. In order to increase the accuracy of parameter estimates at this site, Luo and Illman (2016) concluded that a more sophisticated groundwater flow model that considers heterogeneity in K and S_s, as well as better accounting of the forcing functions (i.e., initial and boundary conditions as well as source/sink terms), is needed to obtain more accurate parameter estimates.

There are a number of approaches to map K heterogeneity. A conceptually simple approach is to map the K heterogeneity through the interpolation of small-scale K estimates including permeameter tests, slug tests and single-hole tests, but a large amount of data is required. For example, Rehfeldt et al. (1992) estimated that about 400,000 K measurements would be required to accurately predict the transport of tracers in an alluvial aquifer at the Columbus Air Force Base in Mississippi, USA, commonly known as the Macrodispersion Experiment (MADE) site. Thus, it would be prohibitively expensive and time-consuming to perform such analyses at a municipal wellfield.

An integrated stochastic-deterministic approach is another method to map K heterogeneity through calibrating a flow model, especially for large-scale systems (Traum et al. 2014; Sampath et al. 2015, 2016; Liao et al. 2020). Generally, the transition probability (TP) approach generates separate "TP zones" and three-dimensional (3D) material distribution based on stratigraphy information from borehole records to create conditional realizations (Carle and Fogg 1997). The realizations are then incorporated into the flow model and calibrated using long-term hydraulic head data. The accuracy of the TP approach is dependent on the density of the boreholes, while the pumping rate is usually simplified to annual/monthly pumping rates (Traum et al. 2014) or calibrated based on the well type (domestic use, public supply, etc.). Thus, such a method is more suitable for large-scale systems with sufficient borehole records and without precise consideration of municipal well operations. Berg and Illman (2015) also examined the TP approach and compared its performance to other approaches such as kriging, geological modeling, and

hydraulic tomography (HT) for capturing heterogeneity at a small-scale field site.

Hydraulic tomography has been developed to map subsurface heterogeneity through inverse modeling of hydraulic heads recorded at multiple locations from sequential pumping tests. Over the last several decades, it has been tested under synthetic (e.g., Yeh and Liu 2000; Bohling et al. 2002; Cardiff et al. 2013a, b), laboratory (e.g., Illman et al. 2007, 2008, 2009; Illman et al. 2010a, b; Illman et al. 2015; Berg and Illman 2011a; Zhao et al. 2016), and field conditions (e.g., Illman et al. 2009; Berg and Illman 2011b; Berg and Illman 2015; Zhao and Illman 2018; Zhao et al. 2019). HT relies on the joint inverse modeling of multiple sets of hydraulic heads from different observation intervals, while pumping and/or injecting water at multiple locations in geologic media. Due to the difficulty in performing sequential pumping tests at municipal wellfields, HT has not been applied previously at such sites.

Previous studies conducted in the study area have mostly relied on pumping tests to estimate hydraulic parameters of various units (e.g., CH2M HILL 2003a, b; Golder Associates 2011; Matrix Solutions Inc, S.S. Papadopulos and Associates 2014a, b), but such tests typically require shutting down the municipal wells and the interpretation is more difficult when the municipal wells are in operation. In order to overcome this difficulty and apply HT analysis without conducting sequential pumping tests, long-term pumping/injection events and water-level variation records obtained during municipal well field operation are used in this study to jointly calibrate a groundwater flow model consisting of homogeneous geological units to estimate the K and S_s. As an initial attempt, geological models are used for HT analysis in this study, as previous studies have shown the importance of geological data in parameter estimation when wells are far apart (Illman et al. 2015; Zhao et al. 2016; Luo et al. 2017; Zhao and Illman 2017, 2018). However, since perfect knowledge of stratigraphy is not available, there is a critical need to assess the impact of various conceptualizations of site geology on groundwater model calibration and HT at the field scale.

Geological uncertainty in groundwater modeling normally originates from: (1) the geological structure; (2) the use of effective model parameters; and (3) model parameters including local-scale heterogeneity (Refsgaard et al. 2012). Zhao et al. (2016) compared the performance of four geological models of different accuracies using laboratory sandbox data by model calibration and validation. Results revealed that it was possible to calibrate geological models both with and without accurate hydrostratigraphy because of the parametric compensation effect (Refsgaard et al. 2012). However, Zhao et al. (2016) found that the calibration of inaccurate geological models led to unrealistic parameter estimates in some geological units and poor model validation results; thus, while an inaccurate model can be well calibrated, this does not necessary result in a robust model that is suitable for making accurate predictions of groundwater flow.

The overall goal of this study is to examine the feasibility of conducting HT analysis of existing water level records influenced through municipal well field operations. Specifically, the objectives of the study are to: (1) demonstrate the usefulness of long-term pumping/injection and monitoring well records obtained through municipal well field operations for estimating hydraulic parameters (i.e., K and S_s) of geological units; (2) investigate the impact of different geological conceptualizations on the performance of groundwater model calibration and validation; and (3) explore the importance of geological data in improving the results of HT analysis at a large-scale field site.

Site description and geology

This study focuses on the Mannheim East Municipal Well Field (Fig. 1b) located in the southwest area of the city of Kitchener, Ontario, Canada. In order to minimize the effect of boundary conditions on simulated groundwater levels, the model is constructed for a larger area (5 km × 5 km) with the Mannheim East Well Field located approximately in the center of the simulation domain.

Information on more than 500 wells is recorded in the Region of Waterloo's WRAS+ database (Regional Municipality of Waterloo 2014) which contains various data sets such as well depths, screen intervals, static water levels, lithology, etc. There are 13 water-supply wells and 19 monitoring wells with 28 screens completed at various depths within the study area, with continuously monitored pumping/injection rates and water-level records. The Mannheim East Well Field (around 3.6 km²) is subdivided into three smaller well sites, which are identified as Mannheim East, Peaking, and Aquifer Storage and Recovery (ASR; shown in Fig. 2), while the distribution of water supply and monitoring wells utilized in this study is also shown in Fig. 2. A detailed description of these wells is provided in Table S1 of the electronic supplementary material (ESM).

The site sits on top of the Waterloo Moraine, which is a Quaternary kame and kettle complex formed by numerous advances and retreats of ice lobes during the Wisconsinan glaciation (shown in Fig. 1c), that has been studied extensively by Karrow (1993). The resulting glaciofluvial sediments consist of a variety of materials including clay, interbedded tills, fine sand, sandy gravel, and coarse gravel, which are normally stratified and poorly sorted (Martin and Frind 1998; Golder Associates 2011).

Four relatively continuous till units have been identified within the Waterloo Moraine, including Pre-Catfish Creek Tills, Catfish Creek Till, Maryhill Till and Tavistock/Pork Stanley Till. The Pre-Catfish Creek Tills, which include the first till units deposited in the area, are generally hard, stony silts to clayey silt tills (Karrow 1993). These till units,

including the Canning Till and several other tills, were formed during the Wisconsinan glacial events and locally overlie the bedrock (Martin and Frind 1998).

The Catfish Creek Till, which is the next oldest unit, was deposited by a major glacial advance across southern Ontario, is an extremely dense, stony silt till and commonly referred to as "hardpan" by local, experienced, water-well drillers (Golder Associates 2011).

The Maryhill Till, which separates the upper and deeper aquifers, is a clay-rich low-K natural infiltration barrier. Previous studies have identified three separate ice advances which resulted in the Upper, Middle and Lower Maryhill Till (Karrow 1993; Paloschi 1993; Bajc and Shirota 2007).

The youngest till units, within the Tavistock/Pork Stanley Till, overlie large portions of the upper aquifer. The Tavistock Till is a dark brown clayey silt till, similar to the Maryhill Till, while the Port Stanley Till is recognized as a sandy silt to silty sand till (Golder Associates 2011).

Description of geological models

Modeling can provide valuable insights on the Waterloo Moraine groundwater system and practical advice for source water protection and management for the Region of Waterloo. As models evolved from a simple, layer-cake concept to a fully 3D distribution of geological units, the focus has changed in scale from the well field scale to the scale of the entire Waterloo Moraine system to solve more sophisticated problems such as the assessment of well vulnerability and wellhead protection areas (Frind et al. 2014).

45 years of modeling the Waterloo Moraine

The first groundwater flow models of the Waterloo Moraine were developed as a simple, two-dimensional (2D), finite element, layer-cake system by Emil Frind in 1973 (International Water Supply Ltd. 1973). The model was calibrated to hydraulic head values at different observation wells and then used for the prediction of aquifer responses under various pumping conditions.

A quasi-3D model was then successfully developed and utilized by Rudolph (1985) and Rudolph and Sudicky (1990) at the Greenbrook well field to capture the complexity of the Waterloo Moraine system. The Waterloo North Aquifer System Study (Terraqua Investigations Ltd 1992) and the Study of the Hydrogeology of the Waterloo Moraine (Terraqua Investigations Ltd 1995) were conducted to define the major aquifer and aquitard units and regional recharge zones. Then, a fully 3D Waterloo Moraine model was created by Martin and Frind (1998) based on the application of WATFLOW (Molson et al. 1995). The groundwater model utilized triangular, prismatic, finite elements and allowed for
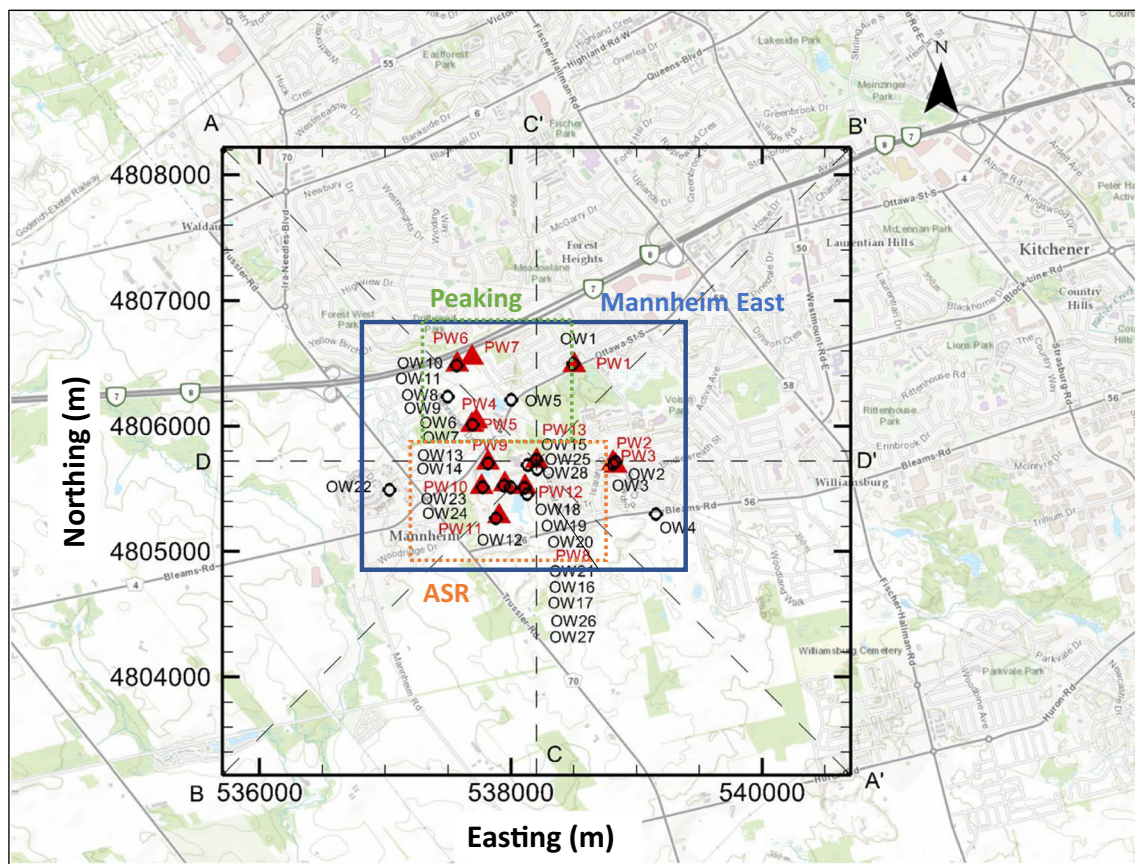
Fig. 2 The distribution of water-supply and monitoring boreholes in the study area. The red triangles indicate water-supply wells, rectangles indicate municipal well fields, and black circles indicate monitoring wells. Names of water supply and monitoring wells are modified for security purposes

grid refinement, which resulted in the better handling of complex geometries and representation of irregular and sloping layers (Callow 1996). The boundaries of the model were defined as natural features including rivers, creeks, and swamps.

Bajc and Shirota (2007) constructed a new geological model of the Waterloo Moraine, applying a basin analysis approach to data collection and interpretation, which provided details to various geological units, including information on the distribution, thickness, geometry and other attributes. The model was built mainly based on geological information and subsurface sediment structure including geological data from a regional borehole database (Farvolden et al. 1987; Bajc and Newton 2007), published information on Quaternary geology, downhole geophysical logs, and identification of available sediment exposures. Since hydrogeological data including hydraulic head and hydraulic test observation data were not used in model layer interpretation, the model layers were considered stratigraphic layers, which may not be consistent with hydrogeological data at each well field (Blackport et al. 2014). Refinements to this model were made within various municipal well fields through subsequent studies (Stantec Consulting Ltd. 2009, 2012a, b, c; Golder Associates 2011;

Blackport Hydrogeology Inc. 2012a, b; Matrix Solutions Inc, S.S. Papadopulos and Associates 2014a, b).

Development of a new geological model

In this study, a new geological model was constructed based on the lithology of wells installed within the study area using Leapfrog Geo (ARANZ Geo Ltd. 2015). Leapfrog Geo constructs 3D geological models using borehole records and GIS data based on the Fast Radial Basis Function (RBF) method. RBF is an interpolation method first used by Hardy (1971) to interpolate scattered topographic data. The smooth interpolation surface is created by Leapfrog Geo using available data points based on weighted linear combinations of covariance functions, which is widely used in geological modelling (Carr et al. 2001; Frank et al. 2007). In total, lithology information from 250 wells were utilized for the construction of the new geological model. The distribution of these borehole records at the study site is illustrated in Fig. 3. For each borehole record, lithology information was obtained from the RoW's WRAS+ database and summarized based on the three main materials identified for each core sample. The topography of
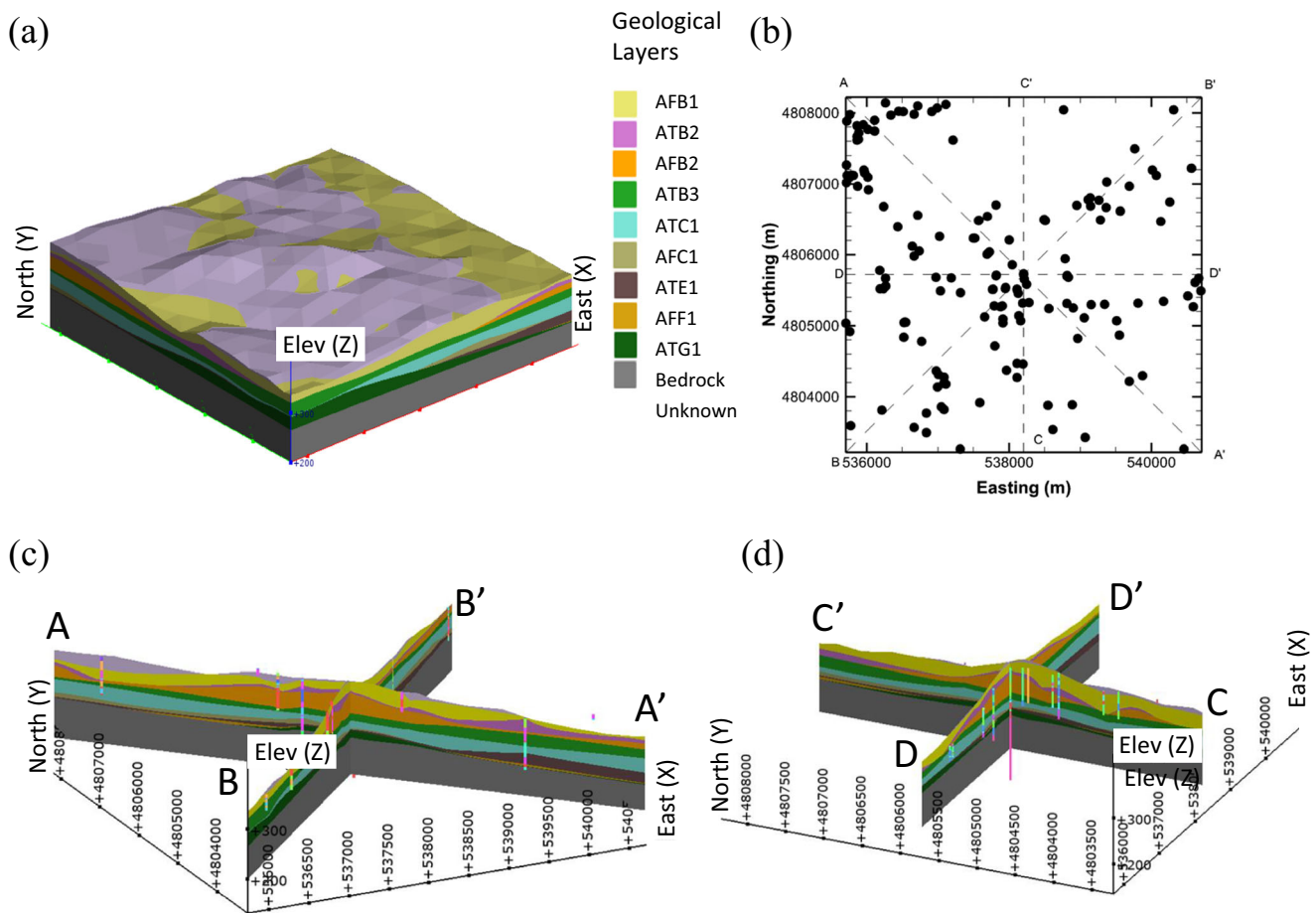
(a)

Geological Layers

	AFB1
	ATB2
	AFB2
	ATB3
	ATC1
	AFC1
	ATE1
	AFF1
	ATG1
	Bedrock
	Unknown

(b)

(c)

(d)

Fig. 3 a 3D geological model of the site constructed using Leapfrog Geo. **b** Distribution of selected wells within the study area along with locations where cross-sections are provided. **c** Cross-sections along A–A′ and B–B′. **d** Cross-sections along C–C′ and D–D′

the geological model is imported into Leapfrog Geo using the digital elevation model (DEM) data (30-m resolution) from Ontario Ministry of Natural Resources and Forestry. Leapfrog Geo creates a smooth interpolation surface between two adjacent geological layers based on lithology data from boreholes. In total, 11 groups of geological units are identified based on the conceptual hydrogeologic model of the Waterloo Moraine constructed by Bajc and Shirota (2007) and Matrix Solutions Inc, S.S. Papadopulos and Associates (2014a, b). The nomenclature of the Ontario Geological Survey (OGS) is adopted here for layer identification, in which AT refers to an aquitard, while AF refers to an aquifer. Following AT or AF, letters and numbers are used to identify the sequence of units, with "A" as the youngest grouped sequence followed by "B", and "1" as the youngest unit in the group followed by "2"—for example, ATB1 refers to the youngest Aquitard in the B sequence, whereas AFF1 refers to an older Aquifer of the F sequence.

Figure 3 shows the resulting 3D geological model with four cross-sections. The dimensions of the geological model are 5 km × 5 km in X (east) and Y (north) directions with an elevation of 200 masl (meters above sea level) as the bottom

and the topography as the top. The bottom of the model is set at 200 masl because no data are available below 200 masl, while Aquitard ATE1 and ATG1 are regionally extensive. It is assumed that the well field is not affected by deeper groundwater flow in the bedrock since there are two extensive aquitards (ATB3 and ATE1/ATG1) covering nearly the entire study area. However, due to lack of data in the deeper units, further studies are necessary to test this assumption. In total, 11 geological layers were identified, listed from youngest to oldest, these units are: ATB1, AFB1, ATB2, AFB2, ATB3, ATC1, AFC1, ATE1, AFF1, ATG1 and Bedrock.

In comparison to the conceptual hydrogeologic model (Fig. 1c) of the Waterloo Moraine, some layers were merged (ATC1 and ATC2 were combined as ATC1 and AFF1 and AFD1 were combined as AFF1) in the newly constructed geological model. This is because: (1) these geological layers are thin and consist of similar materials, and (2) they are located at low elevations where geological data from borehole logs are limited in order for one to accurately separate these layers.

Examination of Figs. 1c and 3 reveals that ATB1 is a thin and patchy aquitard that lies on top of the study area, while

AFB1 is an unconfined aquifer present throughout the study area with considerable recharge from precipitation that appears to take place in the central and eastern portions of the site. ATB2 is a thin aquitard that separates AFB2 and AFB1 in most of the study area. AFB2 is the primary water-supply aquifer in the Mannheim East wellfield. The AFB2 aquifer is evident in the central area of the study site, where the municipal well field was developed with a maximum thickness of approximately 40 m; however, the thickness decreases as it extends to the edges of the geological model. Beneath the AFB2 aquifer, the ATB3 aquitard is continuous across the study area followed by the aquitard ATC1. These two aquitards with extremely low K separate the upper aquifers (AFB1 and AFB2) from the lower aquifer/aquitard system. Between the ATC1 aquitard and the bedrock, four geological layers (AFC1, ATE1, AFF1, and ATG1) have been further identified. These layers are found to be thin and discontinuous within the study area.

Description of geological models

In total, four geological models are utilized in this study including: (1) the 5-layer model; (2) the 11-layer model; (3) the Waterloo model; and (4) the Regional model for model calibration as well as model validation. Detailed layer information of each model is provided in Table S2 of the ESM and cross-sections of each model with screen information are shown in Fig. 4. The 11-layer geological model, the Waterloo model and the Regional model all divide the study domain into 11-layers, with main differences in the layer thicknesses of the upper aquifer/aquitard units and layer classification for the lower aquifer/aquitard units. Groundwater modelers typically fixate on a single conceptual model, but in this study, four geological models will be utilized for model calibration and validation. The results from the four models are compared and discussed next.

Model 1. 5-layer geological model A simplified geological model has been developed by merging some of the layers with similar material, specifically ATB1 as AT1, AFB1, ATB2 and AFB2 as AF1, ATB3 and ATC1 as AT2, AFC1, ATE1, and AFF1 as AF2, ATG1 and bedrock as AT3 (shown in Table S2 of the ESM).

Model 2. 11-layer geological model Detailed information of the 11-layer geological model is explained in the previous section. While the five-layer geological model mainly reflects the contrast in low and high K zones, the 11-layer geological model incorporates more detailed stratigraphy information resulting in higher-resolution representation of local heterogeneity and hydraulic connectivity compared with other large-scale models treated in this study.

Model 3. Waterloo and Model 4. Regional models Two additional models are used in the study for groundwater flow model calibration and validation, including the Waterloo model, built by Bajc and Shirota (2007), and the Regional model (Matrix Solutions Inc, S.S. Papadopulos and Associates 2014a, b), refined based on the Waterloo model. The Waterloo model is constructed based on subsurface information from RoW monitoring wells, urban geological database, field mapping data, cored boreholes, Ministry of the Environment, Conservation, and Parks (MECP) water well records and geophysical databases, while refinements have been made to the Regional model with available hydrogeological data including municipal pumping data, hydraulic head data, water quality data, isotopic data, and wellfield shutdown data (Blackport et al. 2014).

Data used for groundwater flow model calibration and validation

In this study, the dataset used by Luo and Illman (2016) was first studied and a decision was made to utilize a shorter record for model calibration. In particular, the pumping/injection rate records in water supply wells from 1 January 2005 to 31 December 2013 were utilized by Luo and Illman (2016), while this study utilized 1-year data from the year of 2013 to achieve computational efficiency. The importance of the length and number of data from observation records to include in inverse models is discussed by Luo et al. (2020) using a synthetic model based on five different simulation durations. Their results reveal that periods with large water-level variations and continuous data points need to be included to better interpret municipal well data for site heterogeneity characterization. It is also found that longer simulation durations do not always yield better results since the pumping/injection influence from the water-supply wells may propagate beyond the investigated area.

The pumping/injection rate records from 13 water-supply wells (shown in Fig. 5) and water-level records from 19 monitoring locations with 28 screens at different times during the year 2013 are obtained from the WRAS+ database for groundwater model flow calibration. Due to the resolution of the pumping data in the database, pumping/injection rates in water-supply wells are expressed as daily pumped volume in m^3. In reality, pumping/injection events normally operate for a couple of hours throughout a single day. However, the accurate operational time is not provided, thus the pumping/injection rates are simplified as daily pumping/injection rates. Therefore, for each water-supply well, 365 records are extracted from the database within the selected period. It should be noted that pumping/injection rates in these water-supply wells are not constant; instead, they vary frequently in most wells (shown in Fig. 5). Typically, municipal wells are used as
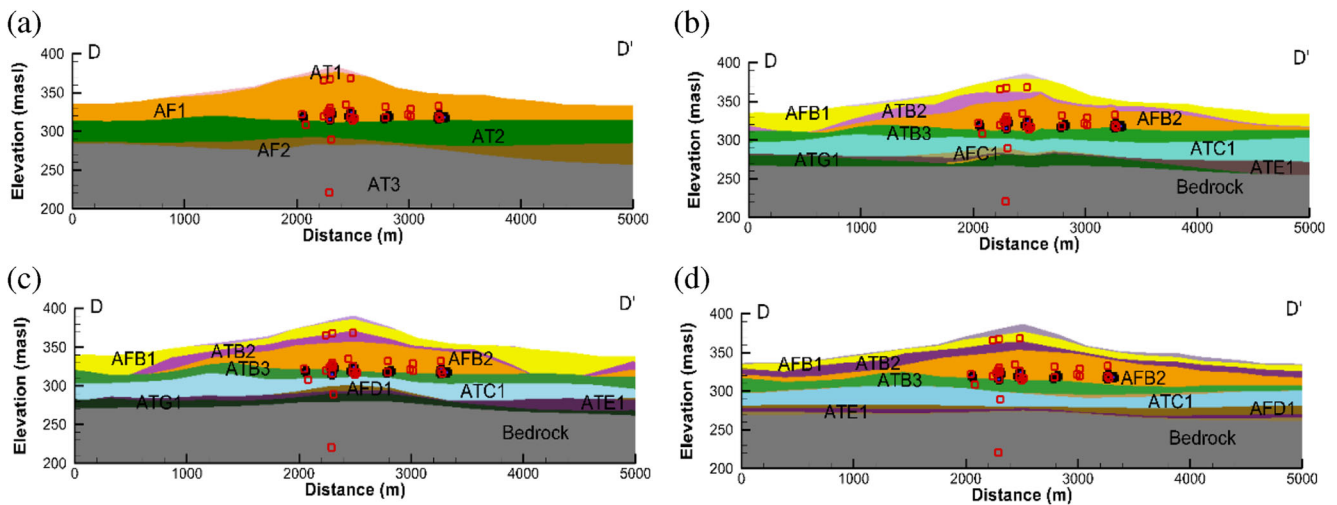
Fig. 4 Cross-sections along D-D′ for: **a** the 5-layer geological model; **b** the 11-layer geological model; **c** the Waterloo model; and **d** the Regional model with screen midpoint elevation information. The black circles indicate water supply wells, while red squares indicate monitoring boreholes

water-supply boreholes, while some ASR wells are designed to inject and store treated water from the Grand River during low water demand periods, and the stored water is extracted during high demand periods. Thus, negative pumping rates shown in Fig. 5 indicate injection.

The screen midpoints of all water-supply wells are located between 315 to 325 masl, while the screen midpoints of the water-monitoring boreholes are located between 185.82 and 368.88 masl. The screens of both water-supply and water-monitoring wells are mainly located at the bottom of AFB2, with few wells installed within the AFB1, AFC1 and Bedrock units based on the 11-layer geological model.

Although the depths of the screened intervals vary widely, well screens installed in the AFB1, AFC1 and bedrock units lack constant monitoring records, based on Table S1 of the ESM, with less than 40 data points available throughout 2013. In addition, there are a large number of water level measurements in many wells, but in some wells, the monitoring record is quite sparse. Most wells are located at the bottom of AFB2 for both the 11-layer geological model and the Regional model, but at the

bottom of AFB2/upper and middle of ATB3 for the Waterloo model. The location of OW16 is at AFC1 for the 11-layer geological model, at ATG1 for the Waterloo model, and ATC1 for the Regional model.

Water levels in water-supply and monitoring wells are measured manually and electronically with pressure transducers. The transducers automatically record the water level every hour, thus the data recorded at the beginning of each day (12:00 am) are used as the water level for each day, thus 365 records are used for 2013. At some wells, water levels are recorded monthly or bi-monthly through manual measurements, so the available data with manual measurements range from 3 to 12 in 2013. The flow rates in all water-supply wells are electronically recorded, thus 365 data points are available for 2013. Table S1 in the ESM summarizes the number of data points used from each well for model calibration as well as model validation.

In this study, drawdown is defined as the hydraulic head change from the initial head and is expressed as:

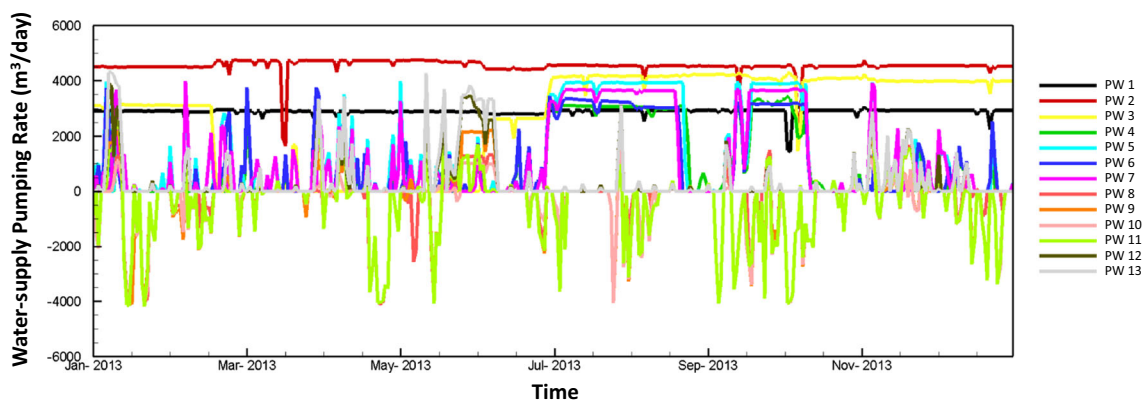$$\Delta h(x, y, z, t) = h_0(x, y, z) - h(x, y, z, t) \tag{1}$$

Fig. 5 Pumping or injection rates in water supply wells during the year 2013. Negative pumping rates indicate injection

where $\Delta h(x, y, z, t)$ is drawdown at some point in the simulation domain defined by the Cartesian coordinate system (x, y, z) at time t due to municipal well operations and $h_0(x, y, z)$ is the initial hydraulic head when t is equal to 0, while $h(x, y, z, t)$ is hydraulic head at time t.

It is important to keep in mind that since groundwater is constantly pumped or injected from water-supply wells, the initial static water level is unknown for each screen. Based on the comparison of simulated and measured drawdown using empirical hydraulic parameters (Martin and Frind 1998) for various geological models during the forward simulation process, the simulated drawdown curves fit well with the measured drawdown data after 20 days for most of the monitoring wells. Therefore, the first 20 data points from monitoring wells are not used for model calibration and validation in order to simulate water level fluctuation as a result of various pumping/injection events. The same strategy was applied by Luo and Illman (2016), which provided optimal matching between simulated and observed data. In total, 4,985 data points are used for model calibration and 5,085 data points are used for model validation (see Table S1 of the ESM) in this study. Data from January to June 2013 are used for model calibration, while data from July to December 2013 are used for model validation.

Description of groundwater models

The groundwater flow model has the same dimension as the geological model. Prior to constructing the 3D groundwater flow model, a 2D grid with 3210 nodes was generated based on the plan view of the simulation domain, as shown in Fig. 6a. Triangular elements with a size of 200 m were applied to discretize the simulation domain. At locations where there are water-supply and monitoring wells, the grid was refined by a factor of five. The element size is determined based on the geological model resolution and computational efficiency. The DEM data are 30-m resolution and the geological model (5 km × 5 km) is built based on the lithology information from 250 wells. As shown later, simulated drawdowns generally capture the observed drawdown curves, so the current model discretization sufficiently captures the observed records.

Layer information identified in the constructed geological model was then introduced to generate the 3D groundwater flow model, as shown in Fig. 6b–d for each geological model. Each geological layer was subdivided into several layers based on the approximate thickness as well as the distance to the pumping wells. In particular, fine grids were assigned to the layer of the water-supply aquifer, while coarse grids were assigned in upper and lower layers. In total, the 3D hydrogeologic model was discretized into 30 layers for both 5- and 11-layer geological models with 188,460 computational elements and 99,510 nodes, 27 layers for the Waterloo

model with 144,482 elements and 76,842 nodes, and 27 layers for the Regional model with 139,412 elements and 74,196 nodes. All four geological models were discretized using the same 2D grid with uniform and isotropic K values of the elements located in the same layer.

All groundwater flow simulations were conducted using the groundwater flow and transport simulator HydroGeoSphere (HGS; Aquanty Inc. 2018) coupled with the model-independent parameter estimation code PEST (Doherty 2005). PEST is unique in that it wraps around any model allowing for automatic calibration using a variant of the Levenberg-Marquardt algorithm to minimize the objective function $\varnothing$, which represents a weighted sum of squared differences between computed and measured calibration targets. Here, the objective function is expressed as:

$$\varnothing(\mathbf{a}) = \left[\mathbf{h}^*(x, y, z, t) - \mathbf{h}(x, y, z, t; \mathbf{a})\right]^{\mathrm{T}} \quad (2)$$
$$\mathbf{W}\left[\mathbf{h}^*(x, y, z, t) - \mathbf{h}(x, y, z, t; \mathbf{a})\right]$$

where $\mathbf{a}$ is a vector of M model parameters to be optimized and $\mathbf{h}$ and $\mathbf{h}^*$ are vectors of simulated and measured hydraulic head values respectively, at N match points in space-time, and $\mathbf{W}$ is an $N \times N$ diagonal weight matrix. Equation (1) is implemented using unit weights, which makes it equivalent to ordinary least squares. The parameters of interest for this paper are K and S_s.

There are five cases considered in the study, including: (1) the five-layer model with uniform initial hydraulic parameters for all layers; (2) the five-layer model with different hydraulic parameters for each layer; (3) the 11-layer model with different hydraulic parameters for each layer; (4) the Waterloo model with different hydraulic parameters for each layer; and (5) the Regional model with different hydraulic parameters.

In case 1 (the five-layer uni model), the initial hydraulic parameters for all layers are set to a uniform value. Specifically, the initial K value for calibrating the five-layer geological model was set as 6.00×10^{-5} m/s with a minimum bound of 1.00×10^{-9} m/s and a maximum bound of 0.01 m/s. Likewise, the initial S_s value was set as 0.0006 m^{-1} with a minimum bound of 1.0×10^{-8} m^{-1} and a maximum bound of 0.1 m^{-1}. Since most observation points are located in AFB2, it is essential to set appropriate initial K and S_s values for the other layers in order to increase the computational efficiency and reliability of results. Doherty (2005) stated that the properly selected initial parameter values will not only increase the optimization efficiency, but also transfer a highly nonlinear model to a reasonably linear model through parameter transformation. In particular, the predominant materials for each layer were identified and used to assign initial K and S_s values, as shown in Table S2 of the ESM. The corresponding K and S_s values for each material were based on Martin and Frind (1998), provided as Table S4 of the ESM. The representative values were identified by Martin and Frind (1998) from the
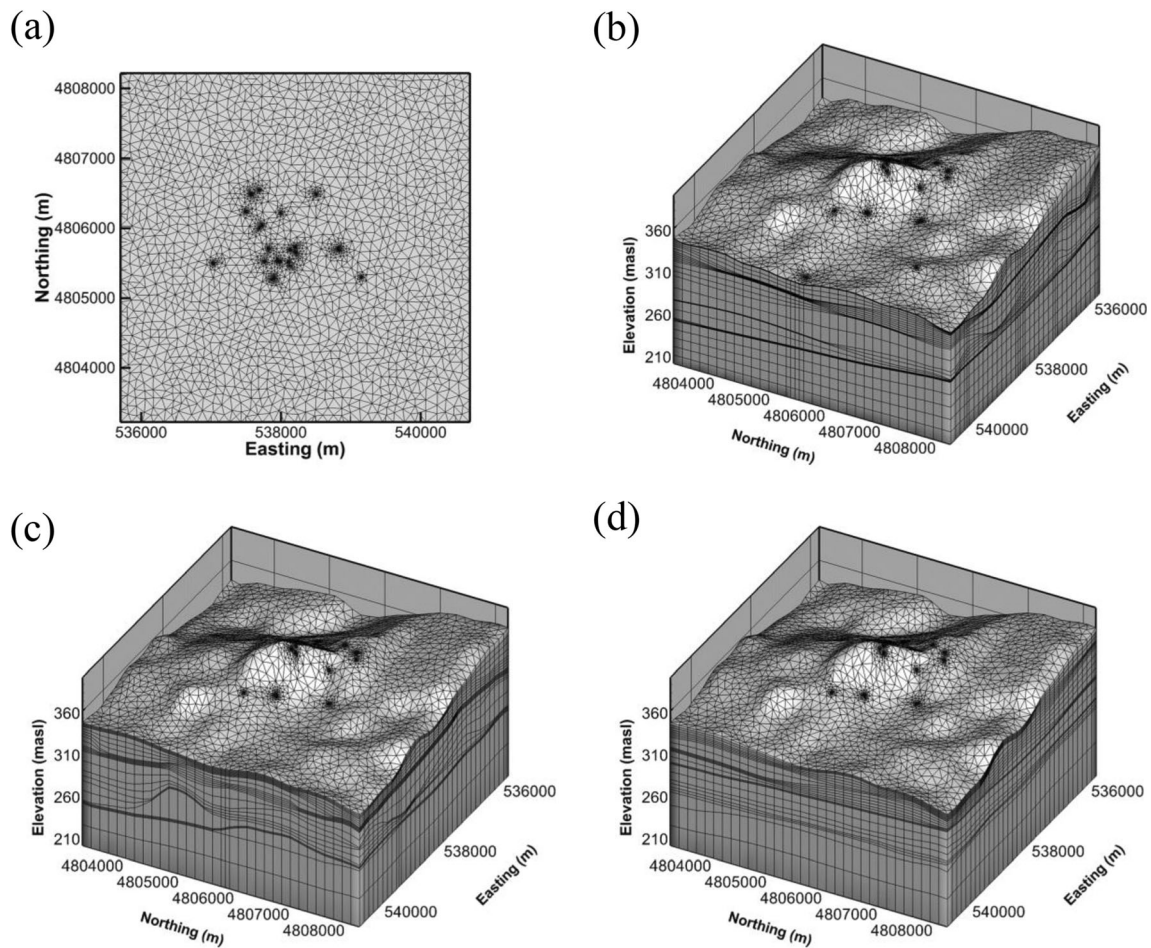
Fig. 6 **a** Generated two-dimensional grid of the study area (plan view). Generated three-dimensional grids for: **b** the 11-layer geological model; **c** the Waterloo model; and **d** the Regional model

literature and also calibrated based on previous pumping and slug test results at the same wellfield site. Since there are several water-supply wells located within the ATB3 unit of the Waterloo model, and in order to increase the computational efficiency, the initial K and S_s values were set the same as AFB2 of the Waterloo model.

Four geological models with appropriate initial K and S_s values for each layer were calibrated and validated as cases 2–5. The minimum and maximum bounds of the estimated parameters in cases 2–5 are set the same as in case 1.

In terms of boundary conditions, the bottom face was defined as a no-flow boundary, since the bottom layer of all four models is bedrock, which underlies the aquitard layer ATG1 for both 11-layer geological model and the Waterloo model. The four side faces were set as constant head boundaries, implying that the hydraulic heads on the boundary faces were not affected by pumping/ injection events. This was examined by plotting the simulated drawdown distribution at the end of the pumping/injection records (see Fig. S3 of the ESM).

The static water level in the monitoring boreholes at the beginning of pumping records from 284 wells (mainly located at AFB2) within the study area was selected and used for kriging of hydraulic head with Tecplot (2011). The resulting hydraulic heads were generally higher at the northwest portion of the study area and lower at the southeast part, indicating that groundwater flows from the northeast to southwest, which is consistent with historical records.

The hydraulic head ranges from 306.28 to 359.25 masl within the study area and were used as initial head values for the simulation domain and constant boundary head for the four boundary faces. It is noted that the hydraulic heads along the vertical direction on the side boundaries were set to be the same since there was no available hydraulic head data for lower aquifers in the study area.

In order to set the boundary condition for the top of the simulation domain, daily precipitation data from 2013 were obtained from the weather station located on the University of Waterloo campus that is 9 km away from the study site. These data were modified as net precipitation (45% of the total

precipitation for each day) and used as nodal fluxes to define the boundary condition at the top face. It was assumed that the effect of evapotranspiration (ET) was constant through the 6-month simulation period. The current study focused on groundwater flow in the phreatic zone and neglected variably saturated flow in the vadose zone. This is due to the fact that most data are collected from AFB2, a geological unit that is 20–50 m from the ground surface and overlain by the topmost aquitard layers ATB1 and ATB2, both of which are aquitards that limit vertical groundwater flow, thus the influence of variably saturated flow can be safely neglected. In addition, transient effects in recharge were not considered in this study because ATB1 and ATB2 have very low K values and the aquitard ATB2 is quite extensive in the study area (Matrix Solutions Inc, S.S. Papadopulos and Associates 2014a, b) which dampens the transient effect in the underlying aquifer.

Guo (2017) studied the relationship between precipitation and ET during the years of 2010, 2011, 2012 and 2014 at the Laurel Creek Watershed near the study area, and found that the average simulated and measured annual ET accounted for 56.5 and 54.3% of the annual rainfall, respectively. Thus, 45% of the total precipitation was used in this study as the net precipitation for each day during the simulation period considering the proximity of the study area to the Laurel Creek Watershed.

Results and discussion

Model calibration results

Inverse modeling of pumping/injection records from 13 water-supply wells was performed on a PC with a six-core CPU and 16 GB of random access memory (RAM) for model calibration. All inverse models for the five cases ran until the convergence criteria were met for either a maximum number of iterations or an observation of no significant improvement between simulated and observed head. The improvement between simulated and observed head was evaluated using: (1) the difference between the current objective function and the lowest objective function achieved to date, and (2) the magnitude of the maximum relative parameter change between optimization iterations.

Calibration of the five-layer geological model took about 24 h to estimate 10 unknowns with 361 model calls for cases 1 and 2, while the calibrations of other three geological models were all completed within 72 h to estimate 22 unknowns with total model calls ranging from 627 for the Regional model (case 5) to 849 for the 11-layer geological model (case 3).

Although cases 3–5 have the same number of unknowns with the same amount of data, the number of model calls varied from case to case. This is due to the fact that appropriate initial parameter values are critical to PEST, since the closer are the initial parameter values to optimal, the faster PEST will converge. They also make optimization possible, especially for highly nonlinear models. Thus, the different number of model calls among cases may be due to the difference between the initial and the optimal parameter values and the optimization process (Doherty 2005).

The estimated K and S_s values for each geological layer are plotted in Figs. 7 and 8, respectively. The estimated K and S_s values and their 95% confidence intervals are summarized in Fig. 9 with numerical values provided in Table S3 of the ESM.

The estimated K and S_S fields shown in Figs. 7 and 8 and corresponding K and S_s values for each geological layer shown in Fig. 9 reveal large differences among four models which highlight the importance of the geological information on groundwater modeling results. In particular, utilizing different geological conceptualizations for inverse modeling with the same data results in dramatic differences in estimated K and S_s values (e.g., K value for ATB3 ranges from 7.89×10^{-12} to 4.71×10^{-4} m/s among cases 3, 4 and 5). The estimated K and S_s values in case 1 are less realistic compared with the ones for the same geological layer in other cases, which highlights the importance of the initial parameter values to the model calibration.

As mentioned previously, a highly nonlinear model may be hard to optimize without properly selected initial parameter values (Doherty 2005). Doherty (2005) also stated that observations can be insensitive to initial parameter values if initial parameter values are not chosen wisely—for example, a parameter may have little effect on model results over a part of its domain, while it may have a much larger effect over other parts of the domain. Thus, the initial parameter values for case 1 may not be within the sensitive part of its domain to yield reasonable results. It was also noticed that the corresponding 95% confidence intervals are extremely large, especially for the lower geological layers. The extremely large 95% confidence intervals may be due to merging layers of different K values and also is a result of insufficient observation points, both of which have been suggested by Zhao and Illman (2018) in relation to a different site.

The remaining four geological models with appropriate initial K and S_s values (cases 2–5) are all well calibrated with more realistic estimations and much smaller 95% confidence intervals compared with case 1. Thus, the reliability of the estimated hydraulic parameters can be greatly increased by using appropriate initial K and S_s values as prior information. As previously noted, most data points are collected from the monitoring boreholes located at the bottom of AFB2, so similar K and S_s values were obtained from model calibration for the 11-layer geological model (case 3) and the Regional model (case 5).
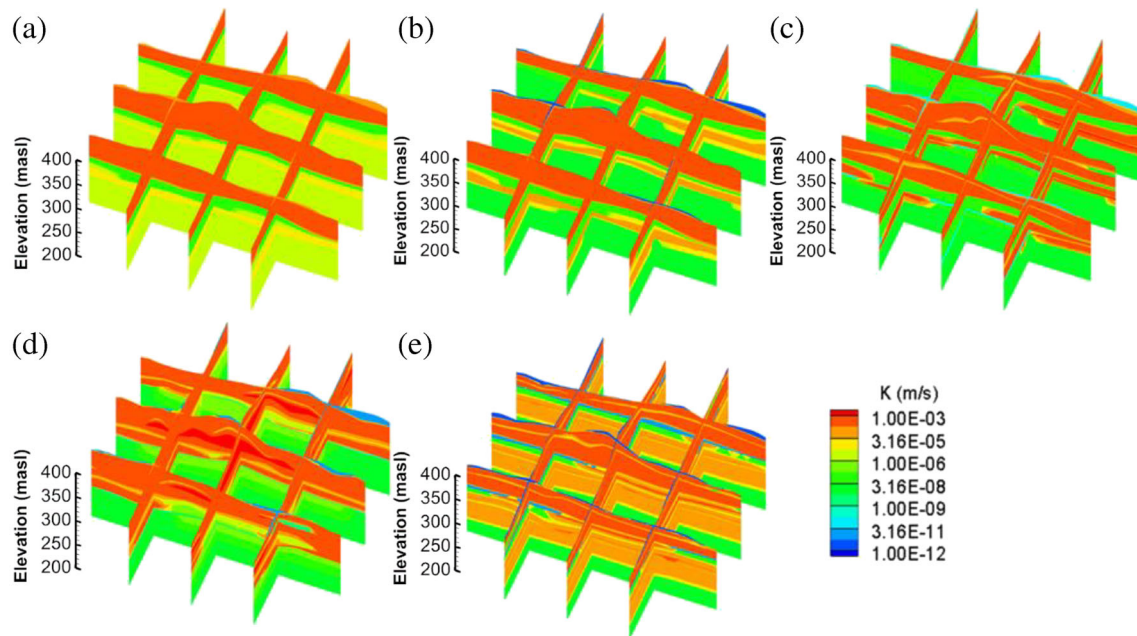
Fig. 7 Estimated K fields from the inversion of pumping and injection events by 13 water-supply wells during January to June, 2013 for: **a** the 5-layer uni model; **b** the 5-layer geological model; **c** the 11-layer geological model; **d** the Waterloo model; and **e** the Regional model

The K estimated for AFB2 is 8.14×10^{-4} m/s with the 11-layer geological model (case 3), while a value of 5.17×10^{-4} m/s is estimated for the 5-layer geological model (case 2), which is a consequence of using one layer to represent multiple soil types. The estimated K of AFB2 (1.21×10^{-3} m/s), ATB3 (4.71×10^{-4} m/s), and ATC1 (2.12×10^{-6} m/s) for the Waterloo model (case 4) is

relatively high, which may be due to the inaccurate classi-fication of the geological layers. Since most of the screens are located within AFB2 and ATB3, the large estimated value of K for the aquitard enables the groundwater to be pumped from it and match the corresponding observed drawdowns at monitoring wells. This is also one drawback of the geological approach in groundwater modeling when
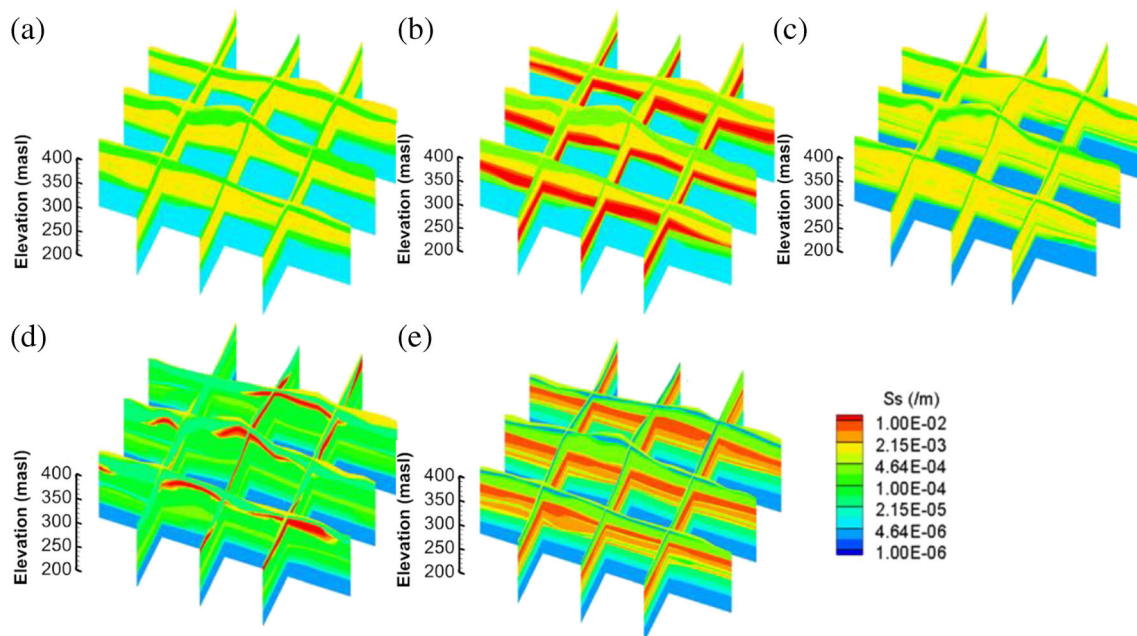

Fig. 8 Estimated S_s fields from the inversion of pumping and injection events by 13 water-supply wells during January to June, 2013 for: **a** the 5-layer uni model; **b** the 5-layer geological model; **c** the 11-layer geological model; **d** the Waterloo model; and **e** the Regional model
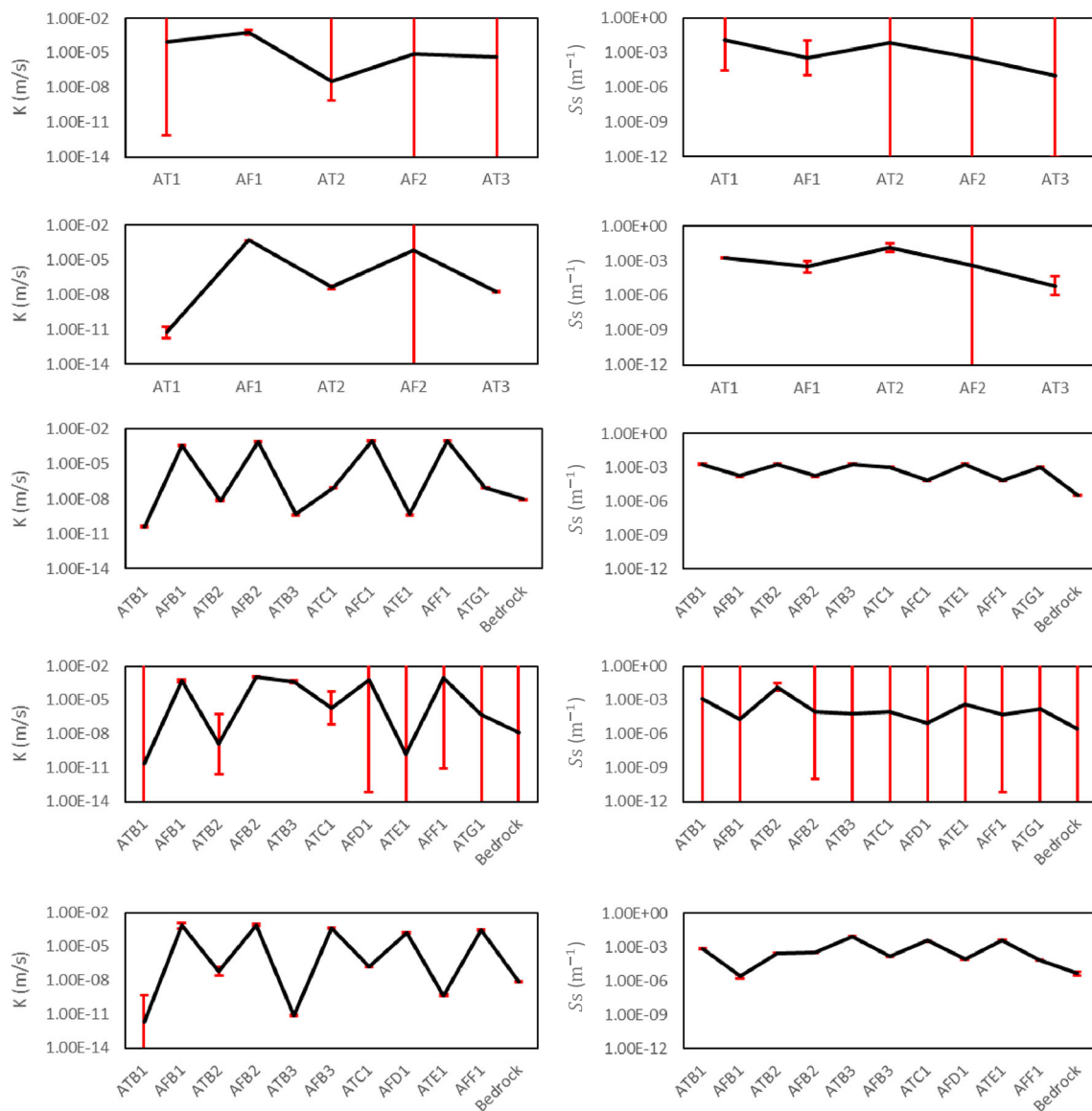
Fig. 9 Estimated K and S_s values as well as their corresponding 95% confidence intervals through PEST calibrations for different geological models: **a** the 5-layer uni model; **b** the 5-layer geological model; **c** the 11- layer geological model; **d** the Waterloo model; and **e** the Regional model. Some confidence intervals exceed the range of the vertical axis due to its large size. Values are provided in Table S3 of the ESM

the layer morphology is fixed, as unrealistic hydraulic parameters could be estimated (Zhao et al. 2016). Golder Associates (2011) summarized the K values at each water-supply well (Table S5 of the ESM), ranging from 9×10^{-4} to 1×10^{-1} m/s from well to well, which is similar to the estimated K values for AFB2 in this study.

The 95% confidence intervals for the estimated K for the upper part of the domain including ATB1, AFB1, ATB2, AFB2, ATB3, and ATC1, are relatively small in cases 2–5. The lower aquifer layer of the five-layer geological model is constructed by combining AFC1, ATE1 and AFF1 and assigning the initial K based on the property of two aquifers, but in reality, AFC1 and ATE1 are two discontinuous shallow aquifers with a thicker aquitard (ATE1) lying

between these two aquifers. Such a merged layer would affect the reliability of the K estimate, which is evident through the extremely large 95% confidence intervals.

The 95% confidence intervals are calculated in PEST on the basis of the linearity assumption encapsulated in the Jacobian matrix, so it is highly dependent on the assumptions underlying the model. If the confidence intervals are exaggerated from a breakdown in the linearity assumption, PEST will not truncate the confidence intervals by assigning narrower parameter bounds in order to avoid an unduly optimistic impression of parameter certainty, which will result in large confidence intervals.

There are three potential reasons for high levels of parameter uncertainty, including a poor fit between model outcomes

and field observations, a high level of parameter correlation, or insensitivity on the part of certain parameters (Doherty 2005). Since most data are concentrated in AFB2, the parameters of other layers may not be sufficiently sensitive to observation data, which results in large confidence intervals. Doherty (2005) stated that the high levels of parameter uncertainty resulting from excessive correlation or from insensitivity of parameters can be reduced by including more measurement data, so more data are needed from the deeper layers for more accurate parameter estimation.

The estimated S_s values are similar to K in that more reliable results are obtained for the shallower part of the models and large confidence intervals are mainly found in the lower layers of the five-layer uni model (case 1), the five-layer geological model (case 2) and the Waterloo models (case 4). The estimated S_s values are found to vary in the range of 3.32×10^{-6} to 1.95×10^{-3} m^{-1} for the 11-layer geological model (case 3), while for the 5-layer geological model (case 2), S_s varies between 6.78×10^{-6} and 1.81×10^{-3} m^{-1}. The S_s values estimated for AFB2 vary from 8.35×10^{-5} m^{-1} for the Waterloo model to 3.16×10^{-4} m^{-1} for the Regional model.

It is of interest to note that compared with previous estimates of S_s by Luo and Illman (2016) which ranged from 1.0×10^{-4} to 3.6×10^{-2} m^{-1} with a geometric mean of 4.0×10^{-3} m^{-1}, the range is greatly reduced and the values are much smaller for this study. This suggests that the S_s estimates of Luo and Illman (2016) could be impacted by a scale effect, which could be due to the use of the Theis (1935) solution by Luo and Illman (2016), which neglects heterogeneity and borehole storage effects (Luo and Illman 2017; see also Neville 2017). In contrast, flow parameters estimated through this study considered a larger simulation domain and the estimation was achieved simultaneously for all pumping/injection and observation well pairs considering layer-by-layer heterogeneity. It is also important to note that flow processes are simulated more accurately for the numerical model which may have caused the estimated storage parameters to be smaller. Similar findings have been found by Vesselinov et al. (2001) who discovered that scale effects in flow parameters became suppressed when the model used to conduct parameter estimation more rigorously considered heterogeneity in estimated parameters.

Performance of model calibration

The effect of different geological conceptualizations on model calibration is evaluated by comparing the simulated drawdowns versus observed drawdowns from 28 observation locations used for the model calibrations, as plotted in Fig. 10. A linear model is fit for each geological model case for performance evaluation.

Generally, the fit greatly improves from cases 1–5, with the slopes of the linear model ranging from 0.76 to 1.15 and values

of the coefficient of determination (R^2) increasing from 0.42 to 0.77. The R^2 value alone cannot be used to assess the fit, but it shows how close the data are to the fitted regression line and it is also used as an indication of data scatter. It is noted that the fits of the four geological models with appropriate initial K and S_s values (cases 2–5) are quite similar to the slopes of the linear model, ranging from 0.76 to 0.86 and values of R^2 increasing from 0.74 to 0.77. Although the estimated K and S_s values are quite different in the four cases, the simulated drawdown matches the observed drawdown for all four geological models quite well. Quantitative assessment is conducted by computing the mean absolute error norm (L_1) and the mean square error norm (L_2). Those quantities are computed as:

$$L_1 = \frac{1}{n} \sum_{i=1}^n \left| \chi_i - \widehat{\chi}_i \right| \tag{3}$$

$$L_2 = \frac{1}{n} \sum_{i=1}^n \left(\chi_i - \widehat{\chi}_i \right)^2 \tag{4}$$

where n is the total number of drawdown data, i indicates the data number, χ_i and $\widehat{\chi}_i$ represent the estimates from simulated and measured drawdowns, respectively.

The calculated L_1 and L_2 values are shown in Fig. 10. In particular, the calibration result based on the Regional model yields the smallest L_1 and L_2, while the five-layer uni model yields the largest L_1 and L_2. The L_1 and L_2 values for the five-layer model and the 11-layer model (case 3) are similar which may be due to the similar hydraulic properties of AF1 and AFB2, where most of the observations are located.

Large errors are mainly observed at wells where rapid changes in water levels are observed, but all the models fail to capture this fluctuation. In a previous study, Luo and Illman (2016) explained such poor matches by suggesting the potential existence of a high K pathway between some of the water-supply wells and water-monitoring boreholes.

Hydraulic tomography based on geological models, as presented in this study, treats each geological layer as homogeneous and isotropic, but in reality, both aquifers and aquitards are highly heterogeneous and could be anisotropic (Berg and Illman 2011b, 2013, 2015; Zhao and Illman 2017, 2018), thus a more sophisticated groundwater model that considers heterogeneity and anisotropy in each geological layer may be needed for future work to overcome this difficulty.

Another factor that may be contributing to the inconsistency is that the pumping/injection events normally extend for a couple of hours, but this study used daily pumping/injection rates for each water-supply well, which could decrease the actual pumping/injection rate for the simulation process.

Model validation results

The performances of different models in their ability to predict drawdown at monitoring wells were evaluated using the

Fig. 10 Scatterplots of observed versus simulated drawdowns for model calibration (black) and validation (red) based on 28 observation locations for: **a** the 5-layer uni model; **b** the 5-layer geological model; **c** the 11-layer geological model; **d** the Waterloo model; and **e** the Regional model. The solid line is a 1:1 line indicating a perfect match. The dashed lines are the best fit lines for model calibration (black) and validation (red). The linear model fit results are also included on each plot
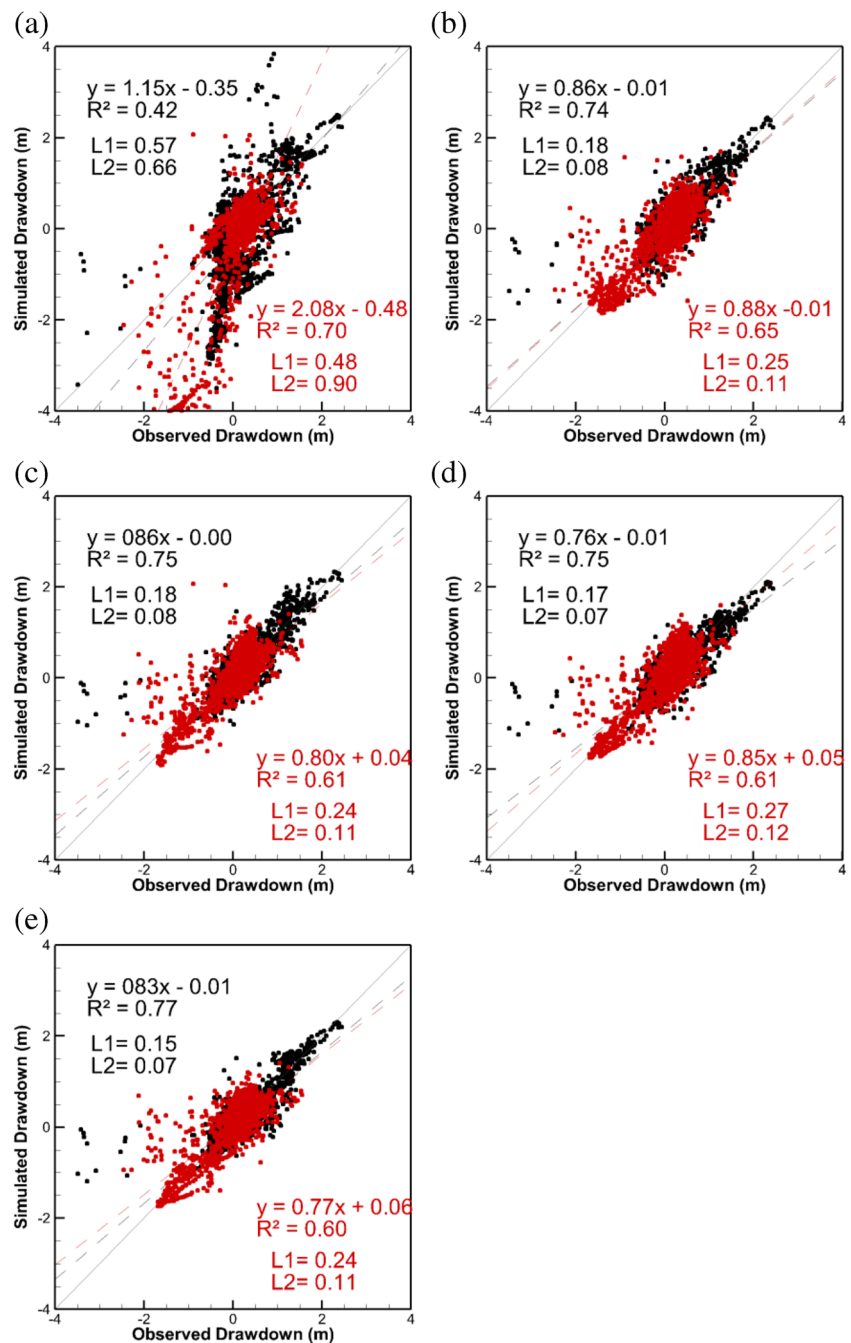
pumping/injection records from July to December, 2013. Simulated drawdowns are compared with corresponding observed drawdowns to provide quantitative evaluation. Similar to the calibration results, the validation results for four geological models with appropriate initial K and S_s values (cases 2–5) are similar in their overall shape in terms of the point distribution and the slope of the fit lines, while the five-layer uni (uniform) model in case 1 yields the worst results with biased prediction.

In a previous study, Zhao et al. (2016) found that as the number of pumping and monitoring points decreases, the performance gap among these models was reduced. Thus, the data set used in the calibration and validation processes may not be large enough to produce dramatic differences in the scatterplots among individual models. The slope of the linear model for the five-layer uni model is highest among the four models, and it also yields the largest L_1 and L_2 values. The 11-layer geological model and the Regional model both have the smallest L_1 and L_2, with a higher slope and a larger R^2 for the 11-layer geological model.

While the validation scatterplots show great similarity among individual models (cases 2–5), the simulated and

observed drawdowns for each observation well are provided (see Figs. S1 and S2 in the ESM) to better examine the performance of calibrated models in predicting drawdowns. The calibrated models for cases 2–5 can generally capture the drawdown curves for most observation wells, while larger differences are mainly shown for manually measured observation wells. There are not much data available from these wells for model calibration, and some of these wells are located in deeper layers (e.g., AFC1) without reliable estimates of hydraulic parameters. Cases 2–5 share more similarities compared with case 1, while the five-layer uni model (case 1) and the five-layer geological model (case 2) can better capture the rapid changes in water levels. The reason behind this may be the higher K values for AT2 and the merging of AFB1, ATB2 and AFB2 into AF1, which resulted in higher-K pathways between the water-supply wells and observation boreholes. Another reason may be the use of an averaged daily schedule of pumping/injection rates, when in reality, the rates vary more abruptly and can impact water level records and corresponding parameter estimates (Luo et al. 2020).

Summary and conclusions

This study investigated the usefulness of hydraulic tomography (HT) analysis based on geological models at a municipal well field to estimate the spatial distribution of hydraulic parameters (i.e., K and S_s) using long-term pumping/injection and monitoring well records. Pumping/injection rate data from 13 water-supply and water level data from 19 water-monitoring boreholes with 28 screens during the year of 2013 are selected and used for model calibration and validation. Four different geological conceptualizations with varying accuracy are used to examine the importance of geological data in HT analysis.

The study resulted in the following findings and conclusions:

- Model calibration and validation both revealed that hydraulic parameters (i.e., K and S_S) can be estimated using long-term pumping/injection rates and corresponding water-level records from municipal well fields. The estimated parameters are compared with those estimated through independently conducted pumping tests and the hydraulic parameters from both studies are consistent.
- Compared with traditional K and S_s estimation methods, which are difficult to conduct at well fields, the HT approach is successfully applied in this study through the use of long-term pumping/injection rates and corresponding water-level records. The observed drawdown curves are well captured by the calibrated models (cases 2–5) with appropriate initial parameter values. The use of such data for inverse modeling results in reliable hydraulic

parameter estimates, while enhancing cost and time efficiency in terms of site characterization. Therefore, it is suggested that these data be collected and used in future studies at other municipal wellfields.

- The simplified daily long-term pumping/injection rates and corresponding water-level records are used in the study. However, it is advisable to perform HT analysis based on real-time data with considering finer scale fluctuations in pumping records and corresponding drawdown measurements at observation wells. In addition, such pumping and observation data should also be collected from aquitard layers and deeper layers to better examine the heterogeneity patterns of complex groundwater systems and other municipal well fields.
- Dramatic differences are observed among four models in hydraulic parameter estimation which emphasize the importance of geological data and geological model construction on hydraulic estimates and groundwater modeling. Although good matches are obtained for both model calibration and validation for the four models, rapid water-level variations observed in the monitoring wells are not fully captured by all models. The large difference of the hydraulic estimates among four models and poor matches for the rapid water-level variation may be caused by the simplification of geological information and inaccurate classification of the geological layers. Thus, it is essential to apply a more sophisticated inverse model which considers the heterogeneity of individual geological units to better predict water-level changes and to obtain more reliable hydraulic estimates.
- The density of observation points can greatly affect the reliability of estimated hydraulic parameters. The large 95% confidence intervals and inconsistent parameter estimates for deeper geological layers promote the need for deeper well installation as well as hydraulic investigation. In addition, prior information of estimated parameters used in the model can reduce confidence interval widths. In future work, giving the unit a preferred value as prior information in PEST should be promoted instead of a uniform starting estimate for all layers.
- Hydrogeological data are critical for geological model construction used for the purpose of groundwater modeling. This study found that the geological model constructed with hydrogeological information (case 5) yields the smallest error norms for both groundwater model calibration and validation, so hydrogeological data are essential for HT analysis based on geological models to yield reliable hydraulic parameters and to better capture local heterogeneity.
- The variability in evapotranspiration throughout the year is not considered in this study, but it could become more important for longer simulation periods. It may also be necessary to more rigorously consider evapotranspiration

and other complexities (e.g., surface-water/groundwater interaction, variably saturated flow in the vadose zone, long-term decline of groundwater levels due to dewatering operations, etc.) at other municipal well fields that are not factored into the present study. If important processes are left out in the model used to estimate parameters, it is conceivable that the estimated parameters will be affected. This important topic will require additional studies in the future.

- Finally, this study applied a new approach to estimate hydraulic parameters for a municipal well field using existing long-term pumping/injection rates and water-level monitoring records. A more sophisticated inverse model which considers each aquifer/aquitard unit to be heterogeneous is currently being built for the study site. It is anticipated that improved parameter estimates will be obtained, which should result in more robust predictions of groundwater level variations due to municipal well operations. All of this should benefit well field management and contaminant transport predictions, as well as improve source water protection.

References

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

Aquanty Inc (2018) HydroGeoSphere: a three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. https://www.aquanty.com/hgs-download. Accessed 13 January 2021

ARANZ Geo Ltd (2015) Leapfrog Hydro 2.2.3. 3D geological modeling software. ARANZ Geo Ltd, Christchurch, New Zealand

Bajc AF, Newton MJ (2007) Mapping the subsurface of Waterloo region, Ontario, Canada: an improved framework of quaternary geology for hydrogeological applications. J Maps 219–230. https://doi.org/10.1080/jom.2007.9710840

Bajc AF, Shirota J (2007) Three-dimensional mapping of surficial deposits in the regional municipality of Waterloo, southwestern Ontario. Groundwater Resources Study 3. Ontario Geological Survey, Sudbury, ON

Berg SJ, Illman WA (2011a) Capturing aquifer heterogeneity: comparison of approaches through controlled sandbox experiments. Water Resour Res 47:1–17. https://doi.org/10.1029/2011WR010429

Berg SJ, Illman WA (2011b) Three-dimensional transient hydraulic tomography in a highly heterogeneous glaciofluvial aquifer-aquitard system. Water Resour Res 47. https://doi.org/10.1029/2011WR010616

Berg SJ, Illman WA (2013) Field study of subsurface heterogeneity with steady state hydraulic tomography. Ground Water 51:29–40. https://doi.org/10.1111/j.1745-6584.2012.00914.x

Berg SJ, Illman WA (2015) Comparison of hydraulic tomography with traditional methods at a highly heterogeneous site. Groundwater 53:71–89. https://doi.org/10.1111/gwat.12159

Blackport Hydrogeology Inc (2012a) Tier three water budget and local area risk assessment: River wells, Pompeii, Woolner and Forwell well fields characterization study. Report prepared for the Regional Municipality of Waterloo. Blackport Hydrogeology, Waterloo, ON

Blackport Hydrogeology Inc (2012b) Tier three water budget and local area risk assessment: Waterloo north, William street and Lancaster well fields characterization study. Report prepared for the regional municipality of Waterloo. Blackport Hydrogeology, Waterloo, ON

Blackport RJ, Meyer PA, Martin PJ (2014) Toward an understanding of the Waterloo Moraine hydrogeology. Can Water Resour J 39(2):120–135

Bohling GC, Zhan X, Butler JJ, Zheng L (2002) Steady shape analysis of tomographic pumping tests for characterization of aquifer heterogeneities. Water Resour Res 38. https://doi.org/10.1029/2001WR001176.60-1-60-15

Bouwer H, Rice RC (1976) A slug test for determining hydraulic conductivity of unconfined aquifers with completely or partially penetrating wells. Water Resour Res 12(3):423–428

Callow IP (1996) Optimizing aquifer production for multiple wellfield conditions in Kitchener, Ontario. MSc Thesis, University of Waterloo, Waterloo, ON

Cardiff M, Bakhos T, Kitanidis PK, Barrash W (2013a) Aquifer heterogeneity characterization with oscillatory pumping: sensitivity analysis and imaging potential. Water Resour Res 49:5395–5410. https://doi.org/10.1002/wrcr.20356

Cardiff M, Barrash W, Kitanidis PK (2013b) Hydraulic conductivity imaging from 3-D transient hydraulic tomography at several pumping/observation densities. Water Resour Res 49:7311–7326. https://doi.org/10.1002/wrcr.20519

Carle SF, Fogg GE (1997) Modeling spatial variability with one and multi-dimensional continuous Markov chains. Math Geol 29(7):891–918

Carr JC, Beatson RK, Cherrie JB, Mitchell TJ, Fright WR, McCallumm BR, Evans TR (2001) Reconstruction and representation of 3-D objects with radial basis functions. SIGGRAPH Computer Graphics Proceedings, Annual Conference Series, August 2001, AMC SIGGRAPH, pp 67–76. http://education.siggraph.org/. Accesssed March 2021

CH2M HILL (2003a) Results of construction and testing of the ASR wells at the Mannheim Water Treatment Plant site: the Region of Waterloo. CH2M HILL, Englewood, CO

CH2M HILL, Papadopulos Associates (2003b) Alder Creek Groundwater Study (final report): the Region of Waterloo. CH2M HILL, Englewood, CO

Cooper HH, Jacob CE (1946) A generalized graphical method for evaluating formation constants and summarizing well field history. Am Geophys Union 27:526–534

Doherty J (2005) PEST: model-independent parameter estimation user manual. Watermark, Brisbane, Australia

Farvolden RN, Greenhouse JP, Karrow PF, Pehme PE, Ross LC (1987) Ontario Geoscience Research Fund Grant 128, subsurface quaternary stratigraphy of Kitchener-Waterloo, using borehole geophysics. OFR5623, Ontario Geological Survey, Sudbury, ON

Frank T, Tertois AL, Mallet JL (2007) 3-D reconstruction of complex geological interfaces from irregularly distributed and noisy point data. Comput Geosci 33:932–943

Frind EO, Molson JW, Sousa MR, Martin PJ (2014) Insights from four decades of model development on the Waterloo moraine, a review. Can Water Resour J 39(2). https://doi.org/10.1080/07011784.2014.914799

Golder Associates (2011) Tier three water budget and local area risk assessment: Mannheim well fields characterization. Report prepared for the regional municipality of Waterloo. Golder, Mississauga, ON

Guo W (2017) Simulating the impact of water infrastructure on groundwater and surface water within the Laurel Creek watershed. UWSpace, Waterloo, ON

Hardy RL (1971) Multiquadric equations of topography and other irregular surfaces. J Geophys Res 176:1905–1915

Harp DR, Vesselinov VV (2011) Identification of pumping influences in long-term water level fluctuations. Ground Water 49(12). https://doi.org/10.1111/j.17456584.2010.00725.x

Hazen A (1911) Discussion: dams on sand foundations. Am Soc Civil Eng 73:199–203

Hvorslev MJ (1951) Time lag and soil permeability in ground-water observations. Bull. No. 36, Waterways Exper. Station, Corps of Eng., US Army, Vicksburg, MI, 50 pp

Illman WA, 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 341:222–234. https://doi.org/10.1016/j.jhydrol.2007.05.011

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

Illman WA, Liu X, Takeuchi S, Yeh TCJ, Ando K, Saegusa H (2009) Hydraulic tomography in fractured granite: Mizunami underground research site, Japan. Water Resour Res 45:1–18. https://doi.org/10.1029/2007WR006715

Illman WA, Zhu J, Craig AJ, Yin D (2010a) Comparison of aquifer characterization approaches through steady state groundwater model validation: a controlled laboratory sandbox study. Water Resour Res 46:1–18. https://doi.org/10.1029/2009WR007745

Illman WA, Berg SJ, Liu X, Massi A (2010b) Hydraulic/partitioning tracer tomography for trichloroethene source zone characterization: small-scale sandbox experiments. Environ Sci Technol 44(22):8609–8614. https://doi.org/10.1021/es101654j

Illman WA, Berg SJ, Zhao Z (2015) Should hydraulic tomography data be interpreted using geostatistical inverse modeling? A laboratory sandbox investigation. Water Resour Res 51:3219–3237. https://doi.org/10.1002/2014WR016552

International Water Supply Ltd (1973) Kitchener-Waterloo groundwater evaluation. In: Report to the regional municipality of Waterloo, International Water Supply, Barrie, ON

Karrow PF (1993) Quaternary geology of the Stratford- Conestogo area, southern Ontario. Report 283. Ontario Geological Survey, Sudbury, ON

Kozeny J (1927) Uber Kapillare Leitung Des Wassers in Boden [Capillary movement of water in soil]. Sitzungsber Akad. Wiss. Vienna, Math. Naturwiss. Kl., Abt.2a 136:271–306

Lake Erie Region Source Protection Committee (2012) Grand River source protection area, Approved assessment report, Lake Erie Region Source Protection Committee, Cambridge, ON

Liao HS, Curtis ZK, Sampath PV, Li SG (2020) Simulation of flow in a complex aquifer system subjected to long-term well network growth. Groundwater 58(2):301–322

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

Luo N, Illman WA (2017) Reply to comments by Christopher J. Neville on "Automatic estimation of aquifer parameters using long-term water supply pumping and injection records". Hydrogeol J 25:2211–2214. https://doi.org/10.1007/s10040-017-1644-7

Luo N, Zhao Z, Illman WA, Berg SJ (2017) Comparative study of transient hydraulic tomography with varying parameterizations and zonations: laboratory sandbox investigation. J Hydrol 554:758–779. https://doi.org/10.1016/j.jhydrol.2017.09.045

Luo N, Illman WA, Zha YY, Park YJ, Berg SJ (2020) Three-dimensional hydraulic tomography analysis of long-term municipal wellfield operations: validation with synthetic flow and solute transport data. J Hydrol 590:125438. https://doi.org/10.1016/j.jhydrol.2020.125438

Martin PJ, Frind EO (1998) Modelling methodology for a complex multi-aquifer system: the Waterloo moraine. Ground Water 36(4):679–690

Matrix Solutions Inc, S.S. Papadopulos and Associates (2014a) Region of Waterloo tier three water budget and local area risk assessment; model calibration and water budget report. August 2014, Matrix, Saskatoon, SK

Matrix Solutions Inc, S.S. Papadopulos and Associates (2014b) Region of Waterloo tier three water budget and local area risk assessment; final risk assessment report. September 2014, Matrix, Saskatoon, SK

Mishra P, Vesselinov VV (2011) WELLS: multi-well variable-rate pumping-test analysis tool. https://gitlab.com/monty/wells. Accessed 7 Feb 2014

Molson JW, Martin PJ, Frind EO (1995) WATFLOW version 1.0: a three-dimensional groundwater flow model, user's guide. Waterloo Centre for Groundwater Research, Waterloo, ON

Neville CJ (2017) Comment on "Automatic estimation of aquifer parameters using long-term water supply pumping and injection records": paper published in *Hydrogeology Journal* (2016) 24:1443–1461, by Ning Luo and Walter A. Illman. Hydrogeol J 25:2207–2209. https://doi.org/10.1007/s10040-017-1643-8

Paloschi GVR (1993) Subsurface stratigraphy of the Waterloo moraine. MSc Project, University of Waterloo, Waterloo, ON

Refsgaard JC, Christensen S, Sonnenborg TO, Seifert D, Højberg AL, Troldborg L (2012) Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. Adv Water Resour 36:36–50. https://doi.org/10.1016/j.advwatres.2011.04.006

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

Rehfeldt KR, Boggs JM, Gelhar LW (1992) Field study of dispersion in a heterogeneous aquifer: 3. geostatistical analysis of hydraulic conductivity. Water Resour Res 28(12):3309–3324

Rudolph DL (1985) A quasi 3-dimensional finite element model for steady-state analysis of multi-aquifer systems. MSc Thesis, University of Waterloo, Waterloo, ON

Rudolph DL, Sudicky EA (1990) Simulation of groundwater flow in complex multiaquifer systems: performance of quasi three-dimensional technique in the steady state case. Can Geotech J 27:590–600

Sampath PV, Liao HS, Curtis ZK, Doran PJ, Herbert ME, May CA, Li SG (2015) Understanding the groundwater hydrology of a geographically-isolated prairie fen: implications for conservation. PLoS One 10(10):e0140430. https://doi.org/10.1371/journal.pone.0140430

Sampath PV, Liao HS, Curtis ZK, Herbert ME, Doran PJ, May CA, Landis DA, Li SG (2016) Understanding fen hydrology across multiple scales. Hydrol Process 30(19):3390–3407

Shepherd RG (1989) Correlations of permeability and grain-size. Groundwater 27(5):633–638. https://doi.org/10.1111/j.1745-6584.1989.tb00476.x

Stantec Consulting Ltd (2009) Tier three water budget and local area risk assessment: Strange street well field characterization study. Report prepared for the regional municipality of Waterloo. Stantec, Kitchener, ON

Stantec Consulting Ltd (2012a) Tier three water budget and local area risk assessment: Erb street well field characterization study. Report prepared for the regional municipality of Waterloo. Stantec, Kitchener, ON

Stantec Consulting Ltd (2012b) Tier three water budget and local area risk assessment: Greenbrook well field characterization study. Report prepared for the regional municipality of Waterloo. Stantec, Kitchener, ON

Stantec Consulting Ltd (2012c) Tier three water budget and local area risk assessment: parkway and Strasburg well field characterization study. Report prepared for the regional municipality of Waterloo. Stantec, Kitchener, ON

Sudicky EA (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

Sudicky EA, Illman WA, Goltz IK, Adams JJ, McLaren RG (2010) Heterogeneity in hydraulic conductivity and its role on the macro-scale transport of a solute plume: from measurements to a practical application of stochastic flow and transport theory. Water Resour Res 46(1). https://doi.org/10.1029/2008WR007558

Tecplot Inc (2011) TECPLOT. Bellevue, WA. https://www.tecplot.com/. Accessed March 2021

Terraqua Investigations Ltd (1992) Waterloo north aquifer system study. Report prepared for the regional municipality of Waterloo. Terraqua InvestigationsWaterloo, ON

Terraqua Investigations Ltd (1995) Study of the hydrogeology of the Waterloo moraine. Report prepared for the regional municipality of Waterloo. Terraqua InvestigationsWaterloo, ON

Theis CV (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 16:519–524

Traum JA, Phillips SP, Bennett GL, Zamora C, Metzger LF (2014) Documentation of a Groundwater Flow Model (SJRRPGW) for the San Joaquin River Restoration Program Study Area, California (no. 2014-5148). US Geological Survey, Reston, VA

Vesselinov VV, Neuman SP, Illman WA (2001) Three-dimensional numerical inversion of pneumatic cross-hole tests in unsaturated fractured tuff: 2. equivalent parameters, high-resolution stochastic imaging and scale effects. Water Resour Res 37(12):3019–3042

Wu CM, Yeh TCJ, Zhu J, Tim HL, Hsu NS, Chen CH, Sancho AF (2005) Traditional analysis of aquifer tests: comparing apples to oranges? Water Resour Res 41:1–12. https://doi.org/10.1029/2004WR003717

Yeh TCJ, Lee CH (2007) Time to change the way we collect and analyze data for aquifer characterization. Ground Water 45(2):116–118. https://doi.org/10.1111/j.1745-6584.2006.00292.x

Yeh TCJ, Liu S (2000) Hydraulic tomography: development of a new aquifer test method. Water Resour Res 36:2095. https://doi.org/10.1029/2000WR900114

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

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

Zhao Z, Illman WA, Berg SJ (2016) On the importance of geological data for hydraulic tomography analysis: laboratory sandbox study. J Hydrol 542:156–171. https://doi.org/10.1016/j.jhydrol.2016.08.061

Zhao Z, Illman WA, Zha Y, Yeh TCJ, Mok CMB, Berg SJ, Han D (2019) Transient hydraulic tomography analysis of fourteen pumping tests at a highly heterogeneous multiple aquifer–aquitard system. Water 11(9):1864. https://doi.org/10.3390/w11091864