

RESEARCH ARTICLE

10.1002/2014WR016402

Key Points:

- Sequential hydraulic tomographical inversion
- Combination of structural information and full signal inversion
- Tracer tests used as independent validation

Correspondence to:

S. Jiménez,
santos.jimenez@erdw.ethz.ch

Citation:

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

Received 18 SEP 2014

Accepted 13 JUN 2015

Accepted article online 17 JUN 2015

Published online 19 JUL 2015

Prediction of solute transport in a heterogeneous aquifer utilizing hydraulic conductivity and specific storage tomograms

S. Jiménez¹, R. Brauchler^{1,2}, R. Hu³, L. Hu¹, S. Schmidt³, T. Ptak³, and P. Bayer¹
¹Department of Earth Sciences, ETH Zurich, Zurich, Switzerland, ²Now at AF-Consult Switzerland Ltd., Baden, Switzerland, ³Geoscientific Centre, University of Göttingen, Göttingen, Germany

Abstract A sequential procedure of hydraulic tomographical inversion is applied to characterize at high resolution the spatial heterogeneity of hydraulic conductivity and specific storage at the field test site Stegemühle, Germany. The shallow aquifer at this site is examined by five short-term multilevel pumping tests with 30 pumping-observation pairs between two wells. Utilizing travel time diagnostics of the recorded pressure response curves, fast eikonal-based inversion is shown to deliver insight into the sedimentary structures. Thus, the structural information from the generated travel time tomogram is exploited to constrain full calibration of the pressure response curves. Based on lateral extrapolation from the measured inter-well profile, a three-dimensional reconstruction of the aquifer is obtained. It is demonstrated that calibration of spatially variable specific storage in addition to hydraulic conductivity can improve the fitting of the model while the structural features are only slightly changed. At the field site, two tracer tests with uranine and sodium-naphthionate were also performed and their concentrations were monitored for 2 months. The measured tracer breakthrough curves are employed for independent validation of the hydraulic tomographical reconstruction. It is demonstrated that major features of the observed solute transport can be reproduced, and structures relevant for macrodispersive tracer spreading could be resolved. However, for the mildly heterogeneous aquifer, the tracer breakthrough curves can also be approximated by a simplified homogeneous model with higher dispersivity. Therefore, improved validation results that capture specific characteristics of the breakthrough curves would require additional hydraulic measurements.

1. Introduction

Nearly all hydraulic tomographic field studies are driven by the need to provide spatial high-resolution parameter fields for solute transport predictions. In fact, tomographic approaches, among others [e.g., Mariethoz *et al.*, 2010], can resolve sedimentary structures or fractures that control preferential flow. Their potential and superiority to traditional field investigation techniques was demonstrated in several previous studies [Gottlieb and Dietrich, 1995; Yeh and Liu, 2000; Illman *et al.*, 2010; Berg and Illman, 2011a, 2015]. However, the effort of data collection and data evaluation is higher for tomographic investigation methods in comparison to conventional methods that avoid spatial assignment of estimated hydraulic parameters. This motivates a strong interest for enhanced tomographic field and inversion techniques [Bohling *et al.*, 2002; Zhu and Yeh, 2006; Lochbühler *et al.*, 2013]. Naturally, the development of new field technologies and field data collection strategies is delayed in time in comparison to the computer-based development of inversion schemes. Numerical studies with virtual aquifers are essential means for motivating, developing, and testing new schemes, but their viability can only be approved by often laborious field experiments. Therefore, especially during the last few years, the number of field studies has been catching up. These started from simplified two-dimensional, depth-integrated characterizations [e.g., Straface *et al.*, 2007b] to arrive at full three-dimensional reconstructions [e.g., Illman *et al.*, 2009; Berg and Illman, 2011b; Cardiff *et al.*, 2013] based on a large number of interference tests.

For field investigations, one of the important and at the same time most challenging tasks is the evaluation of the significance and reliability of the reconstructed hydraulic parameter fields. Independent from the inversion technique, all field studies use the residual error from data fitting as a first measure for the quality of their inversion results. Unfortunately, this information is not sufficient because a large number of

parameter distributions might exist that equally honor the measured data. Hence, independent information and measures have to be exploited to evaluate the quality of the reconstructed parameter fields. Geological information such as deposition information or fault information based on a detailed structural geological study is utilized by *Straface et al.* [2007a] and *Illman et al.* [2009] to support reconstructed parameter fields. *Berg and Illman* [2012] used a large number of permeameter and grain size tests in combination with multi-level slug tests to interpret the estimated hydraulic conductivity (K) fields. The tests were performed at the North Campus Research Site (NCRS), Waterloo, Canada, in a heterogeneous confined aquifer built up by tills and glaciofluvial deposits. The high vertical resolution of multilevel slug tests and direct-push injection-logs were exploited by *Brauchler et al.* [2010, 2013] to interpret reconstructed diffusivity fields estimated at the Stegemühle Site, Göttingen, Germany. This site is characterized by a shallow confined aquifer consisting of fluvial sediments. In comparison to the conditions at the NCRS [variance of log conductivity, $\sigma_K^2 = 1.72$, *Alexander et al.*, 2011], the aquifer at the Stegemühle Site is less heterogeneous ($\sigma_K^2 = 0.2$).

Cardiff et al. [2012, 2013] compared porosity logs and multilevel slug tests with reconstructed parameter fields. They performed a three-dimensional (3-D) transient hydraulic tomographic field experiment utilizing highly flexible packer systems at the Boise Hydrogeophysical Research Site (BHRS), USA. The BHRS site [e.g., *Straface et al.*, 2011; *Cardiff et al.*, 2013] is characterized by a mixed sand/gravel/cobble facies and in comparison to the other two test sites it shows unconfined conditions.

Hydraulic tomographical measurements were performed by *Vasco and Karasaki* [2001, 2006] to reconstruct preferential flow paths at the Raymond Field Site, California, USA. Intensive geological experiments allowed for a comparison of identified preferential flow paths, imaged in the hydraulic tomograms, with borehole conductivity logs and seismic tomograms. Such utilization of independently collected data for comparison with the reconstructed hydraulic tomograms can be a challenge due to different observation scale and mismatch in resolution [e.g., *Brauchler et al.*, 2012]. *Huang et al.* [2011] successfully applied independent validation pumping tests to field data recorded at the test site of the National Yunlin University of Science and Technology in Taiwan. The site consists of fluvial sediments with mean hydraulic conductivity values of around 10^{-4} m/s, which are comparable to those found at the Stegemühle Site. A more direct way to validate tomograms is to use direct visual comparisons between inverted hydraulic conductivity and laboratory experiments [e.g., *Illman et al.*, 2007, 2010].

Although the main motivation of hydraulic tomographic field studies is to provide high-resolution information for solute transport predictions, only a small number of field studies were published that employ tracer test data to interpret or validate reconstructed tomograms. *Bohling et al.* [2007] utilized a solute tracer test to evaluate the capability of steady-state tomography. They show that the tracer test could support the existence of a highly conductive layer, which was reconstructed by hydraulic tomography. For further verification, a large number of small-scale hydraulic tests were performed at the Geohydrologic Experimental and Monitoring Site (GEMS), USA, but none of these revealed the presence of this high-conductivity zone. The GEMS site is a heavily studied alluvial confined aquifer that consists of 11 m of sand and gravel overlain by silt and clay. Another field example utilizing tracer test data and flow data for inversion was presented by *Vasco and Finsterle* [2004] at the Grimsel Rock Laboratory in Switzerland. In contrast to the work of *Bohling et al.* [2007], the tracer test data were not used for validation. Instead, the different data types were inverted together to improve the significance of the reconstructed hydraulic tomograms. *Illman et al.* [2012] showed that estimated hydraulic conductivity tomograms predicted better tracer distribution patterns during a dipole tracer test than other traditional methodologies (i.e., effective parameter/macrodispersion approach or heterogeneous approach using ordinary kriging based on core samples). They also emphasize the difficulties of capturing details of the tracer breakthrough due to intrinsic methodological limitations, such as effects of noise in head measurements and “the less diffusive nature of the tracer which demands a much higher resolution mapping of the K -field.” *Ni et al.* [2009] compared the predictive capabilities of a 2-D tomographic reconstruction to that of a homogeneous model. In their theoretical study, they showed that the tomographic variant was capable of reproducing tracer breakthrough curves (BTCs) independently of the transport distance. In contrast, a homogeneous advection-dispersion model with empirically estimated dispersivity could not match the BTC. When calibrated to the BTC, the homogeneous model would properly reproduce the BTC, but obtained dispersivity would need to be raised with transport distance.

In this paper, our main objective is to validate subsurface reconstructions from hydraulic tomographic inversion for predicting solute transport in the field. Therefore, we choose the hydraulic tomography procedure

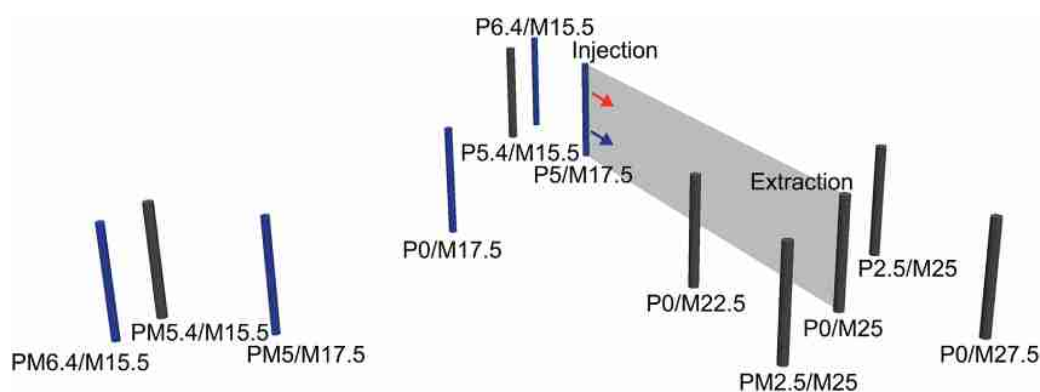


Figure 1. Monitoring well network at the Stegemühle test site. The 2" wells are colored black, and multichamber wells are colored in blue. The plane between P5/M17.5 and P0/M25 corresponds to the eikonal inversion domain.

developed by Jiménez *et al.* [2013], which combines eikonal and pilot point-based inversion approaches. The procedure was originally presented for the reconstruction of a K -field and theoretically assessed by application to a virtual aquifer. Here we further develop and adapt it to the requirements of the simultaneous 3-D reconstruction of specific storage and hydraulic conductivity. We apply it to the inversion of short-term pumping tests at the Stegemühle site. Between the same wells originally used for the hydraulic tomographic investigations by Brauchler *et al.* [2013], a forced gradient tracer test is performed. Two fluorescent tracers are injected in two different depths of the aquifer and the BTCs are recorded in the observation well over the entire thickness of the aquifer. The tracer test results are contrasted with the new findings from hydraulic inversion, and we reveal, to what extent and at which accuracy it is feasible to reconstruct structures relevant for subsurface transport. Similarly to Ni *et al.* [2009], we also compare the BTC prediction by reconstructed model to that of a homogenous one.

2. Material and Methods

2.1. Field Site and Experiments

2.1.1. Stegemühle Site

The experimental field test site Stegemühle is located south of the city of Göttingen, Lower Saxony, Germany (Figure 1). In order to carry out hydrogeological and hydrochemical field research under controlled natural conditions, five 1"-observation, twenty-one 2"-observation (five of them are multichamber wells), and three 6"-observation wells were installed during the period of 2006–2011. The composition of the shallow subsurface was determined by a variety of methods such as inspection of sediment cores, grain size analysis, direct-push electrical conductivity logging, borehole gamma-ray logging, electrical resistivity tomography, and seismic travel time inversion [e.g., Hu, 2007; Vogt, 2007; Meyer *et al.*, 2014; Hu, 2011]. The aquifer is composed of unconsolidated fluvial sediments (sand and gravel) of Quaternary age (Weichsel Glaciation). These sediments have a varying thickness of 1.0–3.3 m and are overlain by alluvial clay. The aquifer bottom is at a depth of 1.9–7.0 m below land surface with erosional contact to the underlying clay stone formation of Middle Keuper Age. In the middle of the field site, which is the focus area of this study, the aquifer exhibits confined conditions. Here Hu [2011] and Brauchler *et al.* [2013] applied multilevel slug tests and observed vertically varying hydraulic conductivities with higher values at the bottom of the aquifer.

2.1.2. Field Implementation of Hydraulic Tests

A series of cross-well multilevel pumping tests were performed at the test site Stegemühle, implementing a tomographic array along a straight line between a pumping well (P0/M25) and an observation well (P5/M17.5) (Table 1). The distance between these two wells is 9 m (Figure 1). During each pumping test, the water was partially pumped out of the pumping well P0/M25 by employing double packer systems with a screened interval of 0.25 m. The tube connected to the pump has an internal diameter (ID) of 0.031 m. The observation well P5/M17.5 (Figure 2) is a multichamber well constructed with the Continuous Multichannel Tubing (CMT) System [Einarson and Cherry, 2002]. This well consists of a pipe with six continuous separate

Table 1. Basic Information of the Two Wells P0/M25 and P5/M17.5 Used for Hydraulic Tomography Inversion and Tracer Testing

	P0/M25	P5/M17.5
Type	Single screen	Multichamber
Aquifer thickness (m)	1.98	1.99
Well height (m)	0.87	0.65
Elevation of the well top (m.a.s.l.)	152.23	151.54
Surface elevation (m.a.s.l.)	151.36	150.89
Well bottom (m.a.s.l.)	145.28	145.3
Aquifer top (m.a.s.l.)	147.382	147.415
Aquifer bottom (m.a.s.l.)	145.401	145.419

channels ($ID = 0.014$ m), which are arranged in a honeycomb shape and lead to different depths. This design allows for the measurement of water level changes at different depths of the aquifer.

For the profile between the pumping well and the observation well, five short-term pumping tests were carried out. For every short-term pumping test and every pumping interval, the pressure changes in the six different depths of the multichamber wells

were recorded at a frequency of 50 Hz with the pressure transducer (PDCR 35/D-8070) connected to a data logger (Campbell Scientific® CR 3000). By varying the pumping interval, a total number of 30 (5×6) draw-down curves for the profile were recorded [Brauchler *et al.*, 2013]. The pumping tests in series produced a pattern of crossing trajectories between test and observation well, similar to the paths of a radar or seismic experiment. The travel times and hydraulic attenuations between the wells P0/M25 and P5/M17.5 thus can be utilized for eikonal-based cross-sectional reconstruction of hydraulic parameter distributions.

2.1.3. Field Implementation of Tracer Tests

Tracer test data are used for independent validation of the derived aquifer model. Nonreactive tracer tests are effective means to identify preferential flow paths or integral transport parameters of the subsurface, such as porosity and dispersivity. They can be conducted under natural gradient or forced gradient conditions. Compared to the natural gradient tracer test, the forced gradient tracer test is hydraulically well

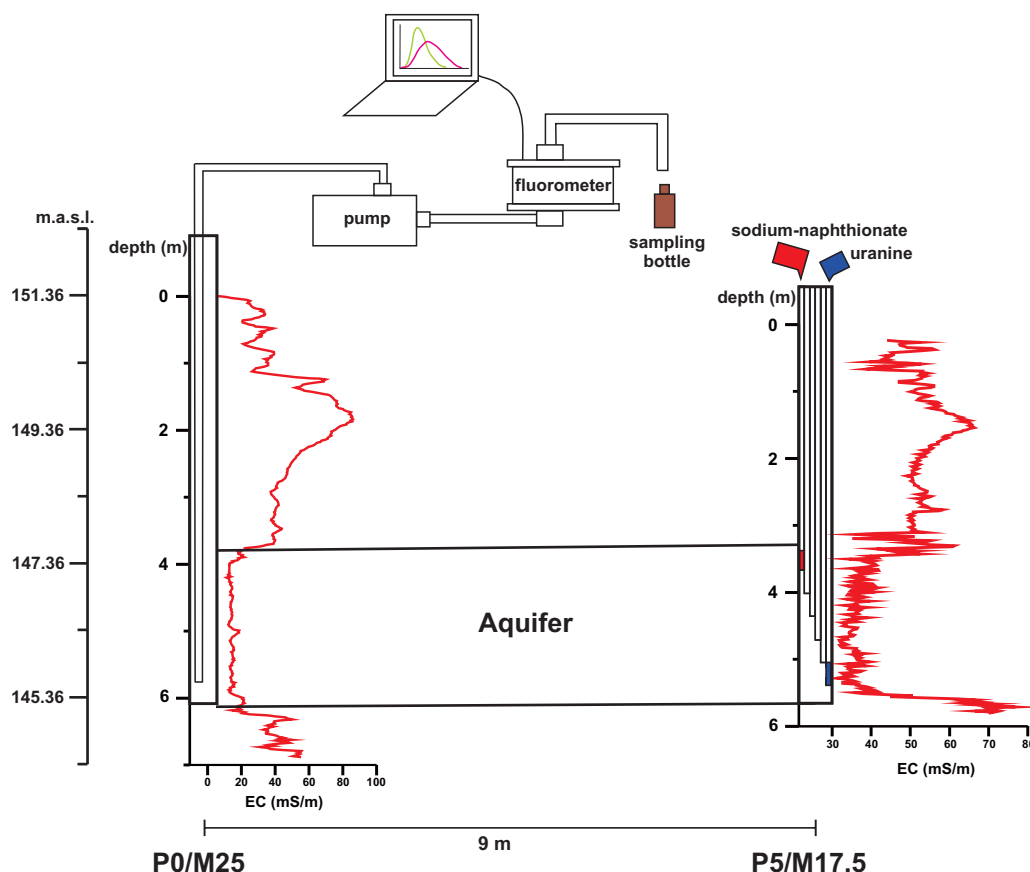


Figure 2. Setup of the tracer experiment at Stegemühle site with well configuration and illustration of injection levels for uranine and sodium-naphthionate. Additionally, electrical conductivity (EC) measurements are depicted (in red) for both wells, which delineate the aquifer boundaries.

Table 2. Experimental Design of Two Tracer Tests at the Stegemühle Field Site

	Uranine	Sodium-Naphthionate
Pumping rate (L/s)	0.301–0.305	
Injection start date and time	18 Oct 2011, 13:45	18 Oct 2011, 13:53
Injection mass (g)	15	150
Injection chamber	6	1
Height of chamber (m.a.s.l.)	145.761	147.261
Injection duration (s)	300	180

controlled and not constrained to a given natural flow field. Moreover, higher mass recovery rates can be achieved [e.g., Ptak *et al.*, 2004]. Consequently, forced gradient conditions were also favored at the field site.

The experiment was conducted between the same wells previously used for the pumping tests (P0/m25, P5/m17.5). We selected two different tracers with similar transport properties, the fluorescent dye

tracers uranine and sodium-naphthionate [Leibundgut *et al.*, 2009], in order to allow for a robust analysis and to mitigate the effects of possible measurement errors, data noise, or tracer-specific transport behavior.

Prior to tracer injection, a steady radial flow field was established by continuous water extraction from the single screen well at a constant rate of ~ 0.3 L/s. A mass of 15 g uranine and 150 g sodium-naphthionate dissolved in water were injected into the multichamber well, each in a different depth (Figure 2). The injection periods of uranine and sodium-naphthionate were 5 min and 3 min, respectively, and by this a pulse-like injection was realized. The different injection periods were employed due to technical reasons. They have negligible influence on the results as they are very small compared to the time of tracer arrival. At the monitoring well depth-integrated concentrations were measured every 30 s by a flow-through field fluorometer (type GGUN-FL30), calibrated to the specific tracers and local groundwater conditions. For quality control, at some time points the pumped-out water was also manually sampled and subsequently analyzed in the laboratory. The entire tracer experiment lasted 2 months, and the water levels at the two wells were measured every few days to ensure the flow field was steady. Detailed information on experimental settings is summarized in Table 2.

2.2. Tomographic Inversion

The inversion procedure we use here encompasses the sequential application of an eikonal-based and pilot point-based inversion scheme. It was presented in detail by Jiménez *et al.* [2013] and makes use of the strengths of both inversions and minimizes drawbacks. On the one side, eikonal-based inversion is a fast, well-tested methodology capable of providing insight into hydraulic parameters and aquifer structure. Strictly speaking it represents an approximation, since it treats the parabolic flow equation as a wave equation. In addition, the eikonal approach, as presented in this paper, only uses a time diagnostic from the whole pressure signal leaving the rest of the information unused. As a consequence, the eikonal-based inversion is highly efficient for fast detection of structures, but as an approximation it is less accurate in estimation of hydraulic parameter values. On the other side, pilot point-based inversion commonly works by fitting the flow equation, using the full recorded pressure signal and thus making full use of the measured information. Therefore, hydraulic parameter values can be determined, but this is computationally demanding, especially for 3-D reconstructions.

Jiménez *et al.* [2013] showed how to link both schemes in a synergetic way (Figure 3). Eikonal-based inversion is utilized to extract structures from reconstructed diffusivity fields (D -tomograms) but not parameter values. Since we are interested mainly in K , the original procedure is refined here by utilizing eikonal-based estimates of K -distribution rather than diffusivity fields. As a bridging step, specific storage (S_s)-tomograms are developed by attenuation tomography [Brauchler *et al.*, 2013]. Given $D = K/S_s$, a fully eikonal-based K -tomogram can be derived from the D -tomograms and S_s -tomograms. The reliability of the tomographic models is assessed by means of null-space energy maps. A null-space energy map represents a measure of the reliability of a tomogram. It relates the trajectory distribution to the mesh used for the discretization of the investigated area [e.g., Böhm and Vesnaver, 1996] and comprises a singular value decomposition of the tomographic matrix.

We consider the eikonal-based K -distribution as a proxy, but carrying valuable insight into subsurface structures. For extracting this information, it is converted into a zonal image (conceptual map) using a k -means clustering algorithm. Pilot points serve as auxiliary variables that are often combined with regularization

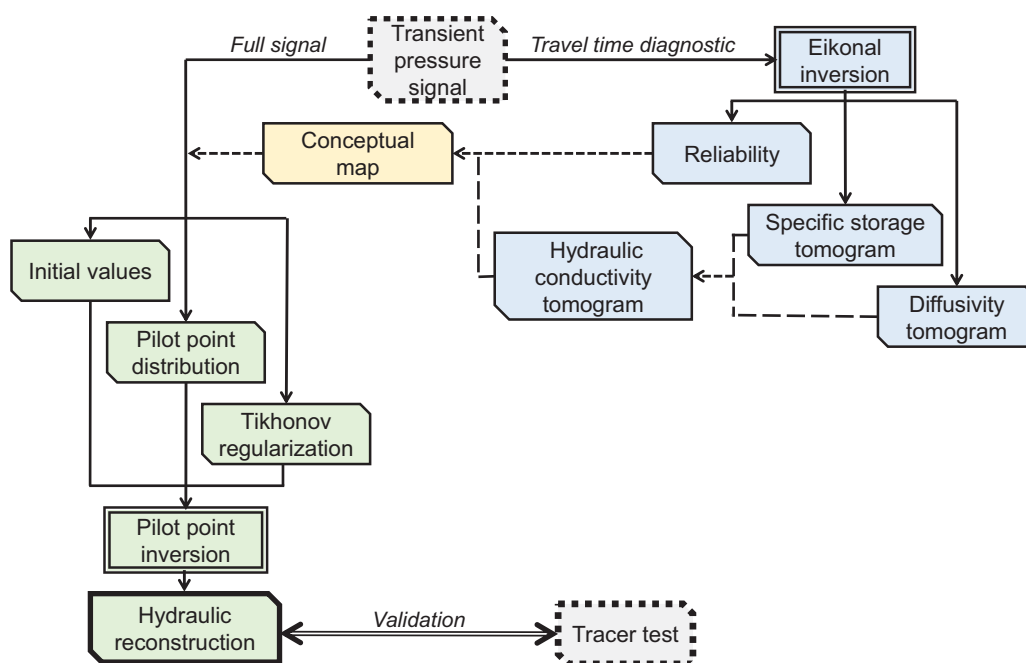


Figure 3. Main elements of sequential inversion procedure: the transient pressure signal is inverted by an eikonal-based approach to deliver a conceptual map. This structural information is used to constrain full pressure signal inversion based on pilot points.

techniques to fill a model parameter field. During the pilot point-based inversion procedure here, the information content of the conceptual maps is exploited as follows [Jiménez et al., 2013]:

Initial Values. Each individual pilot point is assigned an initial value equal to that of the corresponding cluster centroid.

Pilot Point Positioning. Higher pilot point density is favored at locations where parametric variability is suspected, i.e., at cluster boundaries. This step is performed using a finite element mesh generator. For each cluster a mesh is designed, and at each node a fixed pilot point is assigned. This procedure automatically leads to a refinement at the cluster boundaries.

Regularization. Interpolation among the pilot points is based on the conceptual maps as well. Three conditions are proposed, which have to be fulfilled so that two pilot points are correlated: (i) both pilot points pertain to the same cluster; (ii) the distance between the pair of pilot points is smaller than the average length of the cluster in horizontal direction; and (iii) there must be no other pilot point from a different cluster within a given space of influence [Jiménez et al., 2013]. For the regularization implementation, i.e., the spatial relationships among the pilot points, a graph theoretical concept is adopted [e.g., Bhark et al., 2011]. Initially, prior to the calibration, each pair of pilot points is examined and an adjacency matrix is developed:

$$\begin{pmatrix} 0 & \cdots & a_{ij} \\ \vdots & \ddots & \vdots \\ a_{ij} & \cdots & 0 \end{pmatrix} \leftarrow C(p_i, p_j) = \begin{cases} 1, & \text{conditions fulfilled} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $i=j=1, \dots$, number of pilot points, a_{ij} is a boolean indicator, p denotes a pilot point, and $C(\cdot)$ is the adjacency matrix. The adjacency matrix dictates if two pilot points are connected in a graph or not, based on the three conditions listed above, and it is used to calculate the regularization function, Φ_r .

Note that we want to arrive at a 3-D parameter field, but the conceptual map gives only insight into structures in a 2-D vertical slice between source and receiver well. For 3-D extrapolation, variograms along the horizontal and vertical axes are constructed from the eikonal-based diffusivity tomogram. Along the tomogram, pilot point values are considered as hard data, and the values for each cell of a given numerical model grid are derived from 3-D kriging.

The hydraulic parameters with spatial heterogeneity, which are addressed by the inversion procedure, are K and S_s . On one hand, the resulting K -field and S_s -field must honor the recorded pressure response data. This is evaluated by implementation in a flow model and comparison to the field data. On the other hand, the parameter fields are constrained by regularization. This is described by an objective function, which is solved based on Lagrange multipliers [Doherty, 2010; Jiménez et al., 2013].

2.3. Numerical Modeling

2.3.1. Hydraulic Test Simulation

Inversion methods such as the pilot point-based approach use a large number of iterative flow model runs. Therefore, any possibilities for minimizing simulation time are of interest. An appealing option is using locally refined grids, and accordingly hydraulic cross-well simulations were performed with MODFLOW-LGR [Mehl and Hill, 2005; Vilhelmsen et al., 2012], a transient, three-dimensional groundwater flow code. MODFLOW-LGR allows local refinement of a finite difference grid, as an extension to the classical MODFLOW code. It couples two or more finite difference grids called parent and child. A parent grid can cover a large area in order to accommodate regional flow and boundaries. A much more refined child grid can be used to study more local phenomena, for instance, the hydraulic effects in the vicinity of a pumping well.

2.3.2. Tracer Test Simulation

By numerical simulation of the tracer test with the reconstructed aquifer, the suitability of the hydraulic tomography approach for predicting solute transport can be evaluated. For this purpose, the reconstructed 3-D aquifer is implemented in a flow and transport model, and the simulated results are contrasted with those by a homogeneous case. The heterogeneous model represents the full heterogeneity in K and S_s obtained from the tomographical inversion procedure. In the homogeneous model, K and S_s are set constant, taking the mean of the estimated values. For flow modeling, the forward model used for pilot point-based full signal inversion is selected. The code MT3DMS [Zheng and Wang, 1999] is chosen for solving solute transport.

By comparing measured and, by these models, simulated tracer BTCs, we can assess the gain from resolving aquifer heterogeneity and also validate the inverted model. Still, additional parameters need to be specified before the transport models can be run. Crucial unknowns are dispersivity and effective porosity. Since these parameters cannot be determined separately with sufficient accuracy, we consider their possible value ranges, and estimate the most likely parameter ranges for homogeneous and heterogeneous models through a Bayesian approach, a Markov Chain Monte Carlo (MCMC) sampling procedure. It utilizes the Metropolis-Hastings algorithm [Metropolis et al., 1953; Hastings, 1970] to sample realizations of longitudinal dispersivity and effective porosity, separately for heterogeneous and homogeneous models. For simplification, transversal dispersivity is set 1/10 the value of the longitudinal one, which is a rough but common assumption in related work [Molina-Giraldo et al., 2011]. For generating new realizations within the MCMC framework, (i) one of the parameters is selected randomly, (ii) a new parameter value is proposed using a Gaussian random walk, and (iii) the acceptance ratio α is computed:

$$\alpha = \min \left(1, \frac{L(m^{new})}{L(m^{old})} \frac{g(m^{old} \rightarrow m^{new})}{g(m^{new} \rightarrow m^{old})} \right) \quad (2)$$

where L is the likelihood function, g is the proposed distribution, and m denotes a model parameterization. Finally, a random number u is drawn from a uniform distribution on $[0, 1]$ and the realization is accepted if $\alpha > u$, and rejected otherwise. As search criterion, the RMSE between measured and modeled tracer BTCs is selected. The function that maps from RMSE to likelihood is $L = 10^{-\frac{RMSE}{2\sigma^2}}$ with σ equal to 0.2.

3. Results

3.1. Eikonal-Based Inversion of Hydraulic Tests

Brauchler et al. [2013] reconstructed a diffusivity (D) and specific storage tomogram (S_s) utilizing the eikonal-based inversion approach. The derived tomograms displayed in Figures 4a–4e are shortly discussed in the following; however, for more details we refer to Brauchler et al. [2013].

For the eikonal-based inversion, a starting 2-D model domain of 45 cells was applied. Utilizing the method of staggered grid, the mesh was shifted four times in the horizontal direction and three times in vertical

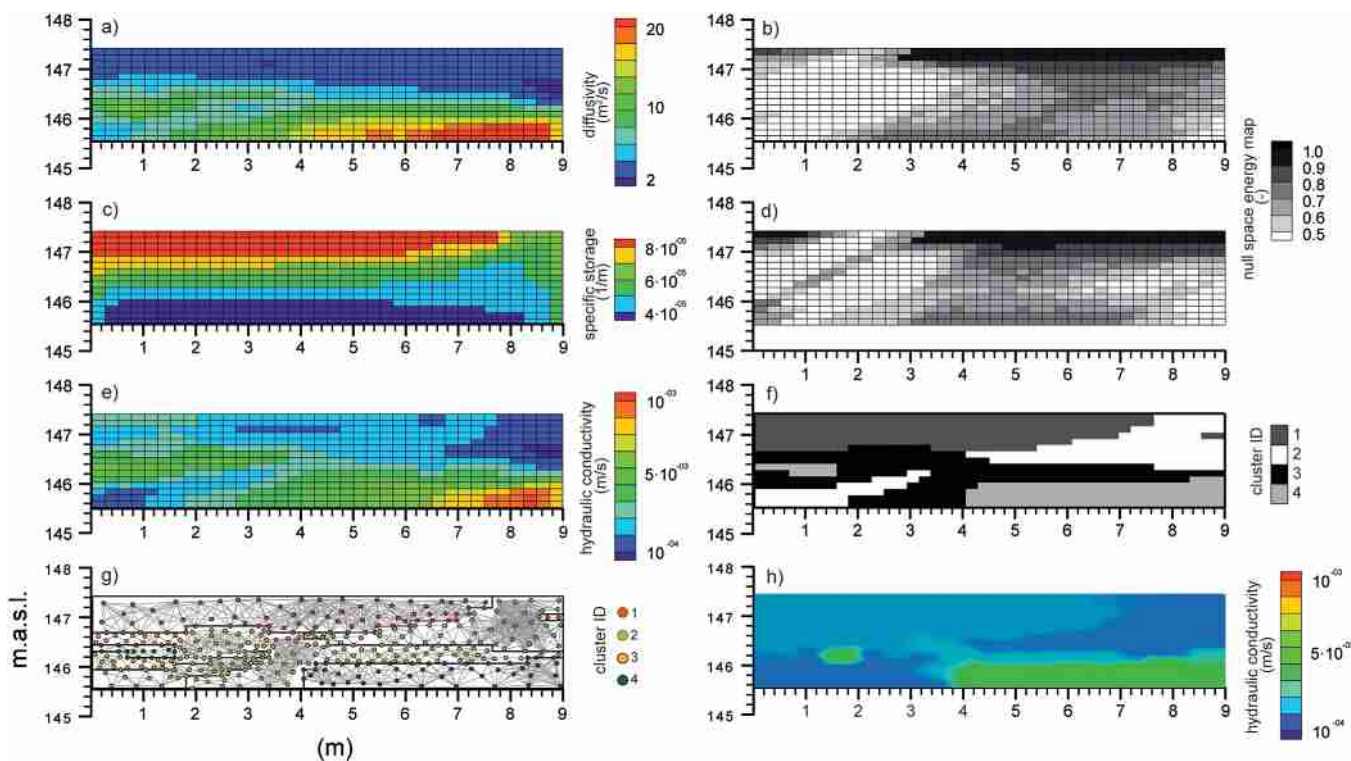


Figure 4. (a and b) Reconstructed diffusivity tomogram and the associated null-space energy map. (c and d) Reconstructed specific storage tomogram and the associated null-space energy map. (e) Computed hydraulic conductivity tomogram using $D = K/S_s$ [see Brauchler et al., 2013]. (f) Resulting cluster distribution based on hydraulic conductivity tomogram. (g) Pilot point distribution, cluster IDs, and Tikhonov regularization connections (gray lines). (h) Resulting hydraulic conductivity from pilot points inversion using K and S_s as a parameter (longitudinal slice of the 3-D domain, Figure 8b, for comparison purposes).

direction, which led to the final resolution of 540 pixels imaged in Figure 4. For the diffusivity reconstruction displayed in Figure 4a, the 50% travel time diagnostic was employed for the inversion. The travel time and attenuation inversion, including the staggered grid calculation, took less than 1 min on a conventional notebook. For quantifying the reliability of the D -tomograms and S_s -tomograms (Figures 4a and 4c), null-space energy maps are provided in Figures 4b and 4d. These maps illustrate the uncertainty associated with the eikonal-based inversion. A value of 1 of the null-space (black color) means the lowest possible confidence on the values obtained for a given cell in a tomogram, and a value of 0 (white color) indicates highest confidence.

The D -tomograms and S_s -tomograms indicate horizontal layering (Figures 4a and 4c), with values of D between 2 and 20 m^2/s , and of S_s between 3×10^{-5} and 10^{-4} m^{-1} . The K -tomogram (Figure 4e) is obtained by $D = K/S_s$. It shows a similar structure, with moderate heterogeneity and values ranging from $K = 10^{-4}$ to 10^{-3} m/s , which are considered typical for sand and gravel aquifers.

3.2. Conceptual Map and Pilot Points Configuration

The conceptual map to support the subsequent pilot point inversion was developed from the eikonal-based results and is displayed in Figure 4f. The number of clusters was set to four, which, after visual inspection, was considered the maximum possible to keep the main structures of the tomograms. Following Jiménez et al. [2013], cells associated with high null-space energy values (here >0.9) were ignored during clustering. These gap cells, which are mainly located close to the top boundary, were filled by nearest neighbor interpolation from adjacent cells.

The conceptual map is used to guide pilot point positioning and setting initial values for hydraulic parameter (K and S_s) calibration. This involves transforming the conceptual map into a vector graph in order to obtain a digital image of the cluster interfaces. The latter are displayed as black lines in Figure 4g. Then a mesh is assigned to the vector graph using a finite element mesh generator. The mesh nodes are translated

Table 3. Number of Pilot Points and Connections for Each Cluster

Cluster ID	Number of Pilot Points	Number of Connections
1	75	358
2	96	301
3	125	369
4	67	238
Total	363	1266

into pilot point positions. Each pilot point an initial value equal to the value of the corresponding cluster centroid is assigned.

The selected mesh generator uses Delaunay triangulation. The utilization of the mesh generator leads to a higher pilot point density along the cluster boundaries. This distribution is favorable because along these boundaries the largest contrasts in hydraulic properties are expected. A maxi-

imum element size of 0.7 m and a resolution curvature of 0.3 are selected. Both parameters control the mesh (i.e., pilot point) density and how fast it declines away from the cluster boundaries. The maximum element size determines how big each grid element can be and the resolution curvature limits the mesh size along a curved boundary [COMSOL Inc., 2012]. Accordingly, the lower the values for each parameter are, the more pilot points are allocated. The maximum number of nodes is set to 1200. Table 3 lists the number of pilot points assigned to each cluster. The maximum element number, which controls the total number of pilot points, is a compromise between available computational resources, available observations and desired resolution.

The regularization step for interpolation among the pilot points is also guided by the conceptual maps. Based on their neighborhood relationships, pilot points are connected in pairs, yielding separate networks that delineate the structures observed in the conceptual maps. In Figure 4g, these networks are illustrated as gray lines. For the maximum distance between a pair of pilot points 3 m were chosen, which equals the average length of the clusters in horizontal direction, and for the calculation of the space of influence an angle of 30° was set [e.g., Jiménez et al., 2013].

For interpolating between the pilot points and lateral extrapolation in 3-D, ordinary kriging is applied. The diffusivity tomogram displayed in Figure 4a is utilized to derive the underlying semivariograms depicted in Figure 5. As expected for fluvial sediments, we find a larger range in the horizontal (2.3 m) than in the vertical direction (1 m). During the subsequent calibration, the K and S_s values at the pilot point cells are tuned, and by kriging the values of the other cells are filled. In the numerical model, pilot points exist only in the vertical tomogram slice between the source and receiver well. For lateral extrapolation, we assume stationary, horizontally isotropic geostatistical properties and consistently use kriging with the same semivariograms. However, it is clear that due to the missing field data away from the well couple, the resulting 3-D field will lose reliability in lateral distance. Alternatively, several tomograms from different source and receiver wells may be collected from the field and combined for 3-D inversion, such as shown in Berg and Illman [2012].

3.3. Simulation of Pumping Tests

In the numerical groundwater flow model domain, the refined child grid (13 m × 4 m × 2 m) is embedded in the coarser parent grid (60 m × 60 m × 2 m). The selected refinement ratio between parent and child

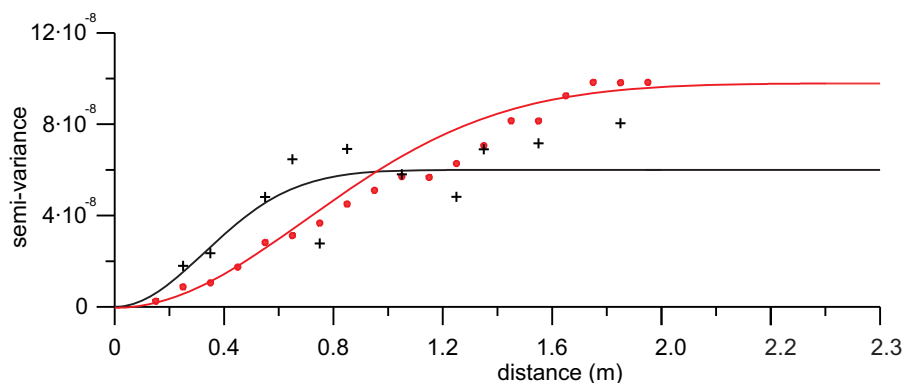


Figure 5. Semivariogram derived from the diffusivity tomogram with horizontal (red) and vertical search direction (black).

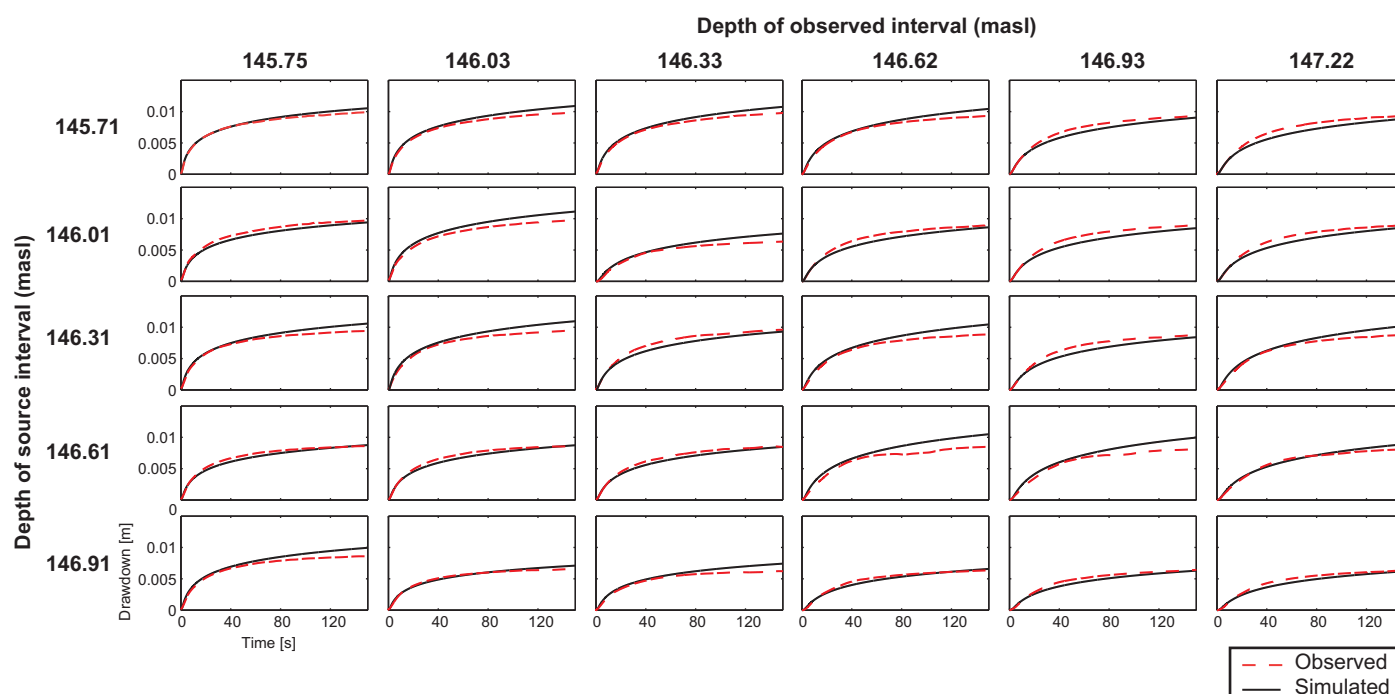


Figure 6. Observed and modeled pressure responses by adjusting only hydraulic conductivity (K) during 3-D pilot point-based inversion.

grid is 5:1, resulting in total around 1 million cells. Fixed hydraulic heads are implemented to enforce the regional hydraulic gradient (0.004) measured in the field.

The child grid simulates the source (P0/M25) and receiver well (P5/M17.5) from the field test. Along the source borehole, five vertical screens are implemented. This resembles those screens in the field at which the pumping tests were performed. At the receiver an array of six observation points is defined in the model to simulate the multichamber well configuration.

For the calibration, the repeated pumping tests at the source well are simulated by transient flow modeling. Each test simulates a 140 s pumping at 0.3 L/s. With five tests at different depths and six receivers, 30 pressure response curves are recorded in total. Each pressure signal is discretized by 300 points. Two choices are compared for fitting these curves to those measured at the field site. In the first one, only K is calibrated, and S_s is set a constant value of $S_{s,mean} = 5.5 \times 10^{-4} \text{ m}^{-1}$ equal to the mean of the tomogram (Figure 4c). The initial values for K are specified equal to the centroids of the clusters. In the second one, both parameters are independently adjusted, which means that the number of decision variables is doubled. However, to save computational time, here the initial values of K are selected according to the results from previous calibration of K only.

The selected optimization algorithm (Levenberg-Marquardt) is suited for parallelization, and a cluster with a desktop (Intel i7 3.4 GHz, 16 GB RAM) and a workstation (Intel Xeon E5 3.1 GHz, 64 GB RAM) was used. The number of parallel runs was 20, distributed on both machines. A single model run took approximately 5 min. A total of 1130 models runs were needed. The full pilot points inversion took approximately 5 h.

3.4. Calibrated Hydraulic Parameter Fields

The hydraulic parameters at the Stegemühle site show low variability in comparison with the conditions at other test sites such as NCRS, where similar experiments were conducted [e.g., Berg and Illman, 2012]. This is reflected in the pressure response curves (Figure 6) which all show a similar behavior. Despite that, in order to be able to resolve heterogeneous structures, we suggest to make full use of the measured information while exploiting the degrees of freedom in the hydraulic model. This means, the calibration procedure is applied to fit all and the complete pressure response curves, and this is achieved by not only calibrating the spatial distribution of K but also S_s .

Figure 6 compares the 30 measured pressure response curves with those calibrated by K adjustment only. The fitting error (root-mean-squared error, RMSE) is minimized to $5 \times 10^{-4} \text{ m}$, and it is shown that most

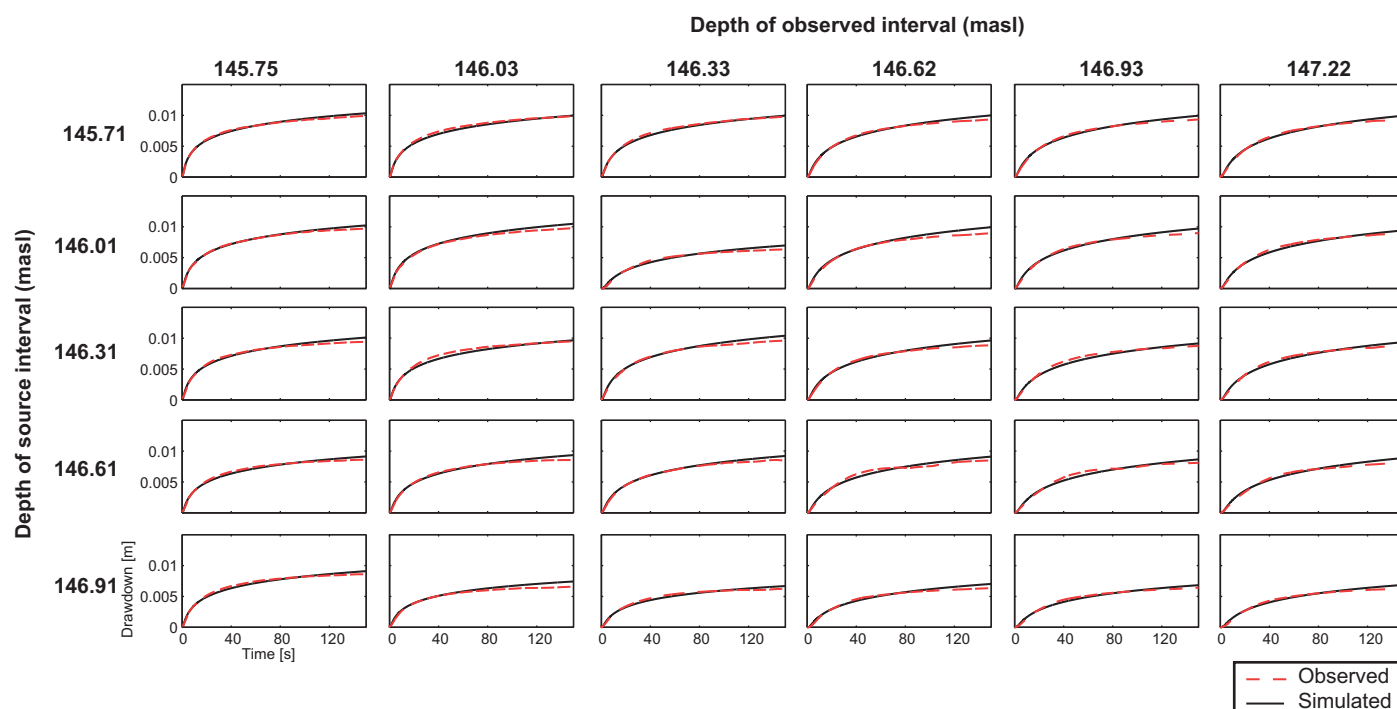


Figure 7. Observed and modeled pressure responses by adjusting hydraulic conductivity (K) and specific storage (S_s) during 3-D pilot point-based inversion.

curves are properly reproduced. This fitting error seems to be the lowest possible with the current parametrization. In some cases, the later stages are not ideally captured. This can be seen for most of the responses from the fifth interval at the receiver. This is improved by also including S_s as a decision variable. The resulting modeled pressure curves, as depicted in Figure 7, fit better to the measurements, and we reduce the RMSE to 3×10^{-4} m.

The resulting K -fields are visualized in Figure 8. Kriging variance or estimation error of the hydraulic parameters assigned to cells increases rapidly from the vertical slice between source and receiver well, which contains the pilot points. Hence, to those cells surrounding the shown central region, mean values are assigned (also for Figure 9). Including S_s in the inversion yields a very similar field, and only a central layer with higher K found in Figure 8a appears less accentuated in Figure 8b. Figure 9 depicts the calibrated distribution of S_s . It is revealed that highest values of around $7 \times 10^{-5} \text{ m}^{-1}$ are characteristic for the upper part of the aquifer, whereas the S_s in the lower ranges around $4 \times 10^{-5} \text{ m}^{-1}$. Comparison of Figures 8 and 9 nicely shows how the structures are related. With regularization and kriging, these clusters stimulate the calibration of zones that can be interpreted as individual sedimentary hydrofacies [e.g., Bayer *et al.*, 2011]. The latter are characterized by specific and fairly constant hydraulic properties, and this is reproduced here by the spatial correlation between the structures for K and S_s in Figures 8 and 9.

The vertical parameter distribution between source and receiver well in the 3-D aquifer model can be compared with the tomograms reconstructed based on travel time diagnostics (Figure 4). It is shown that the basic layer structure is maintained, with higher K values on the lower-right section and low values in the upper section. Same geometries can be recognized in both cluster map and pilot point-based field (Figures 4f and 4h). In comparison with the travel time-based K tomogram (Figure 4e), full signal inversion yields a decrease in the range of K . This is also true for the specific storage ranges. In comparison with the S_s -tomogram (from 4×10^{-5} to $8 \times 10^{-5} \text{ 1/m}$) (Figure 4c), a lower variability is observed in Figure 9 (from 4×10^{-5} to $7 \times 10^{-5} \text{ 1/m}$).

3.5. Validation of the Reconstructed Aquifer With Tracer Test Data

In order to validate the reconstructed hydraulic parameter fields for predicting solute transport, the uranine and sodium-naphthionate tracer tests are used. Applying tracer data to validate an inversion procedure

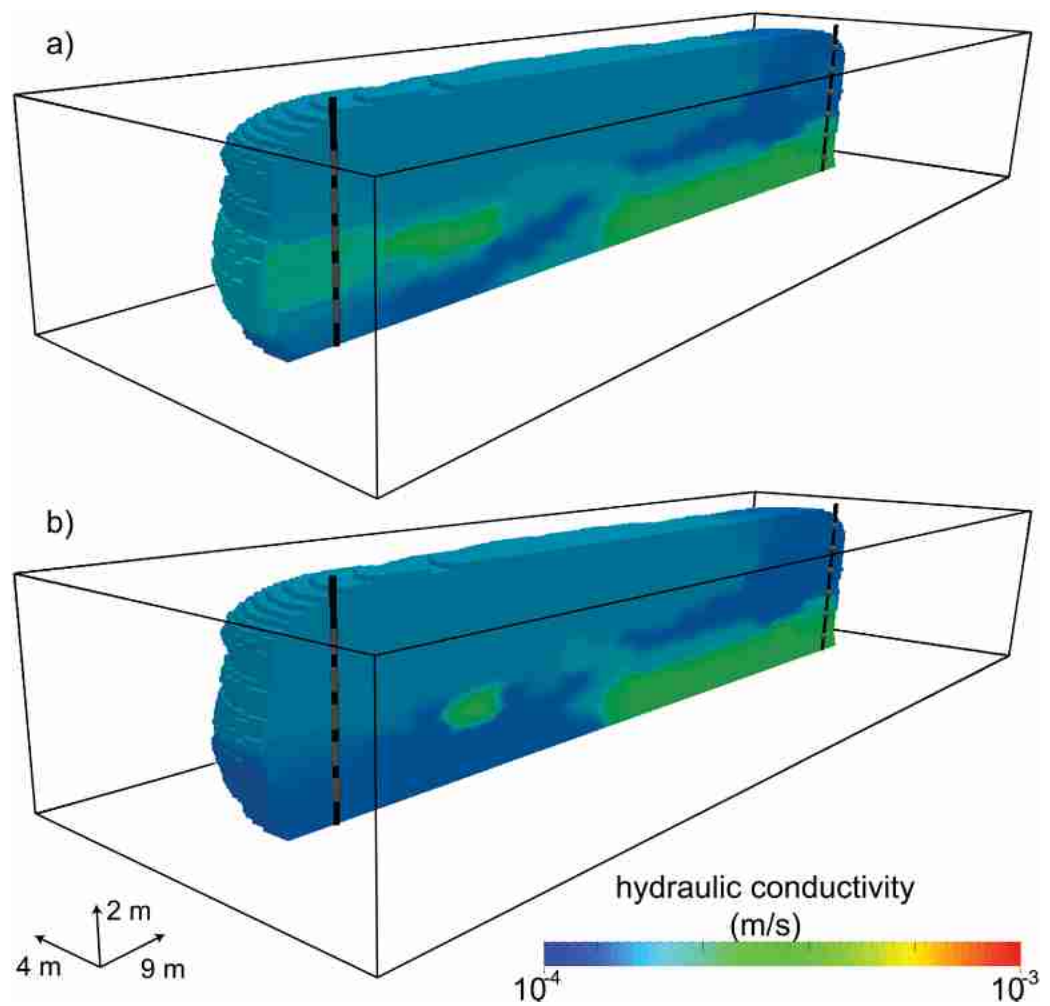


Figure 8. Reconstructed hydraulic conductivity of aquifer using (a) hydraulic conductivity and (b) hydraulic conductivity and specific storage.

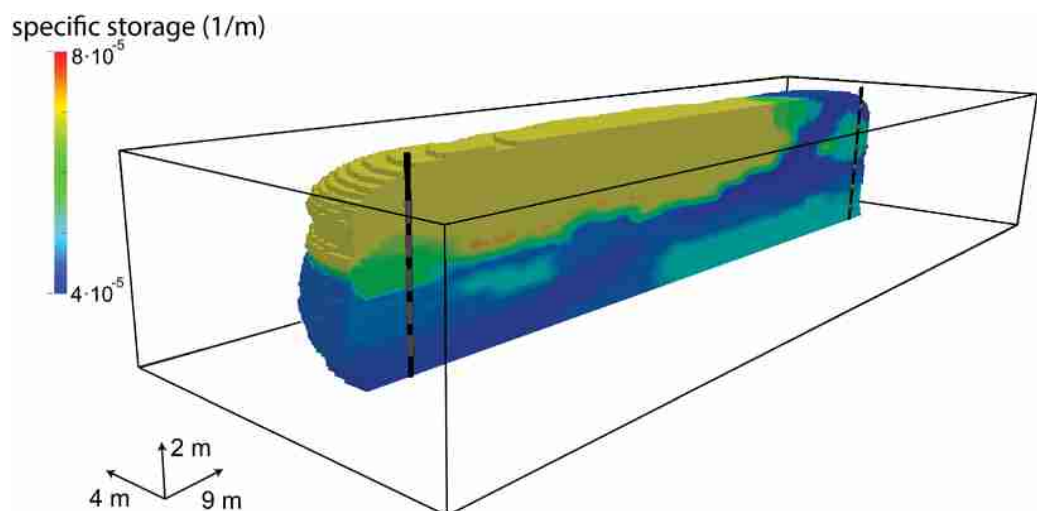


Figure 9. Reconstructed specific storage field using hydraulic conductivity and specific storage tomograms.

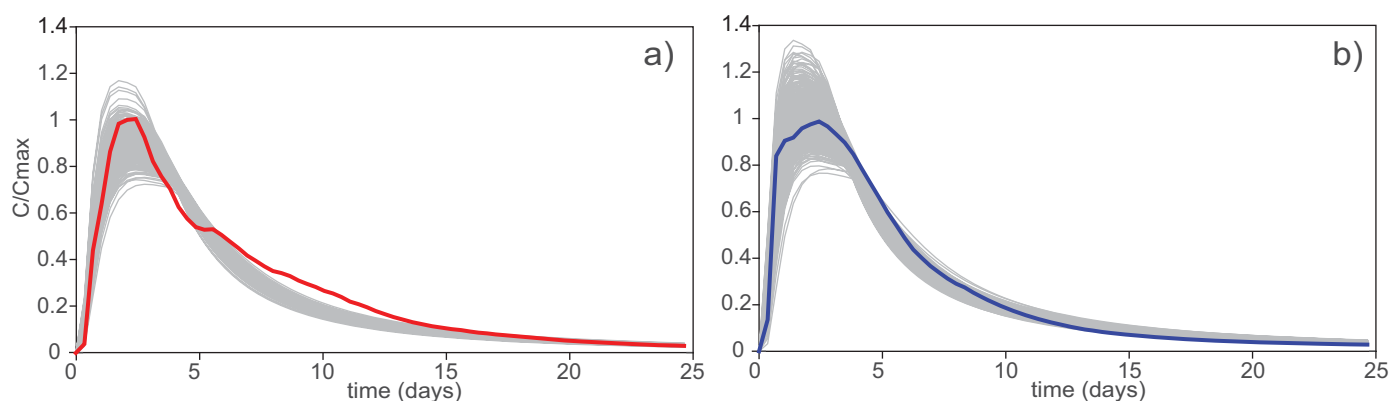


Figure 10. Breakthrough curves (BTCs) measured at the pumping well during the tracer tests and MCMC realizations for the (a) uranine and (b) sodium-naphthionate.

based solely on hydraulic data poses evident challenges. Tracer transport is not fully predictable if only hydraulic data are used. Moreover, the tracer test was performed under a different hydraulic regime, and therefore comparison between tracer simulation and measurements will elucidate the robustness of the hydraulic inversion.

As explained, the two tracers were applied between the source and receiver wells used for the hydraulic inversion. The measured BTCs of the tracers are illustrated in Figure 10. The tracers were monitored for 2 months, but here only the time of breakthrough is shown. A first visual inspection reveals that the curves follow a nearly ideal shape with early steep increase of concentration and, after a peak is reached, tailing sets in. With a closer look, we also recognize nonuniformities in both curves. The sodium-naphthionate BTC shows a small step in the later phase after the peak has passed by. In contrast, before uranine reaches the maximum peak, an apparent local peak already appears which widens the period when the highest concentration is detected.

Our main question is whether the reconstructed heterogeneity is accurate enough for predicting the transport of the tracers. Aside from this, we also ask if this complexity is needed at all. Therefore, results for the reconstructed aquifer are compared to the simplest reference, which is simulation with a homogeneous system. The main steps are implementation in a flow and transport model, specification of unknown transport parameters, and comparison of both model results.

For simulation of the groundwater flow velocity field, the same flow model setup as for the pilot points based inversion was used. The models make use of a 3-D grid in order to accurately capture the heterogeneity of the aquifer and to account for potential transversal spreading of the tracer. Several authors recognize the importance of 3-D models [e.g., Liu *et al.*, 2007]. For example, Illman *et al.* [2008] state “the knowledge of detailed 3-D distributions of K is critical in prediction of contaminant transport.” Steady state conditions are assumed according to the static hydraulic settings during the experiment. The fixed head boundary conditions establish a constant regional flow field, and the pumping well P0/M25 (Figure 2) is configured with an extraction rate of 0.3 L/s to simulate the forced gradient conditions generated in the field. In the heterogeneous model, the reconstructed K -field was implemented, in the homogeneous one, the arithmetic mean of 1.5×10^{-4} m/s of the heterogeneous variant was chosen. By using steady state models, the inverted S_s -fields are not utilized. However, the K -fields are derived by coupled inversion of K and S_s values under transient conditions. This way, the information content of the transient hydraulic experiment is exploited for the steady state flow simulation during the tracer test.

The transport code MT3DMS [Zheng and Wang, 1999] is selected for solving solute transport. The transport model domain covers $10 \text{ m} \times 4 \text{ m} \times 2 \text{ m}$ with a spatial discretization of $160 \times 32 \times 40$ cells, summing up to 204,800 cells. Computational time on a 2015 desktop (i7, 16 GB RAM, 250 GB SSD) was approximately 5 min per model run using the method of characteristics.

The value ranges of two unknown parameters for transport modeling, dispersivity and effective porosity needed to be estimated. For this, two MCMC chains were run, one for the homogeneous model, and other

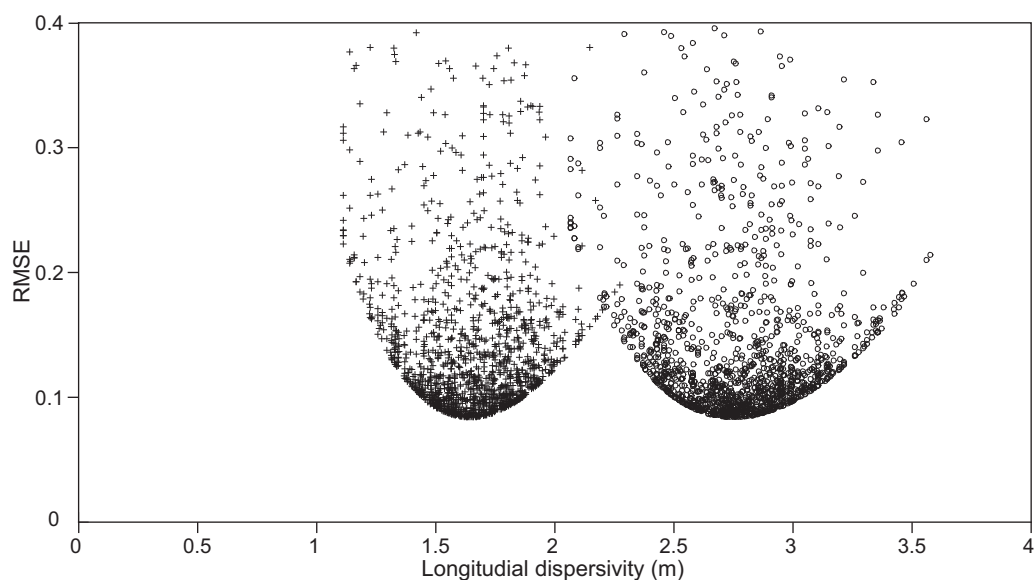


Figure 11. Ensemble of MCMC-based results of longitudinal dispersivities for the homogeneous (circle) and reconstructed heterogeneous model (cross); the RMSE denote the discrepancy of measured and simulated BTCs. The spread of RMSE for a certain dispersivity value highlights that realizations with same dispersivity but different effective porosity values exist.

for the heterogeneous model. Both BTCs are adjusted and a summed up RMSE is computed. Each chain has a length of 3000 model runs, with a burn-in of 1500. BTCs measured at the pumping well during the tracer tests are compared with those resulting from the simulation with the obtained MCMC ensemble. In Figure 10, only those for the heterogeneous model are depicted. The BTCs simulated with the homogeneous model are comparable and not shown here. For both the homogeneous and the heterogeneous model, the BTCs obtained with different dispersivity and effective porosity realizations spread around the observed BTC, and no bias or other systematic error is observed.

The nearly ideal shape of the BTCs indicates that the aquifer exhibits only a moderate heterogeneity, and thus minor differences among homogeneous and heterogeneous model results exist. The homogeneous variant can appropriately capture the general form of the BTCs given properly tuned dispersivity and effective porosity values [see also *Ni et al.*, 2009]. The heterogeneous variant is similarly suitable, which however also means that special characteristics of the BTCs are not resolved. This can be attributed to the limited resolution of the tomograms and the reconstructed fields in order to delineate local hydraulic heterogeneities relevant for solute transport. However, the limitation of an even very detailed reconstruction of aquifer heterogeneity based on a single source-receiver plane certainly plays a major role, because the tracer transport occurs in a 3-D domain. Therefore, the discrepancy between the BTCs and the heterogeneous model simulation may also be caused by the applied lateral extrapolation from the vertical source-receiver plane. For example, the untypical spreading of the uranine BTC around the peak may be due to unseen lateral aquifer heterogeneity that strongly sidetracks the tracer. The irregular behavior of the measured sodium-naphthionate curve after around 5 days may indicate that a portion of the tracer is temporary split apart from the main plume and reaches the pumping well with a time lag. This is observed as local peak and accentuates the tailing of the earlier main tracer mass fraction. As our 3-D reconstruction is only based on extrapolation from one profile, more adjacent and ideally differently oriented profiles would be needed for capturing such lateral heterogeneities.

Lateral tracer loss is also supported by the mass recovery rates of 86% (uranine) and 50% (sodium-naphthionate). The relative low recovery observed for sodium-naphthionate in comparison to uranine during the test cannot be definitively attributed to a certain reason. As the shape of the tracer BTC does not support distinctive retardation of the tracer, sorption processes are unlikely the reason. Due to the relatively long duration of the test and moderate groundwater temperatures (10–12°C), a microbiological degradation of sodium-naphthionate in the aquifer seems possible. Rapid microbiological sodium-naphthionate degradation was observed by *Goldscheider et al.* [2001] for water samples with a certain storage period, depending

on storage temperature. Studies for those kinds of processes are scarce and more research effort in this direction is needed. Nevertheless, as normalized BTCs are employed for this study, a lower tracer recovery does not compromise the results.

Although, the reconstructed model is not superior to a simple homogeneous alternative in delineating the tracer BTCs, it appropriately resolves structures relevant for the transport of the tracers. This is revealed by comparison of the estimated value ranges for the MCMC ensembles. Figure 11 shows that the RMSE for both model variants are similar, and the effective porosities range between 0.1 and 0.27. This is the same for both, and also within the broad range of previously reported values of 0.10–0.25, [Schlie, 1989; Hu, 2011; Meyer, 2011; D. Meischner, Natürliches Einzugsgebiet und Trinkwasserschutzzonen für das Wasserwerk Stegemühle der Stadtwerke Göttingen AG, Survey for Stadtwerke Göttingen AG, unpublished data, 1985]. However, dispersivity values need to be substantially higher when a homogeneous model is used. Best results are obtained for a longitudinal dispersivity of $\alpha_L = 2.67$ m for the homogeneous model. The optimal fit for the heterogeneous case is at $\alpha_L = 1.64$ m, which shows that macrodispersive effects are simulated explicitly and correctly through the reconstructed macroscale hydraulic heterogeneity. The value of $\alpha_L = 1.64$ m is still significant and this denotes that heterogeneities exist at a smaller scale than the resolution of the hydraulic tomography at this site and with this experimental configuration, which strongly influence the tracer spreading. By individual tracer BTC fitting, the estimated values of $\alpha_L = 1.57$ m for uranine and 1.72 m for sodium-naphthionate slightly deviate from the result of combined fitting. These differences are not judged as significant enough to identify clear differences in the tracer-specific transport or associated with the different injection levels.

4. Conclusions

The presented work shows that the sequential travel time and pilot point-based approach can be applied to high-resolution reconstruction of hydraulic parameters at a field site. It is demonstrated, for the only slightly heterogeneous field site that the presented procedure can identify sedimentary structures. However, comparison of model-based predictions with the tracer tests at the site reveals that the reliability of the derived aquifer model also exhibits limitations.

The tracer test delivered two BTCs, which show minor irregularities and this indicates the only moderate heterogeneity at the Stegemühle site. Therefore, even a homogeneous model can provide similarly good predictions as a heterogeneous variant with the reconstructed K -field. A main point is that crucial transport parameters, especially dispersivity, need to be set. We have not prespecified these parameters but analyzed suitable value ranges by applying a MCMC based search. In other words, for minimizing any bias we examined model validity within these ranges. Within these degrees of freedom, the reconstructed model performs similarly well as a homogeneous one. This reflects that even though macroscale heterogeneities are reconstructed, their combined effect on tracer spreading averages. Therefore, the tracer breakthrough curves can also be predicted by a higher integral dispersivity in a much simpler homogeneous model. However, as pointed out in the theoretical study by *Ni et al.* [2009], even if a homogenous model can provide an appropriate fit, it will not capture the scale-dependent increase of dispersion with transport distance [see, e.g., *Molina-Giraldo et al.*, 2011; *Gelhar et al.*, 1992]. In contrast, by reconstructing transport-relevant structures, their effect on macrodispersion is explicitly simulated, and thus the heterogeneous model is more suited for predicting solute transport along shorter or longer distances.

Still, in our application case, neither the homogeneous nor heterogeneous model variant perfectly predicts the recorded tracer concentrations. When measurement errors can be neglected, we interpret inconsistencies caused by unresolved lateral heterogeneity. The proposed sequential approach employs 3-D hydraulic simulation and inversion, but structures are constrained only by the vertical 2-D travel time tomograms. For improved structural reconstruction, additional tomograms between different source and receiver wells would be needed. With these, more reliable results from hydraulic parameter interpolation rather than the presented extrapolation can be expected. Aside from this, as *Illman et al.* [2012] demonstrate in a sandbox experiment, the resolution by HT could be improved with a higher density of sources and receivers. As a result, so far unresolved microstructures could be detected and the value of the dispersivity would be further decreased.

The presented coupled inversion procedure shows to further refine K -tomograms and S_s -tomograms between the investigated source and receiver wells in comparison to travel time-based inversion only. A main observation is that pilot point-based inversion reduces heterogeneity, although homogenization is

not enforced through regularization. This may be due to the fact that the 2-D travel time tomograms are based on a diagnostic of early arrival times, which are accentuated by the existence of high K zones or preferential flow paths. In contrast, the pilot point approach calibrates the full pressure response curves and calibrates a 3-D model, and by this a higher volume of the aquifer is referred to. Further insight could be obtained, for instance, by employing different parts of the response curves for pilot point-based inversion.

As an innovative step, it is shown that including S_s in addition to K in the pilot point-based inversion is beneficial for minimizing model misfit to field data. This has rarely been included in related work [e.g., Castagna and Bellin, 2009]. However, this means also doubling the number of decision variables for the optimization problem. This is potentially not desirable, as this eventually can overparameterize the problem and enhance its ill posedness. In fact, the generated heterogeneous aquifer model can be considered as one solution of many, and further alternative realizations fitting the data could be explored. As future work, we therefore plan to envisage the diversity of several equally probable realizations, based on K with or without S_s as free parameters.

Acknowledgments

The investigations were conducted with the financial support of the Swiss National Science Foundation to the project "A field assessment of high-resolution aquifer characterization: An integrated approach combining hydraulic tomography and tracer tomography" under grant 200021_140450/1 and the CCES funded project "RECORD Catchment." The helpful comments of the Associate Editor and three reviewers are greatly appreciated. Further thanks go to Gabi Moser for language corrections. All data are available by e-mailing the corresponding author (santos.jimenez@erdw.ethz.ch).

References

- Alexander, M., S. J. Berg, and W. A. Illman (2011), Field study of hydrogeologic characterization methods in a heterogeneous aquifer, *Ground Water*, 49(3), 365–382.
- Bayer, P., P. Huggenberger, P. Renard, and A. Comunian (2011), Three-dimensional high resolution fluvio-glacial aquifer analog. Part 1: Field study, *J. Hydrol.*, 405, 1–9.
- Berg, S. J., and W. A. Illman (2011a), Capturing aquifer heterogeneity: Comparison of approaches through controlled sandbox experiments, *Water Resour. Res.*, 47, W09514, doi:10.1029/2011WR010429.
- Berg, S. J., and W. A. Illman (2011b), Three-dimensional transient hydraulic tomography in a highly heterogeneous glaciofluvial aquifer-aquitard system, *Water Resour. Res.*, 47, W10507, doi:10.1029/2011WR010616.
- Berg, S. J., and W. A. Illman (2012), Field study of subsurface heterogeneity with steady-state hydraulic tomography, *Ground Water*, 51(1), 29–40.
- Berg, S. J., and W. A. Illman (2015), Comparison of hydraulic tomography with traditional methods at a highly heterogeneous site, *Ground Water*, 53(1), 71–89.
- Bhark, E. W., B. Jafarpour, and A. Datta-Gupta (2011), A generalized grid connectivity-based parameterization for subsurface flow model calibration, *Water Resour. Res.*, 47, W06517, doi:10.1029/2010WR009982.
- Bohling, G. C., X. Zhan, J. J. Butler Jr., and L. Zheng (2002), Steady shape analysis of tomographic pumping tests for characterization of aquifer heterogeneities, *Water Resour. Res.*, 38(12), 1324, doi:10.1029/2001WR001176.
- Bohling, G. C., J. J. Butler Jr., X. Zhan, and M. D. Knoll (2007), A field assessment of the value of steady shape hydraulic tomography for characterization of aquifer heterogeneities, *Water Resour. Res.*, 43, W05430, doi:10.1029/2006WR004932.
- Böhm, G., and A. Vesnaver (1996), Relying on a grid, *J. Seismic Explor.*, 5, 169–184.
- Brauchler, R., R. Hu, T. Vogt, D. Al-Halbouni, T. Heinrichs, T. Ptak, and M. Sauter (2010), Cross-well slug interference tests: An effective characterization method for resolving aquifer heterogeneity, *J. Hydrol.*, 384(1–2), 33–45.
- Brauchler, R., J. Doetsch, P. Dietrich, and M. Sauter (2012), Derivation of site-specific relationships between hydraulic parameters and p-wave velocities based on hydraulic and seismic tomography, *Water Resour. Res.*, 48, W03531, doi:10.1029/2011WR010868.
- Brauchler, R., R. Hu, L. Hu, S. Jiménez, P. Bayer, P. Dietrich, and T. Ptak (2013), Rapid field application of hydraulic tomography for resolving aquifer heterogeneity in unconsolidated sediments, *Water Resour. Res.*, 47, W03503, doi:10.1029/2010WR009635.
- Cardiff, M., W. Barrash, and P. K. Kitanidis (2012), A field proof-of-concept of aquifer imaging using 3-D transient hydraulic tomography with modular, temporarily-emplaced equipment, *Water Resour. Res.*, 48, W05531, doi:10.1029/2011WR011704.
- Cardiff, M., W. Barrash, and P. K. Kitanidis (2013), Hydraulic conductivity imaging from 3-D transient hydraulic tomography at several pumping/observation densities, *Water Resour. Res.*, 49, 7311–7326, doi:10.1002/wrcr.20519.
- Castagna, M., and A. Bellin (2009), A Bayesian approach for inversion of hydraulic tomographic data, *Water Resour. Res.*, 45, W04410, doi:10.1029/2008WR007078.
- COMSOL Inc. (2012), *COMSOL Multiphysics, Version 4.3, User's Guide and Reference Guide*, Burlington, Mass.
- Doherty, J. (2010), *PEST, Model-Independent Parameter Estimation. User Manual*, 5th ed., Watermark Numer. Comput., Brisbane, Queensland, Australia.
- Einarson, M. D., and J. A. Cherry (2002), A new multilevel ground water monitoring system using multichannel tubing, *Ground Water Monit. Rem.*, 22(4), 52–65.
- Gelhar, L. W., C. Welty, and K. R. Rehfeldt (1992), A critical review of data on field-scale dispersion in aquifers, *Water Resour. Res.*, 28(7), 1955–1974.
- Goldscheider, N., H. Hötzel, and K. Kottke (2001), Microbiological decay of Naphthionate in water samples as a source of misinterpretation of tracer tests, paper presented at XXXI IAH Congress Munich 2001.
- Gottlieb, J., and P. Dietrich (1995), Identification of the permeability distribution in soil by hydraulic tomography, *Inverse Probl.*, 11, 353–360.
- Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57(1), 97–109.
- Hu, R. (2007), *Einrichtung eines Naturmessfelds, hydraulische und hydrogeochemische Untersuchungen, Transportmodellierung*, Bachelor thesis, Geosci. Cent., Univ. of Göttingen, Göttingen, Germany.
- Hu, R. (2011), *Hydraulic tomography: A new approach coupling hydraulic travel time, attenuation and steady shape inversions for high-spatial resolution aquifer Characterization*, PhD thesis, Geosci. Cent., Univ. of Göttingen, Göttingen, Germany.
- Huang, S.-Y., J.-C. Wen, T.-C. J. Yeh, W. Lu, H.-L. Juan, C.-M. Tseng, J.-H. Lee, and K.-C. Chang (2011), Robustness of joint interpretation of sequential pumping tests: Numerical and field experiments, *Water Resour. Res.*, 47, W10530, doi:10.1029/2011WR010698.
- Illman, W. A., X. Liu, and A. Craig (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.

- Illman, W. A., A. J. Craig, and X. Liu (2008), Practical issues in imaging hydraulic conductivity through hydraulic tomography, *Ground Water*, 46(1), 120–132.
- Illman, W. A., X. Liu, S. Takeuchi, T.-C. J. Yeh, K. Ando, and H. Saegusa (2009), Hydraulic tomography in fractured granite: Mizunami underground research site, Japan, *Water Resour. Res.*, 45, W01406, doi:10.1029/2007WR006715.
- Illman, W. A., J. Zhu, A. J. Craig, and D. Yin (2010), Comparison of aquifer characterization approaches through steady state groundwater model validation: A controlled laboratory sandbox study, *Water Resour. Res.*, 46, W04502, doi:10.1029/2009WR007745.
- Illman, W. A., S. J. Berg, and T.-C. J. Yeh (2012), Comparison of approaches for predicting solute transport: Sandbox experiments, *Ground Water*, 50(3), 421–431.
- Jiménez, S., R. Brauchler, and P. Bayer (2013), A new sequential procedure for hydraulic tomographic inversion, *Adv. Water Resour.*, 62, 59–70.
- Leibundgut, C., P. Maloszewski, and C. Külls (2009), *Tracers in Hydrology*, John Wiley.
- Liu, X., W. A. Illman, A. J. Craig, J. Zhu, and T.-C. J. Yeh (2007), Laboratory sandbox validation of transient hydraulic tomography, *Water Resour. Res.*, 43, W05404, doi:10.1029/2006WR005144.
- Lochbühler, T., J. Doetsch, R. Brauchler, and N. Linde (2013), Structure-coupled joint inversion of geophysical and hydrological data, *Geophysics*, 78(3), ID1–ID14.
- Mariethoz, G., P. Renard, and J. Caers (2010), Bayesian inverse problem and optimization with iterative spatial resampling, *Water Resour. Res.*, 46, W11530, doi:10.1029/2010WR009274.
- Mehl, S. W., and M. C. Hill (2005), the U.S. Geological Survey modular ground-water model—Documentation of shared node local grid refinement (LGR) and the Boundary Flow and Head (BFH) Package, *U.S. Geol. Surv. Tech. Methods*, 6-A12, 68 pp.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21(6), 1087–1092.
- Meyer, R. (2011), Aufbau und Einsatz eines online Fluorimeters zur Bestimmung hydrogeologischer Parameter des Testfeldes Stegemühle, Bachelor thesis, Geosci. Cent., Univ. of Göttingen, Göttingen, Germany.
- Meyer, R., M. Sauter, and A. Weller (2014), Aquifer characterization using geoelectrical modelling, a case study, *Geophys. Res. Abstr.*, 16.
- Molina-Giraldo, N., P. Bayer, and P. Blum (2011), Evaluating the influence of thermal dispersion on temperature plumes from geothermal systems using analytical solutions, *Int. J. Therm. Sci.*, 50(7), 1223–1231.
- Ni, C.-F., T.-C. J. Yeh, and J.-S. Chen (2009), Cost-effective hydraulic tomography surveys for predicting flow and transport in heterogeneous aquifers, *Environ. Sci. Technol.*, 43(10), 3720–3727.
- Ptak, T., M. Piepenbrink, and E. Martac (2004), Tracer tests for the investigation of heterogeneous porous media and stochastic modelling of flow and transport—A review of some recent developments, *J. Hydrol.*, 294(1–3), 122–163.
- Schlie, P. (1989), *Hydrogeologie des Grunwasserwerkes, Stegemühle in Göttingen*, Geologische Inst., Univ. Göttingen, Göttingen, Germany.
- Straface, S., T. C. J. Yeh, J. Zhu, S. Troisi, and C. H. Lee (2007a), Sequential aquifer tests at a well field, Montalto Uffugo Scalo, Italy, *Water Resour. Res.*, 43, W07432, doi:10.1029/2006WR005287.
- Straface, S., C. Fallico, S. Troisi, E. Rizzo, and A. Revil (2007b), An inverse procedure to estimate transmissivity from heads and SP signals, *Ground Water*, 45(4), 420–428.
- Straface, S., F. Chidichimo, E. Rizzo, M. Riva, W. Barrash, A. Revil, M. Cardiff, and A. Guadagnini (2011), Joint inversion of steady-state hydrologic and self-potential data for 3D hydraulic conductivity distribution at the Boise Hydrogeophysical Research Site, *J. Hydrol.*, 407(1–4), 115–128.
- Vasco, D. W., and S. Finsterle (2004), Numerical trajectory calculations for the efficient inversion of transient flow and tracer observations, *Water Resour. Res.*, 40, W01507, doi:10.1029/2003WR002362.
- Vasco, D. W., and K. Karasaki (2001), Inversion of pressure observations: An integral formulation, *J. Hydrol.*, 253(1–4), 27–40.
- Vasco, D. W., and K. Karasaki (2006), Interpretation and inversion of low-frequency head observations, *Water Resour. Res.*, 42, W05408, doi:10.1029/2005WR004445.
- Vilhelmsen, T. N., S. Christensen, and S. W. Mehl (2012), Evaluation of MODFLOW-LGR in connection with a synthetic regional-scale model, *Ground Water*, 50(1), 118–132.
- Vogt, T. (2007), Implementation of an experimental field site—Hydraulic and borehole geophysical measurements, numerical groundwater flow modelling, Bachelor thesis, Geosci. Cent., Univ. of Göttingen, Göttingen, Germany.
- Yeh, T.-C. J., and S. Liu (2000), Hydraulic tomography: Development of a new aquifer test method, *Water Resour. Res.*, 36(8), 2095–2105.
- Zheng, C., and P. P. Wang (1999), MT3DMS: A modular three-dimensional multi-species transport model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems; documentation and users guide, *Contract Rep. SERDP-99-1*, 202 pp., U.S. Army Eng. Res. and Dev. Cent., Vicksburg, Miss.
- Zhu, J., and T.-C. Yeh (2006), Analysis of hydraulic tomography using temporal moments of drawdown recovery data, *Water Resour. Res.*, 42, W02403, doi:10.1029/2005WR004309.