

Highly parameterized inverse estimation of hydraulic conductivity and porosity in a three-dimensional, heterogeneous transport experiment

Hongkyu Yoon¹ and Sean A. McKenna¹

Received 16 March 2012; revised 18 September 2012; accepted 23 September 2012; published 20 October 2012.

[1] Assessing the impact of parameter estimation accuracy in models of heterogeneous, three-dimensional (3-D) groundwater systems is critical for predictions of solute transport. A unique experimental data set provides concentration breakthrough curves (BTCs) measured at a 0.25^3 cm^3 scale over the $13 \times 8 \times 8 \text{ cm}^3$ domain ($\sim 53,000$ measurement locations). Advective transport is used to match the first temporal moments of BTCs (or mean arrival times, m_1) averaged at 0.25^3 and 1.0 cm^3 scales through simultaneous inversion of highly parameterized heterogeneous hydraulic conductivity (K) and porosity (ϕ) fields. Pilot points parameterize the fields within eight layers of the 3-D medium, and estimations are completed with six different models of the K – ϕ relationship. Parameter estimation through advective transport shows accurate estimation of the observed m_1 values. Results across the six different K – ϕ relationships have statistically similar fits to the observed m_1 values and similar spatial estimates of m_1 along the main flow direction. The resulting fields provide the basis for forward transport modeling of the advection-dispersion equation (ADE). Using the estimated K and ϕ fields demonstrates that advective transport coupled with inversion using dense spatial field parameterization provides an efficient surrogate for the ADE. These results indicate that there is not a single set of model parameters, or a single K – ϕ relationship, that leads to a best representation of the actual experimental sand packing pattern (i.e., nonuniqueness). Additionally, knowledge of the individual sand K and ϕ values along with their arrangement in the 3-D experiment does not reproduce the observed transport results at small scales. Small-scale variation in the packing and mixing of the sands causes large deviations from the expected transport results as highlighted in forward ADE simulations. Highly parameterized inverse estimation is able to identify those regions where variations in mixing and packing alter the expected property values and significantly improve results relative to the naïve application of the experimentally derived property values. Impacts of the observation scale, the scale over which results are averaged and the number of observations and parameters on the final estimations are also examined. Results indicate existence of a representative element volume (REV) at 0.25 cm^3 , the existence of subgrid scale heterogeneity that impacts transport and the accuracy of highly parameterized models with even relatively small amounts of observations. Finally, this work suggests that local-heterogeneity features below the REV scale are difficult to incorporate into parameterized models, highlighting the importance of addressing prediction uncertainty for small-scale variability (i.e., uncaptured variability) in modeling practice.

Citation: Yoon, H., and S. A. McKenna (2012), Highly parameterized inverse estimation of hydraulic conductivity and porosity in a three-dimensional, heterogeneous transport experiment, *Water Resour. Res.*, 48, W10536, doi:10.1029/2012WR012149.

¹Geoscience Research and Applications Group, Sandia National Laboratories, Albuquerque, New Mexico, USA.

Corresponding author: H. Yoon, Geoscience Research and Applications Group, Sandia National Laboratories, Albuquerque, NM 87185, USA. (hyoon@sandia.gov)

1. Introduction

[2] Accurate identification of heterogeneous material properties in the subsurface is essential for predictive modeling of fluid flow and solute transport. Direct observations of the aquifer properties are limited, and data must be expanded and/or interpolated from the observation locations to populate the model domain. Furthermore, available observations of state variables such as pressure and/or concentration can be brought to bear on estimation of hydraulic parameters

through inverse modeling approaches [*de Marsily et al.*, 1999; *Carrera et al.*, 2005]. Recently, the importance of identifying the spatial distribution of hydrogeological heterogeneity has been increasingly centered on transport processes in the subsurface for applications including environmental remediation [*Murakami et al.*, 2010; *Zachara*, 2010], geological storage of CO₂ [*Baines and Worden*, 2004; *Department of Energy (DOE)*, 2007], and water resources management [*South Florida Water Management District*, 2005; *Young et al.*, 2010]. For strongly nonlinear problems such as reactive transport processes and groundwater flow in highly heterogeneous formations, a majority of inverse parameter estimations are computationally expensive and the impact of small-scale heterogeneity on model prediction can be profound. Additionally, quantification of the uncertainty associated with model predictions resulting from estimated parameters is receiving increased attention in order to better understand different sources of uncertainty [*Renard et al.*, 2010] and provide decision-makers with reliable predictive models [*Keating et al.*, 2010].

[3] Considerable research has been devoted to developing theories and techniques for solving inverse parameter estimation problems in groundwater problems, which have been reviewed thoroughly in the literature [*Carrera*, 1987; *Carrera et al.*, 2005; *McLaughlin and Townley*, 1996; *Hendricks Franssen et al.*, 2009]. These techniques can generally be classified by considering (1) how the variables being estimated are parameterized (e.g., lumped versus distributed parameters), (2) whether or not the result of the inverse procedure is a deterministic best approximation or provides a probabilistic assessment of the uncertainty, and (3) the assumed degree of linearity in the relationship between material properties and observed state variables. Approaches that provide estimations of spatially varying distributed parameters include the generalized likelihood uncertainty estimation (GLUE) method [*Beven and Binley*, 1992], the geostatistical inverse approach [*Kitanidis*, 1996; *Li et al.*, 2005, 2007], Bayesian approaches to formally integrate prior information [*Woodbury and Ulrych*, 2000; *Jiang et al.*, 2004; *Fu and Gómez-Hernández*, 2009], the method of stochastic differential equations [*Rubin and Dagan*, 1987; *Guadagnini and Neuman*, 1999; *Hernandez et al.*, 2006], gradual deformation methods [*Hu*, 2000], ensemble filtering approaches [*Nowak*, 2009; *Bailey and Bau*, 2011], sequential self-calibration [*Gómez-Hernández et al.*, 2003], and pilot point methods [*Lavenue and de Marsily*, 2001] among others. Recently, seven inverse methods commonly used in parameter estimation for groundwater flow have been compared in two synthetic heterogeneous transmissivity fields by *Hendricks Franssen et al.* [2009]. It was concluded that while formulation and parameterization of each method are different, the overall performance differences (i.e., predictive accuracy) do not vary significantly. However, the comparison was limited to the synthetic groundwater flow problems in two dimensional (2-D) heterogeneous domains.

[4] A number of previous works have estimated more than one spatially heterogeneous parameter from the same set of observations in 2-D fields. Most commonly, heterogeneous transmissivity (T) and storativity (S) fields have been estimated from some combination of steady state and transient head observations [*Hendricks Franssen et al.*, 1999; *Li et al.*, 2005, 2007]. Other works have focused on the simultaneous

estimation of a heterogeneous T field and a spatially varying transport parameter such as the distribution of sorption coefficient, dispersivity, or contaminant source terms [e.g., *Huang et al.*, 2004; *Nowak and Cirpka*, 2006; *Tonkin et al.*, 2007; *Hosseini et al.*, 2011]. These approaches have generally been demonstrated on horizontal fields with other researchers estimating multiple properties in 2-D cross-sectional domains [e.g., *Fienen et al.*, 2009]. A question remains in all of these studies regarding the relationship between heterogeneous parameters. Most commonly, independence between the estimated parameter values is assumed, as often necessitated by the lack of observed values for one of the parameters.

[5] Over the past decade, a number of studies have increasingly focused on parameter estimation in heterogeneous 3-D fields. *Lavenue and de Marsily* [2001] used observations of a sinusoidally varying pressure test with pilot point parameterization to estimate heterogeneous hydraulic conductivity (*K*) in a two-layered medium. *Hendricks Franssen and Gómez-Hernández* [2002] estimated *K* in a 3-D fractured medium from steady and transient head observations with the locations of fracture zones specified a priori. *Llopis-Albert and Capilla* [2010] extended this work to account for stochastic representation of complex fracture structures similar to that presented by *Gómez-Hernández et al.* [2001]. *Li et al.* [2008] used the geostatistical inverse approach to estimate a 3-D *K* field from steady state drawdown and vertical flowmeter surveys. *Riva et al.* [2008] used a stochastic Monte Carlo method to describe 3-D random geological facies and hydraulic conductivities in order to capture the features of the depth averaged breakthrough curve (i.e., temporal moments and long tailing). In particular, *Riva et al.* [2008] used an empirical relationship between *K* and porosity (ϕ) to generate ϕ distribution with a dual medium model. Recently, *Schöniger et al.* [2012] improved ensemble Kalman filter (EnKF) approaches by applying nonlinear, monotonic transformations to the observed states in parameter estimation from 3-D hydraulic tomography in multi-Gaussian log *K* fields. *Huber et al.* [2011] used the pilot point method to calibrate a 3-D groundwater flow model of a strongly heterogeneous aquifer with the hydraulic conductivities at a limited number of locations and the leakage coefficients for a number of zones, which were updated in real time with EnKF.

[6] *Doherty* [2003] applied the pilot point method for parameter estimation in the context of underdetermined problems. This allowed pilot points to be distributed through the model domain, resulting in highly parameterized inversion estimation. Inverse estimation of the highly parameterized model with stability and uniqueness was accomplished through regularizations such as the truncated singular-value decomposition (TSVD) and Tikhonov regularization (i.e., hybrid subspace) approaches [*Tonkin and Doherty*, 2005]. In addition, the highly parameterized inversion using pilot points has led to the development of computationally efficient calibration-constrained model prediction uncertainty by *Doherty* and coworkers [*Moore and Doherty*, 2005; *Tonkin et al.*, 2007; *Tonkin and Doherty*, 2009]. Highly parameterized inverse estimation with pilot points also has been used in 3-D domains to simultaneously estimate three heterogeneous parameters: porosity, horizontal and vertical hydraulic conductivity, in a model with 11–14 layers conditioned to both head and concentration data [*Tonkin and*

[Doherty, 2005; Tonkin *et al.*, 2007]. While parameter estimation with pilot points has become relatively common [Alcolea *et al.*, 2006, 2008; Riva *et al.*, 2010; Doherty and Hunt, 2010; Doherty *et al.*, 2010], its application to estimation of 3-D fields has been limited.

[7] Here we focus on highly parameterized inverse modeling of spatially heterogeneous fields for a 3-D transport problem. The high level of parameterization is used to estimate fields of spatially correlated properties, and the pilot point method is used to parameterize these spatial fields. Parameter estimation with pilot points often utilizes multiple locations with existing property measurements to condition the estimated field. The experimental data set examined here is unique in that it provides extremely fine-scale (0.25^3 cm^3) 3-D exhaustive observations on concentrations in a 3-D flowcell over a 10–20 cm scale [Zhang *et al.*, 2007; Yoon *et al.*, 2008]. Contrary to most parameter estimation studies, however, there are no direct measurements of K or ϕ within the model domain, but independently measured K and ϕ values are available [Zhang *et al.*, 2007; Yoon *et al.*, 2008]. The exhaustive 3-D observations available from magnetic resonance imaging (MRI) in the flowcell packed with sands are an uncommon measurement set for inversion, but are analogous to time series imaging of transport in a geologic unit with geophysical techniques. This unique concentration data set provides an opportunity to explore the application of inverse parameter estimation for multiple variables using highly parameterized models in 3-D groundwater flow and solute transport problems.

[8] Objectives of this work are to: (1) Utilize a set of spatially exhaustive 3-D observations to simultaneously estimate heterogeneous K and ϕ fields through application of highly parameterized models. In particular we evaluate the capability of six different approaches for quantifying the relationship between K and ϕ in the estimation process to recover the observed tracer transport. This evaluation also includes the straight forward approach of using the known zonation of the sands and the laboratory measured K and ϕ values of each sand. (2) Examine the ability of highly parameterized estimations of K and ϕ to explain breakthrough curves (BTCs) with and without the addition of a dispersion term in forward solute transport models. If K and ϕ are parameterized at a high resolution, we hypothesize that it should be possible to use a purely advective transport model to match the first temporal moments of BTCs (or mean arrival times, m_1) as a surrogate for more complex transport in the inversion. The estimated fields should then provide accurate transport results when used for forward runs with the advection-dispersion equation (ADE). (3) Determine the impact of varying the number of parameters and the number of observations on the ability of the estimated 3-D fields to match observed data at different scales. Additionally, the impact of the averaging scale of the observations (i.e., 0.25^3 and 1.0 cm^3) on the resulting estimated fields is examined to evaluate the impact of sub-scale heterogeneity on forward transport results with ADE.

2. Experimental Setup and Parameterization

2.1. Experimental Setup

[9] The experimental procedure and data acquisition are given in detail by Zhang *et al.* [2007] and Yoon *et al.*

[2008]. Here, we briefly describe the experimental setup for flow and transport modeling. Five different sizes of sands were used to construct a heterogeneous K field in a 3-D flowcell. The entire flowcell has a dimension of $21.5 \times 9 \times 8.5 \text{ cm}^3$ and the central portion of the flowcell ($14 \times 8 \times 8 \text{ cm}^3$) was packed with 1 cm^3 cubes of five different sand types. The K and ϕ of all five sands were measured independently [Yoon *et al.*, 2008] (Table 1). The sand pattern is based on the distribution of a K field generated using the sequential indicator simulation algorithm (SISIM) [Deutsch and Journel, 1998]. As shown in Figure 1a, there is a distinct, connected, high- K (sands 12/20 and 20/30) feature along the central portion of the flowcell. A snapshot of tracer concentration after 42 min qualitatively illustrates the preferential flowpath along the central portion of the flowcell (Figure 1b). A brass divider ($14 \times 8 \times 1.2 \text{ cm}^3$) with a grid of 1 cm^2 square openings was used to pack the flowcell with dry sands in each 1 cm layer. The rest of the flowcell includes a 4.5 cm zone adjacent to the inlet, a 3 cm zone adjacent to the outlet, a 0.5 cm thick layer on the bottom, and a 0.5 cm thick vertical layer adjacent to the side walls, all of which were filled with the lowest- K sand (50/70). A constant head was maintained at the inlet and outlet reservoirs and a nonreactive, para-magnetic tracer solution was continuously injected through the water-saturated flowcell by maintaining a constant flux. The total time for the tracer solution to flow through the entire heterogeneous region was $\sim 4\text{--}5 \text{ h}$.

[10] The signal intensity of the tracer concentration using MRI was obtained at a resolution of $0.1875 \times 0.1875 \times 0.225 \text{ cm}^3$ at a regular sampling interval of 2.17 min over the MR imaging region ($13 \times 8 \times 8 \text{ cm}^3$), which is slightly smaller than the entire heterogeneous region ($14 \times 8 \times 8 \text{ cm}^3$), and was converted into tracer breakthrough curves (BTCs) [Yoon *et al.*, 2008]. Observed concentration BTCs averaged at both 0.25^3 cm^3 (i.e., 0.016 cm^3 , OBS_{0.016}) and 1 cm^3 scales (OBS_{1.0}) were used. For model comparison, the mean arrival times (m_1) and second central moments (m_{2c}) were calculated from the observed BTCs using the method of temporal moments with a step input [Luo *et al.*, 2006]. Observed data within the top 0.25 cm of the heterogeneous region were not used for model calibration due to decreased imaging accuracy. Actual observed data used are from $x = 4.5$ to 17.5 cm , $y = 0.5$ to 8.5 cm , and $z = 0.5$ to 8.25 cm in the heterogeneous region (Figure 1a) and the total number of observed m_1 data at the OBS_{0.016} and OBS_{1.0} scales are 51,584 and 832, respectively.

[11] In this study, there are no direct measurements of K or ϕ in the flowcell after packing, but independently measured K and ϕ values for all five sand types used for packing are available (Table 1). A priori bounds on K and

Table 1. Properties of Sands

Sand Type ^a	Measured K (cm min^{-1})	Measured ϕ	Mean Grain Size (mm)
12/20	25.08	0.355	1.1
20/30	13.44	0.356	0.71
30/40	6.72	0.355	0.53
40/50	3.78	0.355	0.36
50/70	2.03	0.372	0.26

^aSand type is defined as a sand fraction between two sieve numbers.

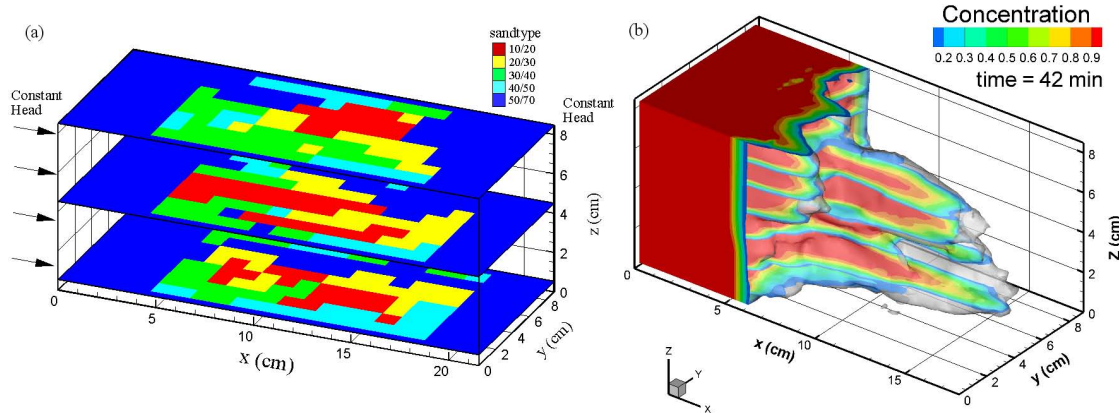


Figure 1. (a) Numerical model domain of 3-D flowcell with hydraulic conductivity distribution of packing patterns shown on three different layers and (b) snapshot of 3-D concentration plume after 42 min. The heterogeneous region spans 4.5 to 18.5 cm, 0.5 to 8.5 cm, and 0.5 cm to 8.5 cm in x , y , and z directions.

ϕ parameters were also independently measured over a mix of two sand types. Porosity varied as a function of volume fraction of each sand type and has a range from 0.25 to 0.45. K values were also measured for sand mixtures with the lowest, median, and highest porosity values, and have a range from 0.3 to 61 cm min⁻¹. In this work, the distribution of the K field used for packing and the independently measured values of K and ϕ on each sand (K_{Pack}) are used for comparison of simulation results with the estimated K and ϕ fields from different parameterizations.

2.2. Parameterization of K and Porosity (ϕ)

[12] Six different approaches for quantifying the relationship between heterogeneous K and ϕ in the estimation process subject to the observed tracer transport were compared. The relationship between K and ϕ is modeled using the following six different parameterizations:

[13] 1. *Independence (K_{IND})*. Following work in a number of previous studies where multiple spatially varying parameters were estimated [e.g., Hendricks Franssen *et al.*, 1999; Li *et al.*, 2005; Tonkin and Doherty, 2005], we consider K and ϕ to be independent of each other. This leads to the largest number of parameters (i.e., $2 \times$ the number of pilot points) to be estimated.

[14] 2. *Independence with Zonal Information (K_{Zone})*. In a variation on the independence parameterization, we add knowledge of the packing distribution of the five different sands as prior information. The packing distribution is referred to as “zonal information”. The spatial estimation of each parameter is limited to only using information from pilot points within the same zone (sand type) within that layer. Zonal information can be analogous to stratigraphic or geological facies information, or lithology reconstruction from other geophysical interpretations. This approach should better represent the discrete boundaries of the sand types in the estimated K and ϕ fields. The number of estimated parameters is equal to that in the previous case.

[15] 3. *Linear model of Coregionalization (K_{Coreg})*. The linear model of coregionalization [e.g., Goovaerts, 1997] enables modeling a linear relationship with a specified correlation coefficient between two variables that also exhibit

spatial correlation. K is estimated at pilot points and the estimated K field is used to modify an independently generated random field for ϕ with the same covariance structure as the K field to produce a specified (K - ϕ) correlation coefficient. Here, we model a linear relationship between $\log K$ and ϕ with the correlation coefficient, positive or negative, in each layer as an estimated parameter. The resulting number of parameters is equal to the number of pilot points where K is estimated plus 24 (the correlation coefficient, mean and standard deviation of porosity in each of eight layers).

[16] 4. *Kozeny-Carman Equation (K_{KC})*. Empirical relationships between K and ϕ are also used to reduce the degrees of freedom for estimation problems [Hu *et al.*, 2009; Riva *et al.*, 2008]. For sandy materials, the Kozeny-Carman formula can be used to develop an empirical relationship between K and ϕ for the five different sands used for packing. Using the characteristic mean grain size and measured K values for the five different sands (Table 1), a linear regression model is developed based on the Kozeny-Carman equation as

$$\phi = 0.0879 \log K + 0.2587. \quad (1)$$

[17] The regression model ($R^2 = 0.989$) is used to estimate ϕ from $\log K$ (cm min⁻¹). This parameterization reduces the number of estimated parameters for porosity to zero.

[18] 5. *Constant Porosity (Domain) (K_{only})*. In the majority of studies where K (or transmissivity, T) is estimated from head and concentration measurements, ϕ is assumed to be constant-valued throughout the model domain [e.g., Fienen *et al.*, 2009; Fu and Gómez-Hernández, 2009; Herckenrath *et al.*, 2011]. We include this parameterization here for comparison. This parameterization reduces the number of estimated parameters for porosity to one.

[19] 6. *Constant Porosity (Layers) (K_{LAY})*. In a variation on the constant porosity parameterization, we add a single estimated porosity value for each layer in the 3-D model. This approach is similar to that used by Kollat *et al.* [2011] and results in eight parameters for the porosity estimation.

Table 2. Summary of the Parameters Used for Six Inversion Models^a

	Inlet Homogeneous Region		Heterogeneous Region ^b		Total Number of Parameters
	K^c	ϕ^d	K	ϕ	
K_{Zone}^e	160 K (20 per layer)	8 ϕ (one per layer)	896 K (one per 1 cm ³ cell)	896 ϕ	1961
K_{IND}^e	160 K (20 per layer)	8 ϕ (one per layer)	896 K (one per 1 cm ³ cell)	896 ϕ	1961
K_{Coreg}	160 K (20 per layer)	8 ϕ (one per layer)	896 K (one per 1 cm ³ cell)	8 correlation coefficients ^f 8 means and s.d. ^f	1089
K_{KC}	160 K (20 per layer)	8 ϕ (one per layer)	896 K (one per 1 cm ³ cell)	—	1065
K_{LAY}	160 K (20 per layer)	8 ϕ (one per layer)	896 K (one per 1 cm ³ cell)	8 ϕ^f	1073
K_{only}	160 K (20 per layer)	8 ϕ (one per layer)	896 K (one per 1 cm ³ cell)	1 ϕ	1066

^aFor all six methods, the head at the inlet (i.e., head difference between inlet and outlet boundaries) is estimated to match the flow rate entering the inlet of the flowcell.

^bPilot points are located at the center of each 1 cm² packing unit per each layer.

^cLower and upper range of Log K values are -1 and 2 , respectively. 20 pilot points per each layer.

^dLower and upper range of ϕ values are 0.25 and 0.45 , respectively.

^eFor the case of calibrating ϕ as independent parameter (K_{IND} and K_{zone}), Tikhonov regularization was additionally employed to regularize ϕ .

^fOne per each layer.

[20] For the highly parameterized model, 896 pilot points were used; one in the center of each 1 cm³ block within the heterogeneous region. Additionally, 160 pilot points with 20 pilot points and one pilot point for K and ϕ , respectively, in each layer are located in the homogeneous inlet region. Table 2 summarizes the main parameters and the number of pilot points for all parameterization methods. Simulation results with the estimated fields of K and ϕ from all six parameterizations are compared to each other and to the results obtained from K_{Pack} .

2.3. Parameter Estimation

[21] Model calibration was performed using the parameter estimation package PEST [Doherty, 2010; Doherty and Hunt, 2010]. In particular, BeoPEST [Hunt et al., 2010], which uses the Message Passing Interface (MPI) protocol suitable for a tightly integrated cluster and cloud computing, was used to improve computational efficiency. Spatially varying parameters considered for estimation are log K and ϕ . Additionally, the head at the inlet (i.e., head difference between inlet and outlet boundaries) is estimated to match the flowrate entering the inlet of the flowcell because the total flowrate varies during calibration of K and ϕ fields. The overall objective function to be minimized is the sum of the measurement (Φ_m) and regularization (Φ_r) objective functions for the cases (K_{Zone} and K_{IND}) where the regularization is employed. The objective function is

$$\Phi = (m_1^{obs} - m_1^{sim})^T Q_m (m_1^{obs} - m_1^{sim}) + \mu(d - Zb)^T Q_r (d - Zb), \quad (2)$$

where the first term is Φ_m which is the weighted sum of squared errors between the observed (obs) and simulated (sim) m_1 , the second term is Φ_r , d is a regularization observation vector, Z is a matrix that relates the parameter values (b) to d , μ is a regularization weight factor, diagonal matrixes Q_m and Q_r represent the square of the observation and regularization weights, respectively, and a superscript, T is the transpose operation. A weight factor of $1000 \sim 5000$ is used for the fixed head at the inlet, compared to a unit weight factor for all m_1 data, to ensure accurate estimation of the flowrate during calibration (i.e., negligible flowrate error). For regularization inversion, a regularization weight

of 300 is used for porosity so that initial Φ_m and Φ_r are similar, and the optimal value for the regularization weight factor was calculated for the iteration during calibration by PEST [Fienen et al., 2009; Doherty et al., 2010].

[22] Both K and ϕ are considered to be spatial fields and the spatial distribution of parameters is represented using pilot points. In this work, kriging with a spherical variogram model was used to interpolate property values over the entire numerical domain based on parameter values at pilot points. The range of the variogram was 4.2 cm for both log K and ϕ , and the sill of log K and ϕ were 0.2 and 0.005. It is noted that in this work the value of the sill has no effect on calibration, but rather the proportional contributions from the Kriging factors modify the heterogeneous field [Doherty et al., 2010]. To reflect the packing procedure in which 1 cm thick physical layers (total 8 layers above the 0.5 cm bottom layer) were packed independently, interpolation of parameter values to the model nodes from pilot points was undertaken for each layer (i.e., 2-D interpolation). The number of neighboring pilot points used in the local kriging interpolation to the model grid was limited to five to account for local heterogeneity by avoiding over-smoothing of the calibrated parameter field [Doherty et al., 2010]. This leads to the pilot point distance of 1 cm that is much smaller than the variogram range, which renders the influence of the variogram range to be negligible [Cirpka and Nowak, 2003] and allows local heterogeneity to be captured by the parameter values at pilot points during calibration [Doherty et al., 2010]. While the variogram parameters (i.e., range and sill) are not estimated in this work, the parameters describing spatial variability of heterogeneity can be estimated by Monte Carlo simulations with pilot point method [Castagna and Bellin, 2009] and moment equations [Riva et al., 2010].

[23] In the 3-D domain, there are four numerical grid cells (0.25 cm) comprising each packing layer, so each grid cell in the vertical direction within each 1 cm layer has the same material properties (i.e., K and ϕ). The initial values for K and ϕ at each pilot point were assigned to the mean of the measured values of five sand materials (Table 1). The pilot points are, in general, uniformly distributed over the heterogeneous region. For the most highly parameterized models, the pilot points are located at the center of each 1 cm³ based on the packing distribution. For the inlet

homogeneous region, a total of 160 pilot points (i.e., 20 pilot points per each layer) were uniformly located to properly calibrate the mean arrival times in the upstream of the heterogeneous region. In this region, $\log K$ is estimated at each pilot point, but ϕ is estimated per each layer due to the expected uniform porosity during packing of a single sand size.

[24] The parameter estimation with the OBS_{0.016} data (51584 m_1) is solved using the truncated singular value decomposition (TSVD) approach within PEST [Doherty, 2003, 2010; Tonkin and Doherty, 2005, 2009]. For the case of calibrating ϕ as an independent parameter (K_{IND} and K_{Zone}), Tikhonov regularization was additionally employed to regularize ϕ . Through Tikhonov regularization, preferred values for ϕ were provided and balancing between misfit and the departure from the preferred values is achieved by selecting an optimal solution according to equation (2). In this study, ϕ was regularized by using a single constant preferred value (0.355) for all five sands based on the measured values (Table 1). In addition, a priori bounds on all K and ϕ parameters were assigned based on independently measured $\log K$ (−1 to 2) and ϕ (0.25 to 0.45) values over different sand mixtures as described earlier. A relatively low truncation criterion of singular values (i.e., 1×10^{-9}) was used due to the different weighting factor on the flow-rate, so that all significant parameters are estimated during calibration. The hybrid Tikhonov and TSVD approach provides a tool to achieve a stable and effective estimation, which is particularly well-suited for highly parameterized models [Doherty *et al.*, 2010].

[25] Parameter estimation with the 1 cm³ observed data (OBS_{1.0}, 832 m_1) is solved using the SVD-assist approach within PEST [Doherty *et al.*, 2010; Tonkin and Doherty, 2005, 2009]. For this case, no Tikhonov regularization is applied, but the same bounds of parameter values used for OBS_{0.016} are employed here. In SVD-assist, the parameter space is decomposed based on primary components from SVD of the sensitivity matrix (i.e., the weighted Jacobian matrix) whose linear combinations are called superparameters in PEST. The calibration solution space estimable by calibration data set is determined through truncating SVD based on the TSVD threshold [Doherty and Hunt, 2010]. For highly parameterized models the TSVD in the SVD-assist technique is numerically stable and computationally efficient since the number of parameters to be calibrated is effectively reduced into a smaller set of the calibration solution parameters, instead of computing sensitivities with respect to all variable parameters. The maximum number of the calibration solution (super-) parameters is set to the number of observed data used for calibration. The threshold value for singular values is set to 1×10^{-9} .

2.4. Numerical Simulation

[26] For parameter estimation, 3-D, steady state, water flow was simulated using MODFLOW-2000 [Harbaugh *et al.*, 2000] and advective tracer transport was simulated using the particle tracking code MODPATH [Pollock, 1994]. A uniform numerical model grid spacing of 0.25 cm was used in the entire 3-D domain. The top, bottom and side boundaries were assigned as zero-flux boundaries, and constant head boundaries at the inlet and outlet were imposed, but a head difference at the inlet and outlet was

calibrated to match the overall flowrate (5.25 cm³ min^{−1}). At the start of each simulation, the tracer is introduced into the steady flow field by placing a total of 78,336 particles (64 particles per model grid cell) uniformly over the cross-section (i.e., y - z plane) at a specified x -location upgradient of the heterogeneous portion of the domain. The number of particles used in this work ensures that all numerical grid cells have particles pass through, such that mean arrival times can be computed. Individual particle locations in the heterogeneous domain are tracked at an interval of 0.5 min, and the forward MODPATH simulations were performed until all particles exited the heterogeneous domain. A histogram is constructed by counting the number of particles that arrive at each grid cell within the time interval. The histograms are converted into BTCs by normalizing over the total number of particles within each grid cell over the entire simulation time.

[27] In addition, forward solute transport (i.e., ADE) was simulated to determine the ability of highly parameterized estimations of K and ϕ with an advective (surrogate) transport model in order to provide accurate transport results compared to the observed BTCs. The estimated K and ϕ fields were input for ADE simulations. The longitudinal dispersivity (α_L) was set equal to the mean grain diameter for each sand type based on the packing distribution. The transverse dispersion coefficient (α_T) was set to a 10th of the longitudinal dispersion coefficient, as discussed in Yoon *et al.* [2008]. The integrated finite difference simulator STOMP (Subsurface Transport Over Multiple Phases) [White and Oostrom, 2006] was used to solve the ADE, and the grid spacing in the x -direction was 0.25 cm and the grid spacing in y and z directions was 0.125 cm to ensure numerical accuracy. A third-order scheme using the total variation diminishing (TVD) technique was used to solve advective transport. The boundary and initial conditions were the same as described above. Additional description of the STOMP simulation is reported in Yoon *et al.* [2008].

2.5. Model Evaluation

[28] In this study, parameter estimations have been performed with two sets of observed m_1 data (OBS_{0.016} and OBS_{1.0}). Observed and simulated BTCs were compared at three different averaging scales: (1) the numerical grid scale of 0.25³ cm³ (0.016 cm³), (2) the intermediate scale of 0.25 cm \times 1 cm² (0.25 cm³) slice in the transverse directions (y - z directions), and (3) the physically based block scale of 1 cm³. These three different averaging scales span 2 orders of magnitude and will be hereafter referred to as AVG_{0.016}, AVG_{0.25}, and AVG_{1.0} scales. For model comparison, m_1 and m_{2c} were calculated from the simulated BTCs using the method of temporal moments as described earlier.

[29] The quality of estimated K distributions compared to the actual sand packing pattern is assessed by mapping accuracy. In this study, mapping accuracy was the fraction of correctly estimated grid cells with regard to the sand type used for physical packing (K_{pack}). In addition, grid cells which match each sand type as an adjacent sand type in order of K value (e.g., matching sand 20/30 as sand 10/20 or sand 30/40) are also counted as accurate. A higher mapping accuracy indicates higher portions of estimated K values matching the corresponding sand type for the packing pattern.

[30] Several statistical measures are used to evaluate the accuracy of the calibrations and ADE simulations: the mean error (ME), the mean absolute error (MAE), and the root mean square error (RMSE). Error is calculated by (observed value – simulated value). Outliers are defined as estimates with ME greater than 15 min or less than –15 min, and are used solely to highlight the number of extreme mismatches. MAE represents the measure of mismatch between measured and estimated results and RMSE emphasizes mismatches of larger magnitudes. In addition, the AIC_c [Hurvich and Tsai, 1989] and KIC [Kashyap, 1982], which are model selection criteria based on model discrimination (or information), are computed after inversion in order to rank different parameterizations. In this work, the model selection criteria for groundwater models are defined as in Poeter and Hill [2007]

$$AIC_c = n \ln(\sigma^2) + 2k + \frac{2k(k+1)}{n-k-1};$$

$$KIC = (n - (k - 1)) \ln(\sigma^2) - (k - 1) \ln(2\pi) + \ln|X^T \omega X| \quad (3)$$

where n is the number of observations, σ^2 is the estimated residual variance ($= \Phi_m/n$), k is the number of parameters ($=$ the number of calibrated model parameters + 1), X is the sensitivity matrix, and ω is the weight matrix. The model selection criteria account for goodness of fit between observed and predicted states using maximum likelihood estimates of model parameters and model complexity based

on the principle of parsimony which penalizes models with a large number of parameters depending on the improvement of model fit with more parameters. The determinant in the last term of KIC is computed using the first term (Φ_m) in equation (2) and the role of this determinant in a Fisher information matrix term can be found in Ye *et al.* [2008]. The model with the lowest value in equation (3) is the best model (i.e., minimizing information loss). In the context of highly parameterized models with the number of parameters exceeding the number of observations, practical and theoretical issues are discussed in Poeter and Hill [2007], Riva *et al.* [2010], and Doherty [2012].

3. Results

3.1. Parameter Estimation With Highly Parameterized Models

[31] Parameter estimations with highly parameterized models have been performed with $OBS_{0.016} m_1$ data using MODFLOW and MODPATH (i.e., advection only). Parameters employed in each method are listed in Table 2. K and ϕ distributions for K_{Pack} and resulting estimations for all six K - ϕ relationships for four of the eight horizontal layers are shown in Figure 2. Estimated K fields for all six methods capture the features of packing where high- K zones are predominantly interconnected through the central regions of the flowcell (Figure 1). Pairs of inverse models with similar parameterizations (e.g., $K_{Zone} - K_{IND}$, $K_{Coreg} - K_{KC}$, and $K_{LAY} - K_{only}$) produce similar estimated K patterns. The quality of estimated K distributions compared to the

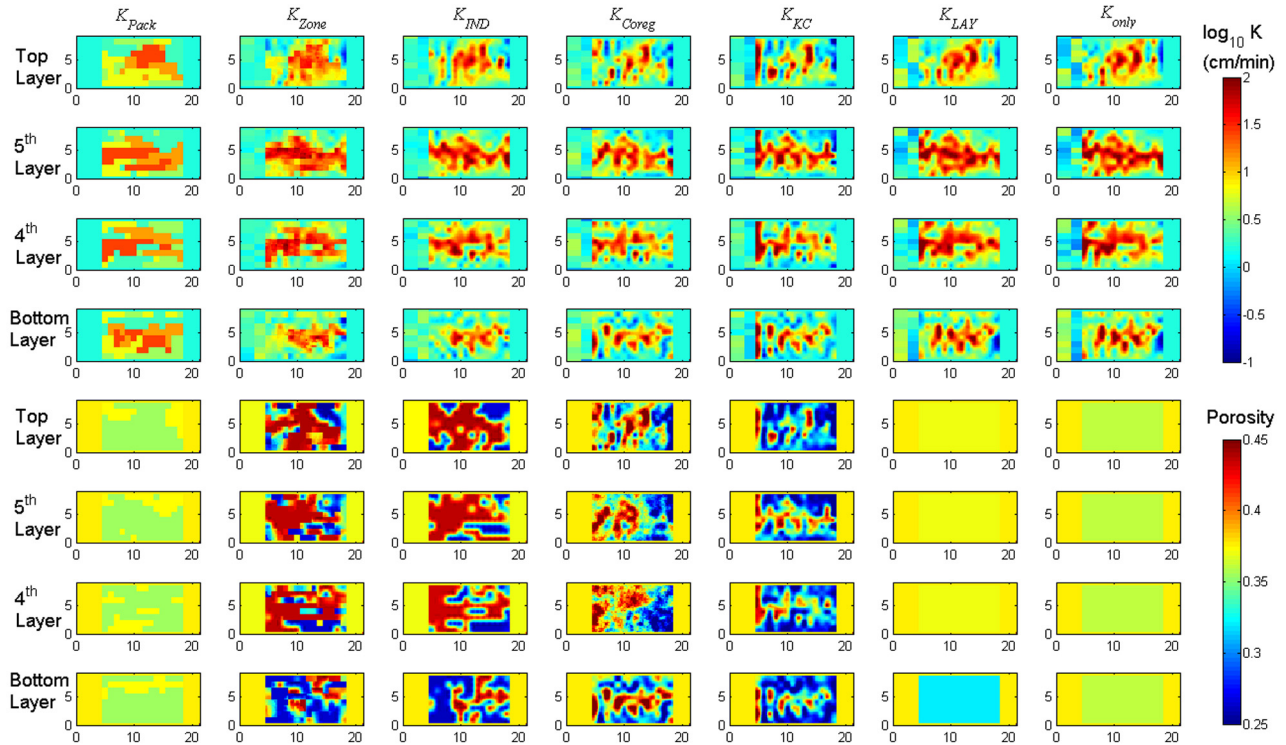


Figure 2. K and porosity distributions for experimental packing pattern (K_{Pack}) and estimation for all six K -porosity relationships in four different horizontal layers. Highly parameterized models with 896 pilot points in the heterogeneous region and 160 pilot points in the homogeneous region were calibrated with observed m_1 data at the 0.016 cm^3 scale (i.e., $OBS_{0.016}$, 51,584 m_1).

original packing pattern is quantitatively assessed by mapping accuracy in Table 3. All six K - ϕ relationships are able to estimate the locations of different sand type distribution used in K_{Pack} , on average, with greater than 50% accuracy. As expected, addition of the zonation boundaries as prior information in the inversion results in the K_{Zone} approach having a much better mapping accuracy than other methods. K_{IND} has a better mapping accuracy than the rest of four parameterizations, but the confidence interval of K_{IND} is slightly overlapped with that of four parameterizations (Table 3), indicating that more parameterization without zonal information leads to slightly better identification of the packing pattern, but the increase in accuracy is not statistically significant.

[32] It is also noted that even using the zonal parameterization, there is a high degree of variability within the zones relative to the assumed internally homogeneous packing pattern. As an example, Figure 2 identifies several regions with K values higher and lower than those used in K_{Pack} . The lower- K regions correspond to areas where coarse sand interfaces fine sand in the original packing and represents regions where fine sand has mixed into the inter-grain volumes within the coarse sand, thus reducing K and ϕ in those areas. The higher- K regions represent relatively interconnected homogeneous regions such as central portions of the flowcell (e.g., dark red color in Figure 2) where loose packing of the sand may result in higher K and ϕ [Zhang *et al.*, 2007, 2008]. This result highlights the ability of the highly parameterized model to capture key features of the heterogeneity.

[33] Statistics of mean arrival times (m_1) for all parameterizations are very similar (Figure 3). As expected, the K_{Zone} and K_{IND} sets have slightly lower MAE and RMSE values than other parameterization sets, primarily due to higher parameterization. RMSE was reduced more than AAE compared to K_{Pack} , indicating that the higher magnitudes of outliers are significantly reduced by all parameterizations as shown in box plots. AIC_c and KIC values for six parameterizations are listed in Table 4. For the OBS_{0.016} data, the rank in both criteria is the same and the K_{Zone} set has the lowest value, indicating that higher parameterization with additional information (i.e., zonal information) leads to an improvement in model fit. The K_{IND} and K_{Coreg} sets are next ranked, and the improvement of model fit in K_{IND} can be considered as a direct result of using more parameters. For the OBS_{1.0} data, only AIC_c is computed due to the complexity of computing the determinant term in KIC and the number of estimated parameters for all six

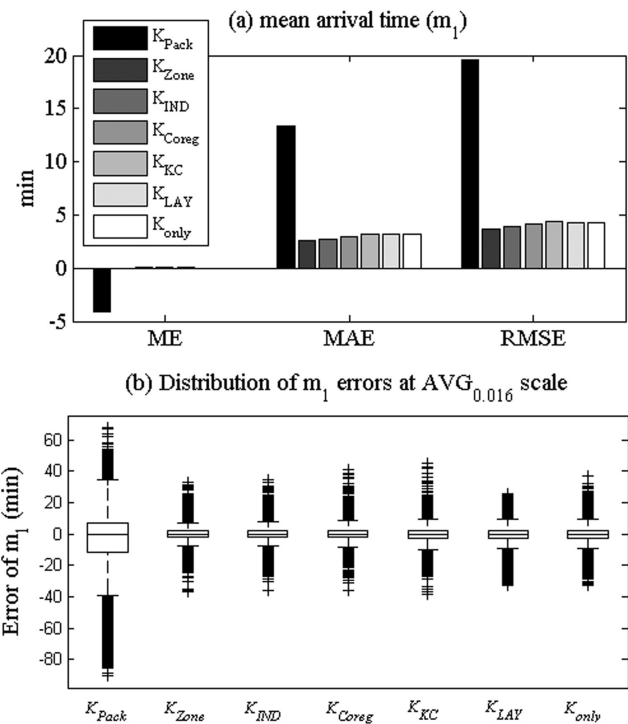


Figure 3. (a) Comparison of ME, MAE, and RMSE values and (b) box plots of m_1 errors at the AVG_{0.016} scale for the packing pattern and six K -porosity relationships. OBS_{0.016} (51,584 m_1) are used for estimating parameters using MODPATH (i.e., advection only).

parameterizations was set to the number of observed data ($n = 833$) through TSVD, making AIC_c strongly determined by goodness of fit (i.e., the first term in AIC_c). AIC_c values for the K_{Zone} and K_{IND} sets are lower than those for other sets. This may demonstrate that the information contained in OBS_{1.0} was best transferred to parameters with highly parameterized models with more parameters such as K_{Zone} and K_{IND} . Further work is needed to provide practical and theoretical guidance for applying the model selection criteria for highly parameterized models.

[34] Compared to K_{Pack} , all inverse results reduce the mismatch of m_1 , resulting in lowering the objective function (equation (2)), demonstrating that the actual experiment and the designed experiment are different and inverse modeling improves the results relative to K_{Pack} relying on

Table 3. Mapping Accuracy of K Distribution Compared to the Packing Pattern^a

	Sand Type					Composite Sands	STD	Confidence Interval (95%)	
	10/20	20/30	30/40	40/50	50/70			Low	High
K_{Zone}	86.9	84.8	81.1	91.2	82.3	85.2	1.2	82.8	87.6
K_{IND}	64.5	60.8	48.5	78.2	57.1	61.7	1.6	58.6	64.8
K_{Coreg}	57.9	47.4	40.8	81.8	49.8	55.4	1.7	52.1	58.7
K_{KC}	58.5	45.6	27.8	84.1	52.7	53.8	1.7	50.5	57.1
K_{LAY}	63.9	62.6	47.3	72.4	45.8	58.0	1.6	54.9	61.1
K_{only}	65.0	59.1	44.4	72.9	45.3	57.0	1.7	53.7	60.3

^aMapping accuracy in percent.

Table 4. AIC_c and KIC Results for Six Parameterizations^a

	k^b	Φ_m^c	AIC _c	KIC
K_{Zone}	1962 (834)	697,370 (132.9)	138,411 (-696,250.9)	144,559.1
K_{IND}	1962 (834)	765,780 (240.6)	143,238.2 (-695,756.5)	149,202.9
K_{Coreg}	1090 (834)	880,520 (749.3)	148,588.3 (-694,810.2)	153,452.4
K_{KC}	1066 (834)	1007400 (1180.0)	155,482.3 (-694,431.9)	160,359.3
K_{LAY}	1074 (834)	941,610 (701.7)	152,015.1 (-694,864.9)	156,716.7
K_{only}	1058 (834)	948,580 (932.9)	152,362.2 (-694,627.7)	156,961.9

^aValues outside and within parenthesis represent results calibrated with OBS_{0.016} ($n = 51,585$) and OBS_{1.0} ($n = 833$) data, respectively. For results with OBS_{1.0} the number of the calibration solution parameters is set to the number of observed data ($n = 833$) by TSVD.

^b k is the number of parameters (=the number of calibrated model parameters + 1) in equation (3).

^c Φ_m is the measurement objective function ($=\sigma^2 \times n$).

the experimental design. Comparison of m_1 profiles for K_{Pack} and six parameterizations in Figure 4 reveals that simulated m_1 profiles matched the trend of observed m_1 well. In particular, estimated m_1 profiles for all inverse parameterizations show high similarity, despite different K

and ϕ distributions (Figure 2). This result indicates that all parameterizations with highly parameterized models used in this work can characterize the moderately heterogeneous field in the flowcell conditional to the m_1 observations, despite different parameterization of K and ϕ .

[35] Comparison of estimated K fields between K_{IND} and K_{only} reveals that the shape of high- K regions is very similar, but K_{only} has relatively higher hot spots of the highest- K values and the size of high- K regions is bigger than K_{IND} (Figure 2). This result indicates that if ϕ is not included in the estimation, the estimation process has to make significant changes to K in order to achieve calibration conditional on solute transport observations; model structure error leads to parameter error. By contrast, parameterization with K_{IND} needed a bimodal ϕ distribution in order to fit to mean arrival times due to local heterogeneity (i.e., larger m_1 to reduce the objective function). Heterogeneity at these unresolved scales is a possible reason for relatively low to moderate correlation of K - ϕ for K_{zone} : correlation coefficients were 0.54, 0.44, 0.59, and 0.22 from top to bottom layers in Figure 2. This pattern is also evident in comparison with results between K_{Coreg} and K_{KC} . K_{KC} has the higher- K contrast, while K_{Coreg} has the higher porosity contrast (Figure 2). K_{KC} can be considered a special case of K_{Coreg} where the perfect correlation between K and ϕ is formulated based on the Kozeny-Carman equation. This

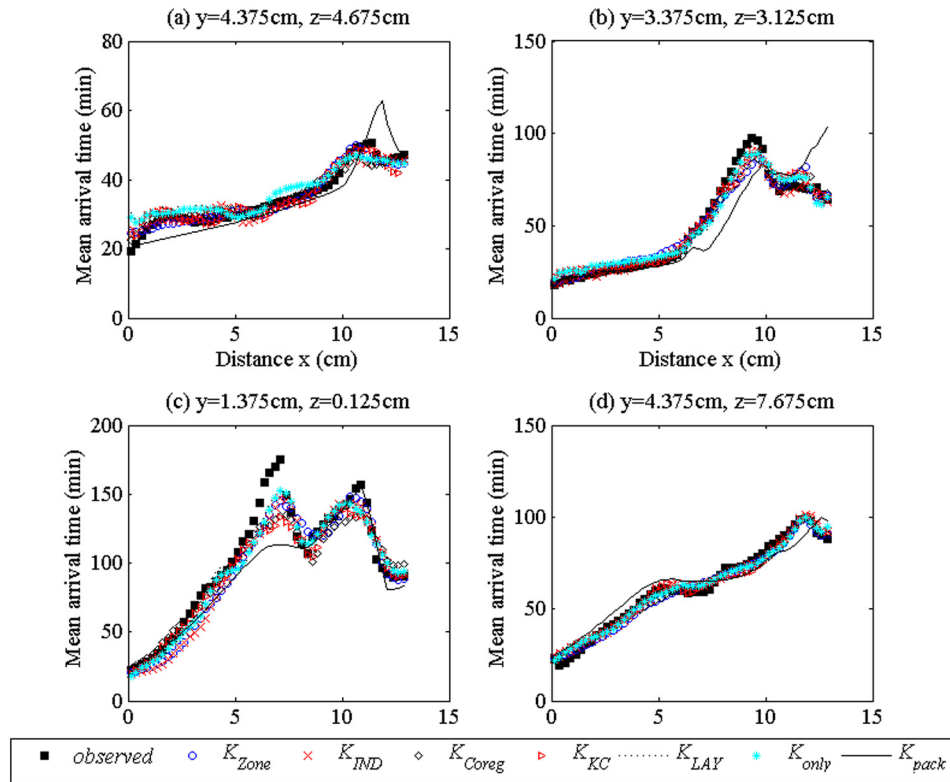


Figure 4. Comparison of mean arrival times (m_1) for the advection only at four different locations at the AVG_{0.016} scale along the x -direction for K_{Pack} and six inverse models. OBS_{0.016} (51,584 m_1) are used for estimating parameters. $x = 0$ represents the location of the beginning of the heterogeneous region, corresponding to $x = 4.5$ cm in Figure 1.

comparison highlights the importance of the local-scale heterogeneity on inverse parameter estimation and the need to account for the impact of the observation scale on transport modeling.

3.2. Transport Modeling and Impact of Scales

3.2.1. Effect of the Estimated Fields of K and ϕ on ADE Results

[36] The calibrated fields of K and ϕ from all six highly parameterized models using OBS_{0.016} and OBS_{1.0} data were used to simulate solute transport with the ADE. Overall, the calibrated fields consist of 12 separate estimated K and ϕ fields (i.e., 6 fields per observed data set). All statistical results of m_1 , m_{2c} , and BTCs over three different averaging scales (AVG_{0.016}, AVG_{0.25}, and AVG_{1.0}) using the ADE are presented in Table 5. The averaging scale refers to the scale at which the volume averaging of concentrations from the ADE results and observations is computed for m_1 and m_{2c} values. First, MAE and RMSE values of m_1 in Table 5 increased compared to those with advection only at the AVG_{0.016} scale (Figure 3). Since dispersivities reflect the small-scale fluctuations (i.e., mechanical dispersion) due to K and ϕ variability at the sub-Darcy scale, it is expected that m_1 estimation with the ADE using calibrated fields based on advection only leads to an increased level of mismatch. Although the mismatch is inevitable, it can be minimized with the highly resolved

estimation of the K and ϕ variability. As shown in m_1 profiles along the main flow x -direction (Figure 5), m_1 errors from the ADE increase with systematic over or under estimation of the observed m_1 values, compared to advective transport results (Figure 4). However, the overall increase of mismatch of m_1 due to dispersion in ADE was not significant. In addition, m_1 profiles for ADE results using calibrated fields from all six different methods are similar. This result shows that using advective transport with the highly parameterized model as a surrogate in inversion on concentration data only results in fields that accurately model transport under the ADE.

3.2.2. Effect of Local Heterogeneity

[37] Inspection of the BTCs at selected points indicates that the majority of the observed BTCs are relatively well matched. However, as indicated by the number of outliers in Table 5, 3–6% of the BTCs at the AVG_{0.016} scale are significantly mismatched. Results of simulated and observed BTCs at the AVG_{0.016} and AVG_{0.25} scales are shown in Figure 6. Estimated fields with OBS_{0.016} and OBS_{1.0} were used for simulations. At all three locations, simulated results match m_1 relatively well. Figure 6a–6f and Table 5 show that highly resolved estimation using only m_1 data is able to accurately reproduce the full BTCs at most locations. However, the agreement between observed and simulated BTCs varies. In particular, observed BTCs at one location (Figures 6g and 6h) show the effect of subgrid scale

Table 5. Statistical Results of Advection-Dispersion Simulations for the Original Sand Packing and Six Calibrated Fields With the Observed Data at the 0.016 cm³ and 1 cm³ Calibration Scales^a

	m_1 (min)			m_{2c} (min ²)		Breakthrough Concentrations, RMSE	Number of Outliers, ^b m_1
	ME (min)	MAE (min)	RMSE (min)	MAE (min ²)	RMSE (min ²)		
<i>0.016 cm³ [51,584]^c</i>							
K_{Pack}	3.41	6.47	9.58	49.9	157.0	0.070	4913
K_{Zone}	−0.03 (−0.23)	5.80 (6.63)	7.74 (8.87)	47.3 (51.1)	144.8 (157.2)	0.066 (0.075)	1913 (2546)
K_{IND}	−0.15 (−0.48)	6.39 (6.80)	8.36 (9.11)	46.9 (54.4)	151.6 (174.4)	0.072 (0.075)	2215 (2639)
K_{Coreg}	−0.16 (−0.93)	6.74 (6.30)	9.03 (8.50)	52.6 (55.1)	171.0 (180.6)	0.075 (0.070)	2802 (2239)
K_{KC}	0.22 (−0.37)	6.18 (5.79)	8.42 (8.14)	51.3 (60.1)	171.2 (195.5)	0.071 (0.066)	2289 (2073)
K_{LAY}	−0.38 (−1.32)	6.87 (6.44)	9.23 (8.64)	46.9 (48.6)	141.8 (147.6)	0.077 (0.072)	2754 (1923)
K_{only}	−0.82 (−0.71)	5.77 (5.74)	7.81 (7.78)	48.5 (49.0)	144.9 (146.0)	0.066 (0.065)	1607 (1628)
<i>0.25 cm³ [3328]</i>							
K_{Pack}	3.59	5.45	7.57	106.5	200.7	0.058	244
K_{Zone}	0.13 (−0.03)	4.56 (4.95)	5.84 (6.44)	100.9 (120.8)	169.4 (198.5)	0.052 (0.059)	29 (55)
K_{IND}	0.01 (−0.26)	5.05 (4.97)	6.41 (6.49)	100.8 (116.5)	175.5 (186.7)	0.058 (0.058)	43 (56)
K_{Coreg}	0.03 (−0.77)	5.42 (4.66)	7.10 (6.16)	114.6 (117.8)	199.2 (205.5)	0.061 (0.054)	110 (69)
K_{KC}	0.35 (−0.31)	4.74 (3.87)	6.23 (5.16)	109.1 (125.9)	195.9 (225.3)	0.056 (0.049)	41 (22)
K_{LAY}	−0.34 (−1.21)	5.69 (4.87)	7.57 (6.56)	114.6 (121.1)	179.1 (186.7)	0.063 (0.056)	104 (42)
K_{only}	−0.79 (−0.70)	4.39 (4.46)	5.84 (5.84)	113.8 (108.2)	180.7 (182.4)	0.049 (0.051)	10 (15)
<i>1 cm³ [832]</i>							
K_{Pack}	3.59	5.40	7.50	106.2	197.5	0.056	58
K_{Zone}	0.13 (−0.31)	4.51 (4.91)	5.77 (6.36)	98.7 (119.0)	164.9 (193.9)	0.051 (0.057)	7 (13)
K_{IND}	0.01 (−0.26)	5.01 (4.93)	6.34 (6.42)	98.9 (114.8)	171.2 (203.0)	0.056 (0.056)	10 (14)
K_{Coreg}	0.03 (−1.10)	5.35 (4.60)	7.02 (6.08)	122.3 (116.1)	195.5 (201.4)	0.059 (0.052)	28 (16)
K_{KC}	0.35 (−0.31)	4.69 (3.80)	6.14 (5.03)	107.0 (124.1)	191.8 (220.9)	0.054 (0.047)	9 (4)
K_{LAY}	−0.34 (−1.39)	5.64 (4.80)	7.50 (6.49)	112.5 (119.7)	175.0 (182.6)	0.061 (0.054)	25 (11)
K_{only}	−0.79 (−0.70)	4.31 (4.39)	5.74 (5.74)	111.8 (106.1)	176.8 (178.4)	0.048 (0.048)	3 (3)

^aThe number of observed m_1 data at the two 0.016 cm³ and 1 cm³ calibration scales are 51,584 and 832, respectively. Statistical values outside and within parenthesis represent results of the forward ADE simulations for the calibrated fields with the observed data at the 0.016 cm³ and 1 cm³ calibration scales, respectively.

^bThe number of outliers is counted if ME is greater than 15 min or less than −15 min.

^cThe number in brackets represents the number of simulated and observed data at each averaging scale.

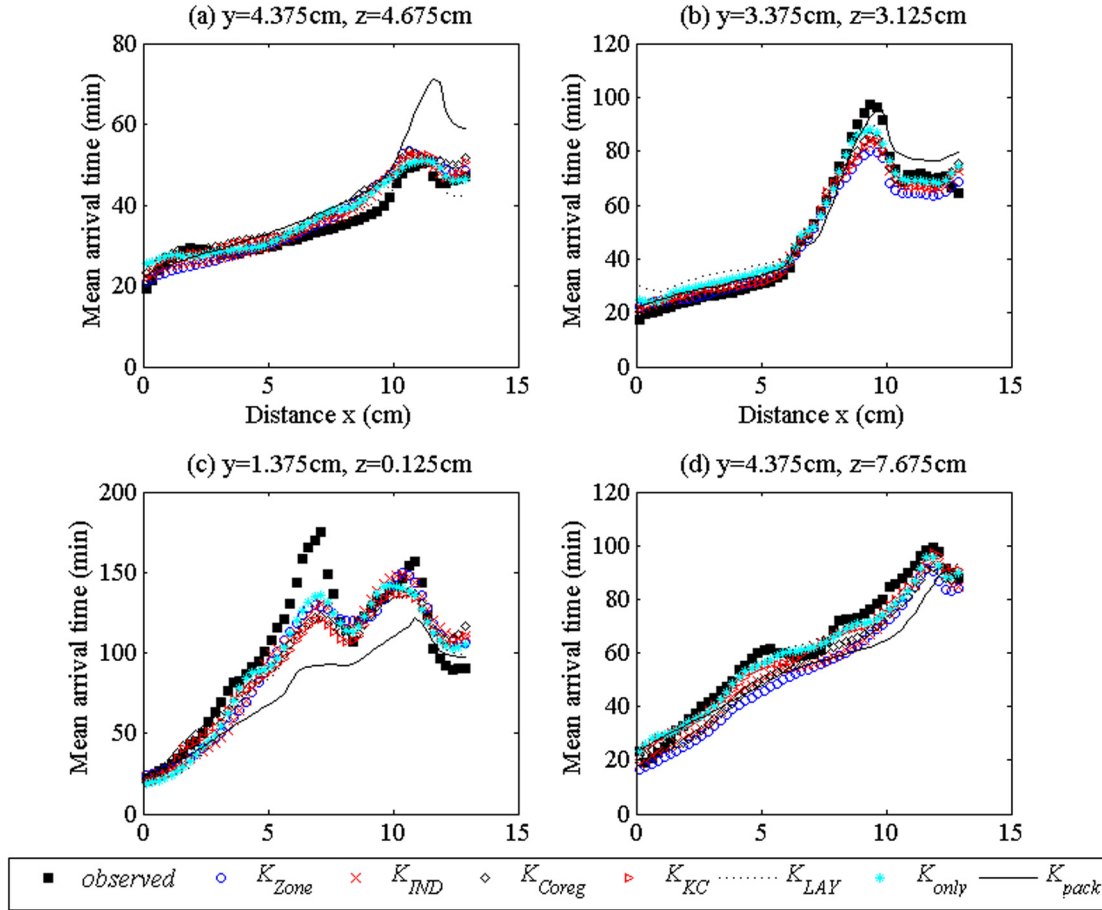


Figure 5. Comparison of mean arrival times (m_1) for the ADE at four different locations at the $AVG_{0.016}$ scale along the x -direction for K_{Pack} and six inverse models. $OBS_{1.0}$ (832 m_1) are used for estimating parameters. $x = 0$ represents the location of the beginning of the heterogeneous region, corresponding to $x = 4.5$ cm in Figure 1.

heterogeneity on BTCs over all scales. Even at the $AVG_{0.016}$ scale, the sharp increase of concentration is followed by gradual increase after ~ 90 min, indicating that this grid cell receives mixed transport signals from upstream regions of differing velocity values, i.e., particles that are younger or older due to their upstream travel history. These particles can flow into one cell by simple discretization effects (e.g., misalignment between the cell boundaries and streamtubes, and/or numerical transverse dispersion), or by actual transverse dispersion. Hence, tailing and irregular shapes of BTCs in Figures 6g and 6h are mainly attributed to local heterogeneity. It is also noted that highly parameterized calibrated models with fixed dispersivities are able to capture some of the effect of local heterogeneities in a 3-D real packing problem.

3.2.3. Impact of Observation Scales

[38] For ADE results in Table 5, the number of outliers (i.e., $ME > 15$ min or $ME < -15$ min) from $AVG_{0.016}$ to $AVG_{0.25}$ scales decreased dramatically by 25 \sim 160 times, compared to the reduction of the averaging volume by a factor of 16. However, all statistics at the $AVG_{0.25}$ and $AVG_{1.0}$ scales are very similar for the calibrated fields with both sets of observed data ($OBS_{0.016}$ and $OBS_{1.0}$), indicating that

the $AVG_{0.25}$ scale can be considered a representative elementary volume (REV).

[39] Comparison of statistics in Table 5 for K_{Coreg} and K_{KC} as well as K_{LAY} and K_{only} shows that MAE and RMSE values of m_1 for K_{KC} and K_{only} were lower than those for K_{Coreg} and K_{LAY} , respectively. Note that K_{Coreg} and K_{LAY} have more calibrated parameters related to ϕ , compared to K_{KC} and K_{only} . In contrast to results with advection only where K_{Coreg} and K_{LAY} have lower objective functions (Table 4), results with ADE simulations reveal that calibration of parameters related to ϕ with advection only should be carefully done to avoid over-fitting. Comparison of the calibrated K - ϕ fields using $OBS_{0.016}$ and $OBS_{1.0}$ m_1 data (Figures 2 and 7) show that calibrated K fields are quite similar, while ϕ fields for K_{Zone} and K_{IND} using $OBS_{1.0}$ m_1 data (Figure 7) represent less contrast because ϕ at a majority of pilot points was not involved in calibration (i.e., not included in the calibration solution space) due to a small number of observations (832 m_1) compared to the number of parameters (1954). It should be noted that statistical results for calibrated fields with $OBS_{1.0}$ (832 m_1) and $OBS_{0.016}$ (51584 m_1) are in a good agreement at all three AVG scales, while individual statistical values for each

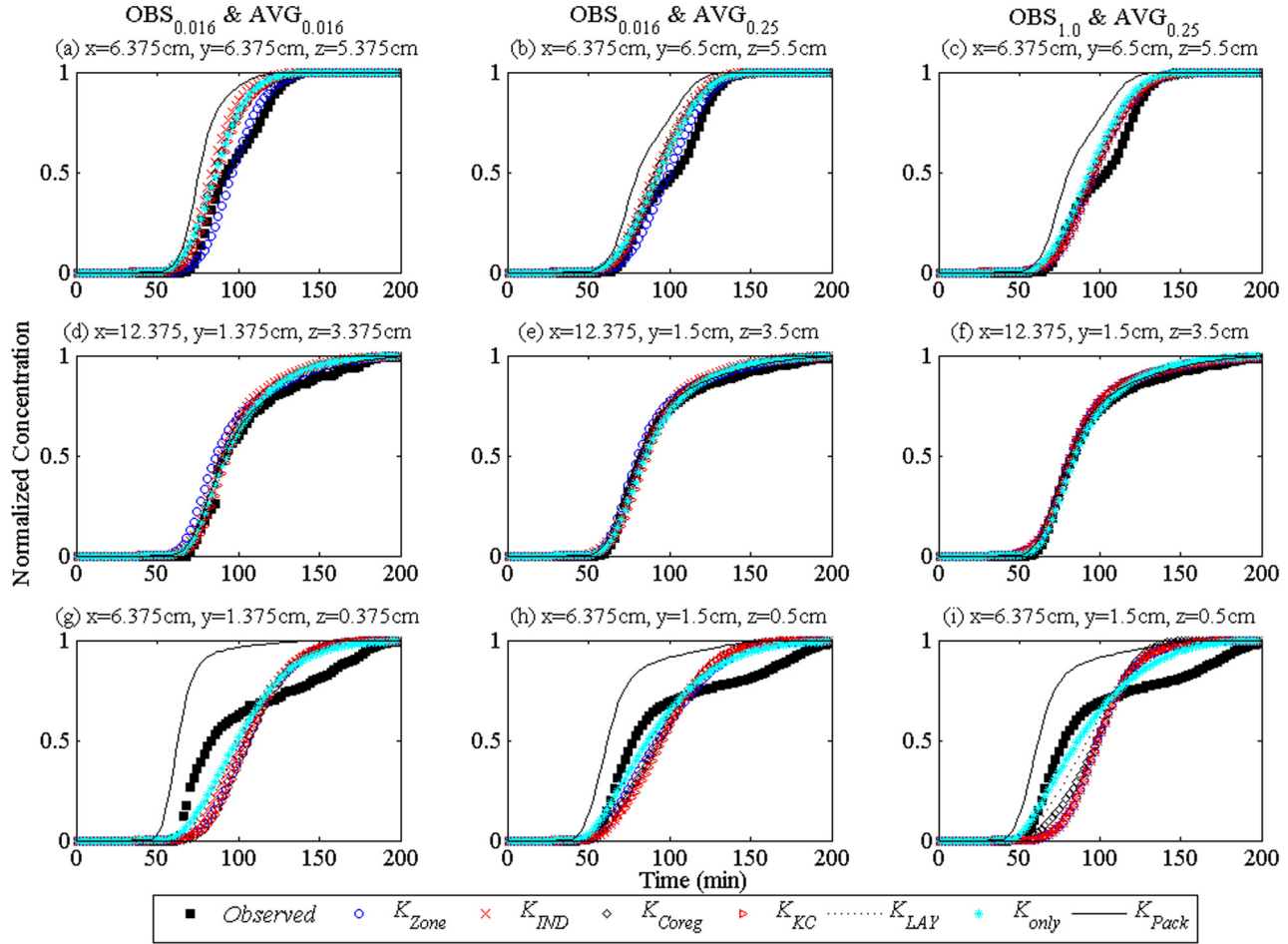


Figure 6. Comparison of measured and predicted BTCs at the $AVG_{0.016}$ (a, d, and g) and $AVG_{0.25}$ scales (b, c, e, f, h, and i). Parameter fields calibrated with $OBS_{0.016}$ (51584 m_1) are used for the forward ADE simulations in the left and middle columns and fields calibrated with $OBS_{1.0}$ (832 m_1) are used in the right column.

K and ϕ relation vary (Table 5). This result indicates that $OBS_{0.016}$ data (below a REV scale) are not necessarily helpful to account for the impact of local heterogeneity on transport.

3.3. Impact of Number of Observations and Parameters

[40] The impact of the number of observations and parameters is investigated by estimating the parameters at the $AVG_{1.0}$ scale, while varying the number of $OBS_{1.0}$ m_1 (48, 96, 224, 416, and 832) and the number of parameters (256, 456, 616, 1065, and 1952) corresponding to 112 to 896 pilot points in the heterogeneous region. All pilot points are uniformly distributed in the 3-D field and K and ϕ are independently estimated at every pilot point (i.e., K_{IND} parameterization with varying pilot points). For example, 256 parameters are located at 120 pilot points with 15 pilot points in the heterogeneous region and one pilot point in the homogeneous region for each layer. For each number of parameters with different observations, initial parameter values are calibrated results from previous estimation with $OBS_{1.0}$ (832 m_1). Calibration of m_1 data was performed with MODPATH (advection only).

[41] RMSE values of m_1 at the $AVG_{0.016}$ and $AVG_{1.0}$ scales are shown in Figure 8. At both scales, RMSE values are distinctly different between 1065 and 616 parameters. For 1065 parameters, pilot points are located every two 1 cm^3 cells; hence, the distance between pilot points was less than 2 cm and shorter than the correlation length (4.2 cm) of the spherical variogram model. For 616 parameters, pilot points are located every four cells, so the distance between pilot points is close to the correlation length. Hence, the number of pilot points within the correlation length is an important criterion for interpolation from pilot points to neighboring numerical cells, as demonstrated in analytical work for the interpolation uncertainty as a function of the data density on regular grids [Cirpka and Nowak, 2003]. A practical guideline on the pilot point spacing is well-summarized in Doherty *et al.* [2010]. The RMSE values with 456 parameters are higher than those with 256 parameters (Figure 8). Overall, Figure 8 shows that for the cases studied in this work pilot point spacing must be less than the correlation length and accuracy in the calibrations is more strongly influenced by the number of parameters than the number of observations.

[42] A subset of observation (96 m_1 data at $OBS_{1.0}$) was used to test a more realistic case of limited observations.

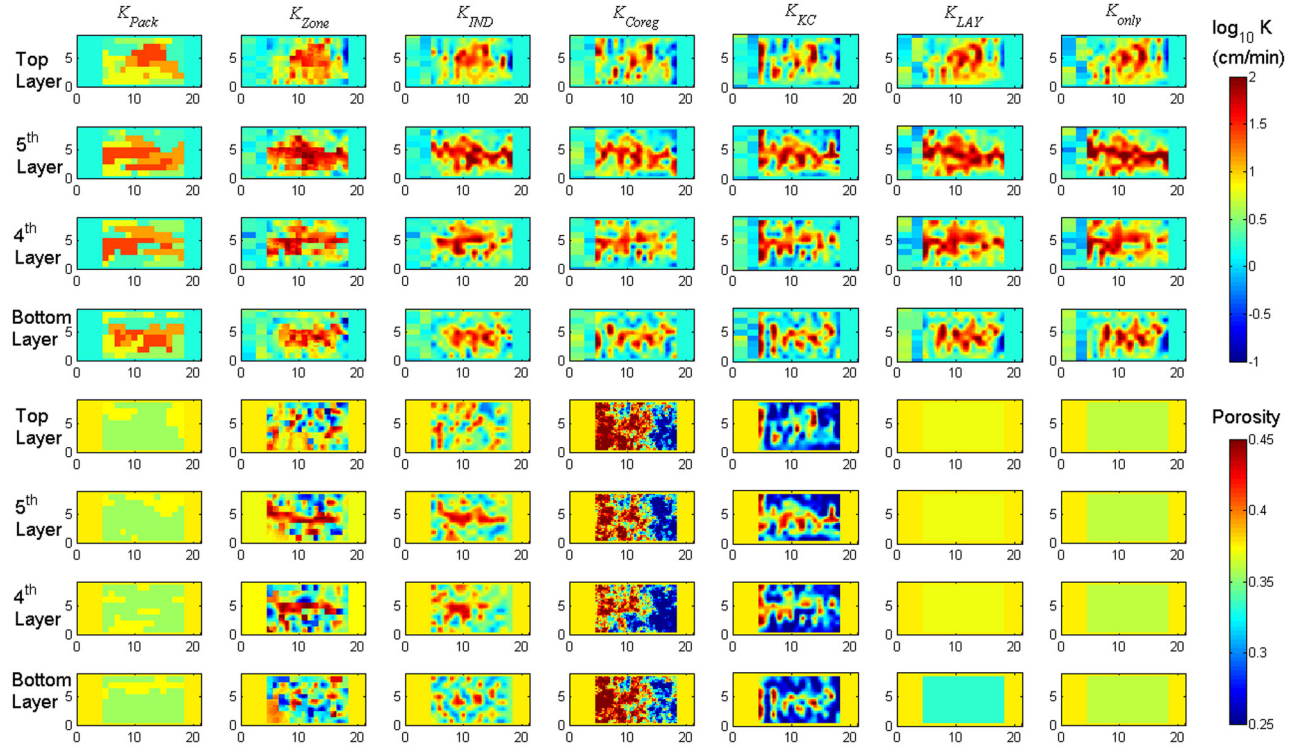


Figure 7. K and porosity distributions for experimental packing pattern and estimation for all six K -porosity relationships in four different horizontal layers. Highly parameterized models with 896 pilot points in the heterogeneous region and 160 pilot points in the homogeneous region were calibrated with $OBS_{1.0}$ (832 m_1).

The subset of 96 observed data mimics 12 multiple observation wells per each layer in the 3-D field. Highly parameterized models with 896 pilot points in the heterogeneous region and one pilot point in the homogeneous region were used to calibrate K - ϕ fields (Figure 9). Observation wells are uniformly distributed. Comparison of m_1 profiles for K_{Pack} and six different methods demonstrates that simulated m_1 profiles matched the trend of observed m_1 relatively well along the main flow direction (Figures 10a and 10c) where observation wells are located. However, the discrepancy between m_1 profiles increased in the center of the flowcell (Figure 10b) where there are no observation wells. In particular, estimated K distributions (Figure 9) show that the resulting K contrast is much smaller than it was for fields calibrated with larger amounts of observation data (Figure 7), but all six highly parameterized models captured the key features of packing patterns relatively well. Relative composite sensitivities, which are the product of composite sensitivity and calibrated parameter value, show that sensitivity values for a majority of K in the heterogeneous region were similar in magnitude, while sensitivity values for ϕ were relatively lower than those for K (results not shown). This comparison indicates that in well-defined or characterized regions (i.e., high K along the center of the domain in this study) more sensitive parameters such as K rather than ϕ would be enough to calibrate the observed data. Hence, the calibration solution space whose dimension was the same as the number of observations (i.e., 96) was mostly related to K . This also explains the different ϕ distributions between K_{Coreg} and K_{KC} because parameters

related to ϕ in K_{Coreg} were not included within the calibration solution space, resulting in correlation coefficients similar to an initial value (i.e., 0.5). Sensitivity and parameter importance or data worth associated with the design of monitoring network and predictive uncertainty reduction will be explored in the future.

4. Summary and Conclusions

[43] Inverse estimation of K and ϕ based on advection only using the unique MRI concentration data set was performed to explore the application of highly parameterized models in a 3-D heterogeneous transport experiment. Comparison of the inverse estimations from the six different K and ϕ relationships show similar statistical measures of mean arrival times (m_1) and very similar m_1 spatial estimates as shown in m_1 profiles along the main flow direction (Figure 4). All six models recovered the main physical feature of the K pattern in K_{Pack} (i.e., highly interconnected high- K sands along the central part of the flowcell), though they differed in specific details of the K and ϕ patterns. This work highlights that complete knowledge of the 3-D field in the experimental design (K_{Pack}) does not provide good predictions, and inverse models with all different parameterizations can considerably improve upon the naïve application of the experimental setup. Model comparison also suggests that site-specific conceptual model information such as zonal information or a particular K - ϕ relationship will help constrain the calibration and create more reliable predictions. Compared to the most commonly used

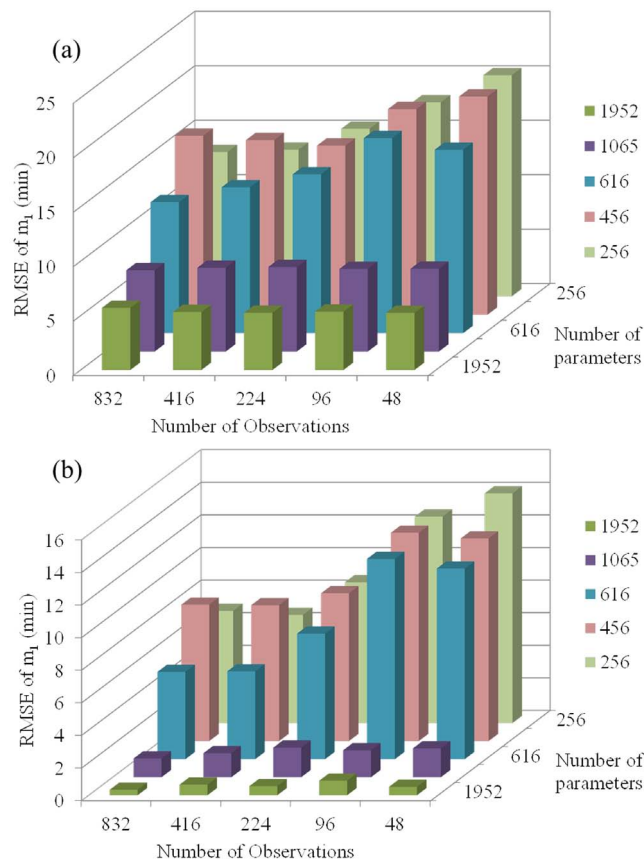


Figure 8. RMSE values of mean arrival times using different number of observations (48 to 832 m_1) and parameters (256 to 1952 K and porosity) at the (a) $AVG_{0.016}$ and (b) $AVG_{1.0}$ scales. Parameter fields (K_{IND}) calibrated with $OBS_{1.0}$ (832 m_1) are used for the advection only using MODPATH. Parameter fields for each set of the number of observations and parameters are calibrated separately.

K_{KC} and K_{only} (and/or K_{LAY}) approaches with less parameterization, the K_{IND} approach within the framework of regularized highly parameterized models can be a practical solution to estimate aquifer properties. The K_{IND} approach produced the estimations most similar to those obtained with K_{Zone} for all cases considered. Model selection criteria such as AIC_c and KIC also show that the rank of models with different parameterizations follows the order of the objective functions. Future work is necessary to provide guidance for the application of the model selection criteria to highly parameterized models.

[44] Forward transport modeling of the ADE using calibrated K and ϕ fields based on advection only also shows that BTCs at a majority of locations as well as statistical results of m_1 , m_{2c} and BTCs for all six relationships are very similar. This comparison shows that there is not a single set of model parameters or a single K - ϕ relationship that leads to a single, best representation of the real experimental condition, and the estimation conditioned only to tracer breakthrough data can be nonunique, but, further study is needed to confirm this, in particular, for cases with direct K and ϕ measurements at limited locations. The

difference between forward model and experimental results of BTCs increased with decreasing the averaging scale of the concentration, demonstrating the significance of small-scale variability in aquifer properties causing uncertainty in transport predictions. Even with the K_{Zone} set where packing (i.e., zonal) information is used as the basis for a highly parameterized model, small-scale variations in packing and mixing of the sands cause significant deviations from the expected transport results as seen in the number of BTC outliers. These results are consistent with similar impacts of fine-scale structures on transport results in heterogeneous media observed in other detailed experimental investigations [e.g., Klise *et al.*, 2008].

[45] In this study, ME, MAE, and RMSE values at the $AVG_{0.25}$ and $AVG_{1.0}$ scales are similar, but different from those at the $AVG_{0.016}$ scale (Table 5), indicating that a REV scale is approximately at $AVG_{0.25}$. Another important feature in these statistical results is that all statistics with the $OBS_{0.016}$ (51584 m_1) and $OBS_{1.0}$ (832 m_1) data are strikingly similar across all averaging scales. This suggests that even with much larger number of observations at the $OBS_{0.016}$ scale local-scale heterogeneity features below the REV are not properly incorporated into parameterized models used in this work. Hence, a proper scale of observation for transport problems (e.g., concentration and/or mass flux) should be considered such as at an REV scale and then local fluctuations can be separately treated as random or correlated deviations [Alcolea *et al.*, 2008; Hu *et al.*, 2009]. This work strongly suggests that uncaptured variabilities during model calibration (e.g., local-heterogeneity features in this study) inevitably exist, highlighting the importance of addressing prediction uncertainty following calibration in modeling practice.

[46] Although transport parameters such as dispersivities were not included during calibration in this study, the forward modeling results with calibrated fields indicate that even in a moderately heterogeneous field, BTC tailing can be observed at a small scale such as $AVG_{0.016}$ (Figure 6). In low to moderate K contrast fields, predicted and experimental BTCs have varied considerably [Barth *et al.*, 2001; Nowak and Cirpka, 2006] and these differences have been attributed to variability in the measured hydraulic conductivity [Barth *et al.*, 2001] and small-scale heterogeneity [Nowak and Cirpka, 2006]. In this study, even at the $AVG_{0.016}$ scale, observed BTCs can exhibit distinct features such as a fast increase of concentration followed by a more gradual increase due to local heterogeneity at high spatial resolution and mixing of flow paths (i.e., stream-tubes). Recently, a limited set of BTCs has been calibrated with different parameterizations such as inclusion of dispersivity as a fitting parameter in 2-D tens of meter-long tank [Nowak and Cirpka, 2006], mass transfer between a mobile and an immobile in a 3-D layered heterogeneous field and spatial variability of porosity [Riva *et al.*, 2008], time-fractional ADE in 2-D slab [Major *et al.*, 2011], and inclusion of nested variograms at two different scales in 2-D synthetic heterogeneous fields [Alcolea *et al.*, 2008]. In particular, this study shows that if the conditions are met to use an advection-only transport in inversion, ADE transport with calibrated fields based on the advection only can be accurately modeled, highlighting computational and practical efficiency of inverse estimation with advection-only as

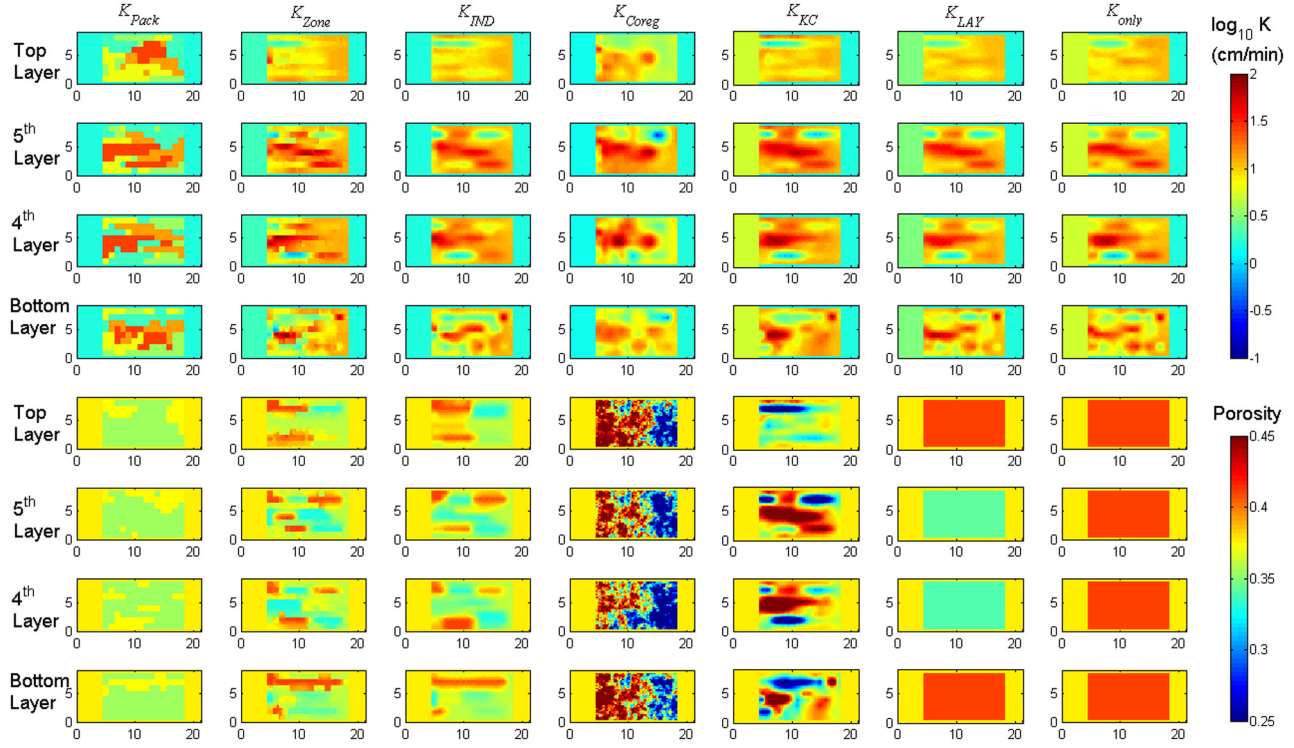


Figure 9. K and porosity distributions for K_{Pack} and estimation for all six K -porosity relationships in four different horizontal layers. Highly parameterized models with 896 pilot points in the heterogeneous region and one pilot point in the homogeneous region were calibrated with a subset of $OBS_{1,0}$ at 12 observation wells (i.e., 96 m_1).

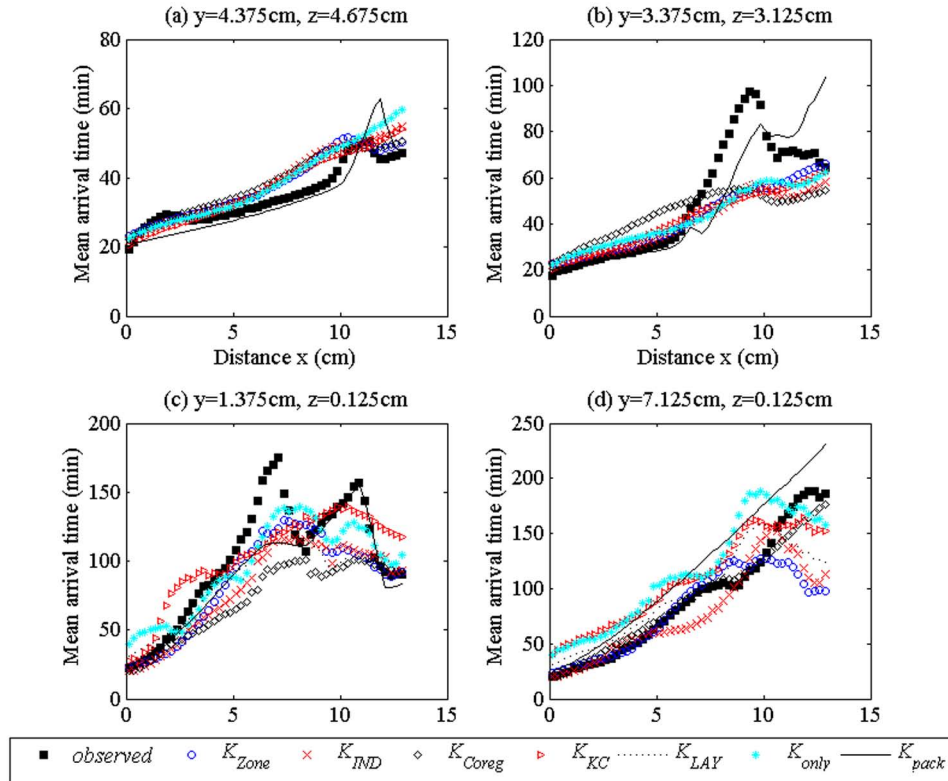


Figure 10. Comparison of mean arrival times (m_1) for the advection only at four different locations at the $AVG_{0.016}$ scale along the x -direction for K_{Pack} and all six K -porosity relationships. A subset of $OBS_{1,0}$ at 12 observation wells (i.e., 96 m_1) is used for estimating parameters using MODPATH. $x = 0$ represents the location of the beginning of the heterogeneous region, corresponding to $x = 4.5$ cm in Figure 1.

a surrogate for more expensive ADE solutions to provide results and initial conditions for further system analysis.

[47] **Acknowledgments.** This material is based upon work supported as part of the Center for Frontiers of Subsurface Energy Security, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0001114. Sandia National Laboratories is a multiprogram laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. We also acknowledge the effort of the Associate Editor, Wolfgang Nowak, Matthew Tonkin, and two anonymous reviewers for their careful and constructive reviews, which led to significant improvement of our manuscript.

References

- Alcolea, A., J. Carrera, and M. Medina (2006), Inversion of heterogeneous parabolic-type equations using the pilot-points method, *Int. J. Numer. Methods Fluids*, 51, 963–980, doi:10.1002/fld.1213.
- Alcolea, A., J. Carrera, and A. Medina (2008), Regularized pilot points method for reproducing the effect of small scale variability: Application to simulations of contaminant transport, *J. Hydrol.*, 355(1–4), 76–90, doi:10.1016/j.jhydrol.2008.03.004.
- Bailey, R. T., and D. Bau (2011), Estimating spatially-variable first-order rate constants in groundwater reactive transport systems, *J. Contam. Hydrol.*, 122, 104–121.
- Baines, S. J., and R. H. Worden (2004), The long-term fate of CO₂ in the subsurface: Natural analogues for CO₂ storage, in *Geological Storage of Carbon Dioxide*, vol. 233, edited by S. J. Baines and R. H. Worden, pp. 59–85, Geol. Soc. of London, London.
- Barth, G. R., M. C. Hill, T. H. Illangasekare, and H. Rajaram (2001), Predictive modeling of flow and transport in a two-dimensional intermediate scale, heterogeneous porous medium, *Water Resour. Res.*, 37(10), 2503–2512.
- Beven, K., and A. Binley (1992), The future of distributed models—Model calibration and uncertainty prediction, *Hydrol. Processes*, 6(3), 279–298, doi:10.1002/hyp.3360060305.
- Carrera, J. (1987), State of the art of the inverse problem applied to the flow and solute transport problems, in *Groundwater Flow and Quality Modeling*, edited by E. Custodio, A. Gurgui, and J. P. Lobo Ferreira, pp. 549–585, D. Reidel, Boston, Mass.
- Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten (2005), Inverse problem in hydrogeology, *Hydrogeol. J.*, 13(1), 206–222, doi:10.1007/s10040-004-0404-7.
- Castagna, M., and A. Bellin (2009), A Bayesian approach for inversion of hydraulic tomographic data, *Water Resour. Res.*, 45, W04410, doi:10.1029/2008WR007078.
- Cirpka, O. A., and W. Nowak (2003), Dispersion on kriged hydraulic conductivity fields, *Water Resour. Res.*, 39(2), 1027, doi:10.1029/2001WR000598.
- de Marsily, G., J. P. Delhomme, F. Delay, and A. Buoro (1999), 40 years of inverse problems in hydrogeology, *Earth Planet. Sci.*, 329(2), 73–87.
- Department of Energy (DOE) (2007), Basic research needs for geosciences: Facilitating 21st century energy systems, report, Off. of Basic Energy Sci., Washington, D. C.
- Deutsch, C. V., and A. G. Journel (1998), *GSLIB: Geostatistical Software Library and User's Guide*, Oxford Univ., New York.
- Doherty, J. (2003), Ground water model calibration using pilot points and regularization, *Ground Water*, 41(2), 170–177.
- Doherty, J. (2010), PEST: Model-independent parameter estimation, Watermark Numer. Comput., Brisbane, Queensland, Australia. [Available at <http://www.pesthomepage.org>].
- Doherty, J. (2012), Addendum to the PEST manual, Watermark Numer. Comput., Brisbane, Queensland, Australia. [Available at <http://www.pesthomepage.org>].
- Doherty, J., and R. J. Hunt (2010), Approaches to highly parameterized inversion: A guide to using PEST for groundwater-model calibration, *U.S. Geol. Surv. Sci. Invest. Rep.*, 2010-5169, 60 p.
- Doherty, J., M. F. Fienen, and R. J. Hunt (2010), Approaches to highly parameterized inversion: Pilot-point theory, guidelines, and research directions, *U.S. Geol. Surv. Sci. Invest. Rep.*, 2010-5168, 36 p.
- Fienen, M., R. Hunt, D. Krabbenhoft, and T. Clemo (2009), Obtaining parsimonious hydraulic conductivity fields using head and transport observations: A Bayesian geostatistical parameter estimation approach, *Water Resour. Res.*, 45, W08405, doi:10.1029/2008WR007431.
- Fu, J., and J. J. Gómez-Hernández (2009), Uncertainty assessment and data worth in groundwater flow and mass transport modeling using a blocking Markov chain Monte Carlo method, *J. Hydrol.*, 364, 328–341, doi:10.1016/j.jhydrol.2008.11.014.
- Gómez-Hernández, J. J., H. J. Hendricks Franssen, and E. F. Cassiraga (2001), Stochastic analysis of flow response in a three-dimensional fractured rock mass block, *Int. J. Rock Mech. Min. Sci.*, 38, 31–44.
- Gómez-Hernández, J. J., H. J. Hendricks Franssen, and A. Sahuquillo (2003), Stochastic conditional inverse modeling of subsurface mass transport: A brief review of the self-calibrating method, *Stochastic Environ. Res. Risk Assess.*, 17, 319–328, doi:10.1007/s00477-003-0153-5.
- Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, Oxford Univ., New York.
- Guadagnini, A., and S. P. Neuman (1999), Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains: 1. Theory and computational approach, *Water Resour. Res.*, 35(10), 2999–3018, doi:10.1029/1999WR900160.
- Harbaugh, A. W., E. R. Banta, M. C. Hill, and M. G. McDonald (2000), MODFLOW-2000, the U.S. Geological Survey modular ground-water model—User guide to modularization concepts and the ground-water flow process, *U.S. Geol. Surv. Open File Rep.*, 00-92, 121 pp.
- Hendricks Franssen, H. J., and J. J. Gómez-Hernández (2002), 3D inverse modelling of groundwater flow at a fractured site using a stochastic continuum model with multiple statistical populations, *Stochastic Environ. Res. Risk Assess.*, 16(2), 155–174, doi:10.1007/s00477-002-0091-7.
- Hendricks Franssen, H. J., J. J. Gómez-Hernández, J. E. Capilla and A. Sahuquillo (1999), Joint simulation of transmissivity and storativity fields conditional to steady-state and transient hydraulic head data, *Adv. Water Resour.*, 23(1), 1–13.
- Hendricks Franssen, H. J., A. Alcolea, M. Riva, M. Bakr, N. van der Wiel, F. Stauffer, and A. Guadagnini (2009), A comparison of seven methods for the inverse modeling of groundwater flow: Application to the characterisation of well catchments, *Adv. Water Resour.*, 32, 851–872.
- Herckenrath, D., C. D. Langevin, and J. Doherty (2011), Predictive uncertainty analysis of a saltwater intrusion model using null-space Monte Carlo, *Water Resour. Res.*, 47, W05504, doi:10.1029/2010WR009342.
- Hernandez, A. F., S. P. Neuman, A. Guadagnini, and J. Carrera (2006), Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media, *Water Resour. Res.*, 42, W05425, doi:10.1029/2005WR004449.
- Hosseini, A. H., C. V. Deutsch, C. A. Mendoza, and K. W. Biggar (2011), Inverse modeling for characterization of uncertainty in transport parameters under uncertainty of source geometry in heterogeneous aquifers, *J. Hydrol.*, 405(3–4), 402–416.
- Hu, B. X., M. M. Meerschaert, W. Barrash, D. Hyndman, C. He, X. Li, and L. Guo (2009), Examining the influence of heterogeneous porosity fields on conservative solute transport, *J. Contam. Hydrol.*, 108, 77–88.
- Hu, L. Y. (2000), Gradual deformation and iterative calibration of gaussian-related stochastic models, *Math. Geol.*, 32(1), 87–108.
- Huang, H., B. X. Hu, X.-H. Wen, and C. Shirley (2004), Stochastic inverse mapping of hydraulic conductivity and sorption partitioning coefficient fields conditioning on nonreactive and reactive tracer test data, *Water Resour. Res.*, 40, W01506, doi:10.1029/2003WR002253.
- Huber, E., H. Hendricks-Franssen, H. Kaiser, and F. Stauffer (2011), The role of prior model calibration on predictions with ensemble Kalman filter, *Ground Water*, 49, 845–858, doi:10.1111/j.1745-6584.2010.00784.x.
- Hunt, R. J., J. Luchette, W. A. Shreuder, J. Rumbaugh, J. Doherty, M. J. Tonkin, and D. Rumbaugh (2010), Using the cloud to replenish parched groundwater modeling efforts, *Ground Water*, 48(3), 360–365, doi:10.1111/j.1745-6584.2010.00699.x.
- Hurvich, C. M., and C.-L. Tsai (1989), Regression and time series model selection in small sample, *Biometrika*, 76(2), 99–104.
- Jiang, Y., A. D. Woodbury, and S. Painter (2004), Full-Bayesian inversion of the Edwards aquifer, *Ground Water*, 42(5), 724–733, doi:10.1111/j.1745-6584.2004.tb02726.x.
- Kashyap, R. L. (1982), Optimal choice of AR and MA parts in autoregressive moving average models, *IEEE Trans. Pattern Anal. Machine Intell.*, 4(2), 99–104.
- Keating, E. H., J. Doherty, J. A. Vrugt, and Q. Kang (2010), Optimization and uncertainty assessment of strongly non-linear groundwater models with high parameter dimensionality, *Water Resour. Res.*, 46, W10517, doi:10.1029/2009WR008584.

- Kitanidis, P. K. (1996), On the geostatistical approach to the inverse problem, *Adv. Water Resour.*, 19(6), 333–342, doi:10.1016/0309-1708(96)00005-x.
- Klise, K. A., V. C. Tidwell, and S. A. McKenna (2008), Comparison of laboratory-scale solute transport visualization experiments with numerical simulation using cross-bedded sandstone, *Adv. Water Resour.*, 31(12), 1731–1741.
- Kollat, J. B., P. M. Reed, and R. M. Maxwell (2011), Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics, *Water Resour. Res.*, 47, W02529, doi:10.1029/2010WR009194.
- Lavenue, M., and G. de Marsily (2001), Three-dimensional interference test interpretation in a fractured aquifer using the pilot point inverse method, *Water Resour. Res.*, 37(11), 2659–2675.
- Li, W., W. Nowak, and O. A. Cirpka (2005), Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown, *Water Resour. Res.*, 41, W08403, doi:10.1029/2004WR003874.
- Li, W., A. Englert, O. A. Cirpka, J. Vanderborght, and H. Vereecken (2007), Two-dimensional characterization of hydraulic heterogeneity by multiple pumping tests, *Water Resour. Res.*, 43, W04433, doi:10.1029/2006WR005333.
- Li, W., A. Englert, O. A. Cirpka, and H. Vereecken (2008), Three-dimensional geostatistical inversion of flowmeter and pumping test data, *Ground Water*, 46(2), 193–201.
- Llopis-Albert, C., and J. E. Capilla (2010), Stochastic inverse modelling of hydraulic conductivity fields taking into account independent stochastic structures: A 3D case study, *J. Hydrol.*, 391(3–4), 277–288.
- Luo, J., O. A. Cirpka, and P. K. Kitanidis (2006), Temporal-moment matching for truncated breakthrough curves for step or step-pulse injection, *Adv. Water Resour.*, 29, 1306–1313.
- Major, E., D. A. Benson, J. Revielle, H. Ibrahim, A. Dean, R. M. Maxwell, E. Poeter, and M. Dogan (2011), Comparison of Fickian and temporally nonlocal transport theories over many scales in an exhaustively sampled sandstone slab, *Water Resour. Res.*, 47, W10519, doi:10.1029/2011WR010857.
- McLaughlin, D., and L. R. Townley (1996), A reassessment of the groundwater inverse problem, *Water Resour. Res.*, 32(5), 1131–1161, doi:10.1029/96WR00160.
- Moore, C., and J. Doherty (2005), Role of the calibration process in reducing model predictive error, *Water Resour. Res.*, 41, W05020, doi:10.1029/2004WR003501.
- Murakami, H., X. Chen, M. S. Hahn, Y. Liu, M. L. Rockhold, V. R. Vermeul, J. M. Zachara, and Y. Rubin (2010), Bayesian approach for three-dimensional aquifer characterization at the Hanford 300 Area, *Hydrol. Earth Syst. Sci.*, 14, 1989–2001, doi:10.5194/hess-14-1989-2010.
- Nowak, W. (2009), Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator, *Water Resour. Res.*, 45, W04431, doi:10.1029/2008WR007328.
- Nowak, W., and O. A. Cirpka (2006), Geostatistical inference of hydraulic conductivity and dispersivities from hydraulic heads and tracer data, *Water Resour. Res.*, 42, W08416, doi:10.1029/2005WR004832.
- Poeter, E. P., and M. C. Hill (2007), MMA, a computer code for multimodel analysis, *U.S. Geol. Surv. Tech. Methods*, TM6-E3, 113 pp.
- Pollock, D. W. (1994), User's guide for MODPATH/MODPATHPLOT, version 3: A particle tracking post-processing package for MODFLOW, the US Geological Survey finite-difference groundwater flow model, *U.S. Geol. Surv. Open File Rep.*, 94-464, 249 pp.
- Renard, B., D. Kavetski, G. Kuczera, M. A. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328.
- Riva, M., A. Guadagnini, D. Fernandez-Garcia, X. Sanchez-Vila, and T. Ptak (2008), Relative importance of geostatistical and transport models in describing heavily tailed breakthrough curves at the Lauswiesen site, *J. Contam. Hydrol.*, 101, 1–13.
- Riva, M., A. Guadagnini, F. De Gaspari, and A. Alcolea (2010), Exact sensitivity matrix and influence of the number of pilot points in the geostatistical inversion of moment equations of groundwater flow, *Water Resour. Res.*, 46, W11513, doi:10.1029/2009WR008476.
- Rubin, Y., and G. Dagan (1987), Stochastic identification of transmissivity and effective recharge in steady state groundwater flow: 1. Theory, *Water Resour. Res.*, 23(7), 1192–1200.
- Schöniger, A., W. Nowak, and H.-J. Hendricks Franssen (2012), Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography, *Water Resour. Res.*, 48, W04502, doi:10.1029/2011WR010462.
- South Florida Water Management District (2005), *Theory Manual, Regional Simulation Model*, West Palm Beach, Fla.
- Tonkin, M. J., and J. Doherty (2005), A hybrid regularized inversion methodology for highly parameterized environmental models, *Water Resour. Res.*, 41, W10412, doi:10.1029/2005WR003995.
- Tonkin, M., and J. Doherty (2009), Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, *Water Resour. Res.*, 45, W00B10, doi:10.1029/2007WR006678.
- Tonkin, M., J. Doherty, and C. Moore (2007), Efficient nonlinear predictive error variance for highly parameterized models, *Water Resour. Res.*, 43, W07429, doi:10.1029/2006WR005348.
- White, M. D., and M. Oostrom (2006), STOMP, Subsurface transport over multiple phases, Version 4, User's Guide, *Rep. PNNL-15782*, Pac. Northwest Natl. Lab., Richland, Wash.
- Woodbury, A. D., and T. J. Ulrych (2000), A full-Bayesian approach to the groundwater inverse problem for steady state flow, *Water Resour. Res.*, 36(8), 2081–2093, doi:10.1029/2000WR000086.
- Ye, M., P. D. Meyer, and S. P. Neuman (2008), On model selection criteria in multimodel analysis, *Water Resour. Res.*, 44, W03428, doi:10.1029/2008WR006803.
- Yoon, H., C. Zhang, C. J. Werth, A. J. Valocchi, and A. G. Webb (2008), Three dimensional characterization of water flow in heterogeneous porous media using magnetic resonance imaging, *Water Resour. Res.*, 44, W06405, doi:10.1029/2007WR006213.
- Young, S. C., J. Doherty, T. Budge, and N. Deeds (2010), Application of PEST to re-calibrate the groundwater availability model for the Edwards-Trinity (Plateau) and Pecos Valley aquifers, *Rep. 2010-1436*, Tex. Water Dev. Board, Austin, Tex.
- Zachara, J. M. (2010), Multi-scale mass transfer processes controlling natural attenuation and engineered remediation: An IFRC focused on Hanford's 300 area uranium plume, *Rep. PNNL-19209*, Pac. Northwest Natl. Lab., Richland, Wash.
- Zhang, C. Y., C. J. Werth, and A. G. Webb (2007), Characterization of NAPL source zone architecture and dissolution kinetics in heterogeneous porous media using magnetic resonance imaging, *Environ. Sci. Technol.*, 41, 3672–3678.
- Zhang, C., H. Yoon, C. J. Werth, A. J. Valocchi, N. B. Basu, and J. W. Jawitz (2008), Evaluation of simplified mass transfer models to simulate the impacts of source zone architecture on nonaqueous phase liquid dissolution in heterogeneous porous media, *J. Contam. Hydrol.*, 102(1–2), 48–59.